

The Impact of Layer Normalisation Methods on Training Effectiveness and Performance of Decoder-Only Transformer Architectures

Research Question:

"How necessary are different normalisation techniques (LN, RMSN, PN) for decoder-only transformers (GPT-2), and which configuration yields the best performance in terms of training effectiveness and downstream tasks?"

Word Count: 3793

Subject: Computer Science [Higher Level] Extended Essay

1. Introduction	4
2. Background Information	6
Transformer Architecture Overview	6
3. Theoretical Framework	11
3.1 Layer Normalisation (LN)	11
3.2 Root Mean Square Normalisation (RMSN)	14
3.3 Power Normalisation (PN)	16
3.4 Comparison of Normalisation Methods	18
4. Experimental Design	19
4.1 Model implementation	19
4.1.1 Model Implementation	20
4.1.2 Normalisation Techniques Implementations	20
Layer Normalisation (LN) and Root Mean Square Normalisation (RMSN) Implementation	
20	
Power Normalisation Implementation	21
4.1.3 Computational Resources and Infrastructure	22
4.1.4 Optimizations for Efficient Training	24
4.2 Ablation Studies and Downstream Tasks	25
4.2.1 Ablation Methodology	27
4.2.2 Downstream Tasks Selection and Fine-Tuning	28
4.2.3 Evaluation Metrics and Methods	30
4.3 - Expected Outcomes	33
Section 5 - Challenges Faced and Mitigations	34
5.1 GPT-4 Tokenizer Issue	34
5.2 Power Normalisation Challenges	36
5.3 Billsum Context Window Issue	38
6. Empirical Results and Data Analysis	39
6.1 Training Curves of Pre-trained Models	39
6.1.1 Layer Normalisation (LN)	40
6.1.2 Root Mean Square Normalisation (RMSN)	41
6.1.3 Power Normalisation (PN)	42
6.1.4 Comparative Analysis	44
6.2 Training Curves of Fine-Tuned Ablated Models	45
6.2.1 Fine-Tuning on Summarisation	45
6.2.2 Fine-Tuning on Question Answering (QA)	46
6.2.3 Comparative Analysis	47
6.3 Internal Gradient Analysis	49
6.4 Downstream Task Evaluations	51
Evaluation Metrics Analysis Methodology	51
6.4.1 Text Generation	53
6.4.2 Summarisation (BillSum Dataset)	58
6.4.3 Question Answering (SQuAD Dataset)	63
6.4.4 Overall Observations	67
Overall Model Rankings Based on Observations	67
Comparison of Normalisation Techniques	67
7. Conclusion	68

7.1 Summary of Findings	68
7.2 Methodological Limitations	69
7.3 Future Research Directions	70
8. References	71
9. Acknowledgements	78
10. Appendix	79
Appendix A	79
A.1 Custom Prompts for Text Generation	79
A.2 TextGen Evaluation Prompts	80
A.3 Prompt Engineering for Summarisation	81
A.4 Prompt Engineering for Question Answering	82
Appendix B - Links to Hugging Face Repository	84
B.1 Pre-trained Models' Weights	84
B.2 not fine-tuned Text-Generation Models' Weights	84
B.3 Fine-tuned Summarisation (BillSum) Models' Weights	85
B.4 Fine-tuned Question Answering (SQuAD) Models' Weights	86
Appendix C - Graphing Code	87
Appendix D - pre training code	121
D.1 Layer Normalisation Pre-training	121
D.2 Root Mean Square normalisation pretraining	134
D.3 Power normalisation pretraining	148
D.4 Tokenization of EduFineweb-10B dataset using GPT-2 tokenizer	163
Appendix E - Ablation Code	165
E.1 - LayerNorm Ablation Code	165
E.2 - RMSN Ablation Code	168
Appendix F	171
F.1 Fine Tuning Summarisation (BillSum)	171
F.2 Fine Tuning Question Answering (SQuAD)	176
Appendix G - Collecting CSV Files	180
G.1 TextGen (GPT-4o Metrics)	180
G.2 Summarisation (Traditional Metrics)	182
G.3 Summarisation (GPT-4o Metrics)	186
G.4 Question Answering (Traditional Metrics)	189
G.5 Question Answering (GPT-4o Metrics)	191
Appendix H - GPT4o API	193
H.1 TextGen	193
H.2 Summarisation (Billsum)	198
H.3 Question Answering (SQuAD)	201
Appendix I - Text Generation GPT-4o CSV Files	206
Appendix J - Summarisation GPT-4o CSV Files	216
Appendix K - Question Answering GPT-4o CSV Files	240

1. Introduction

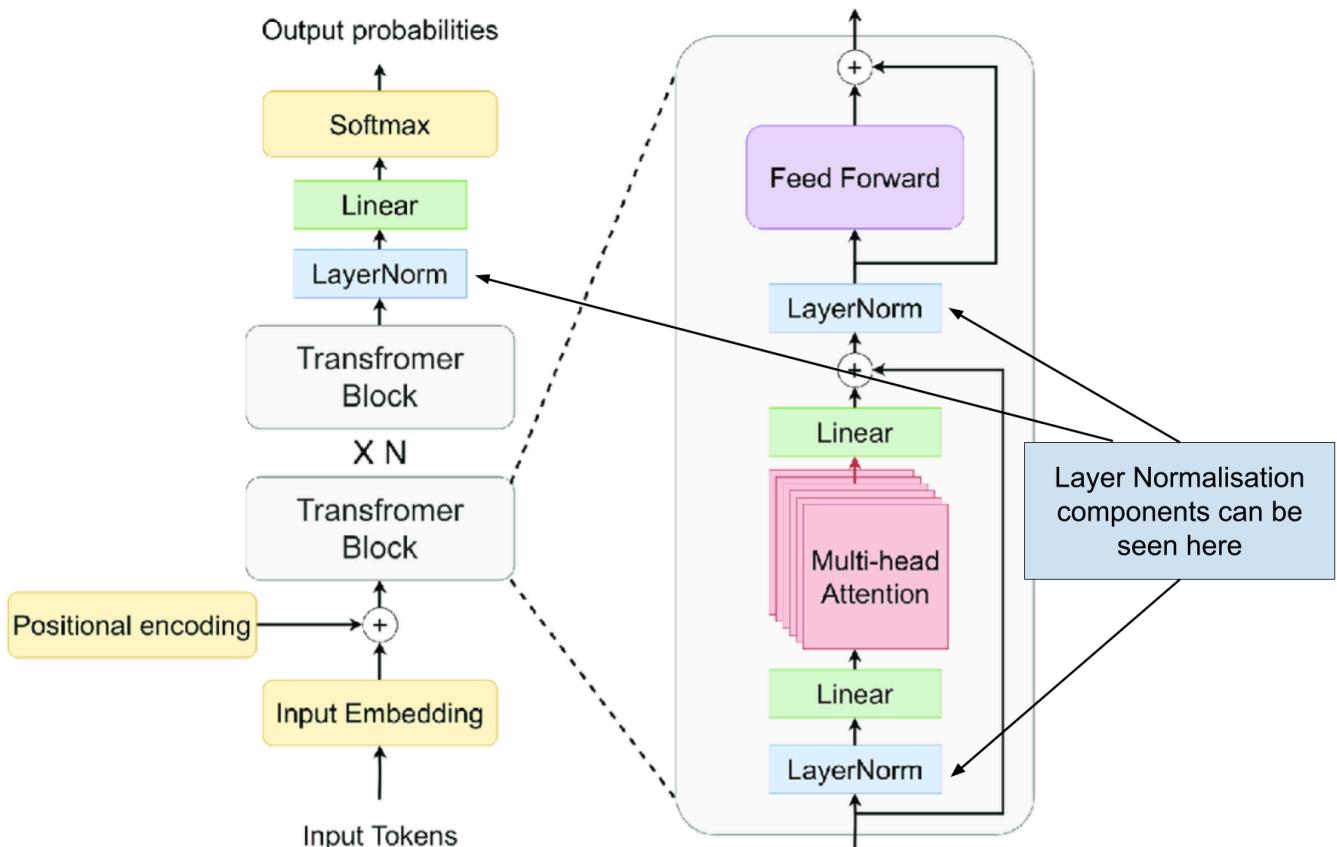


Figure-1: Diagram of Decoder-only Transformer Architecture¹

In 2017, Google's paper "Attention is All You Need" introduced the Transformer architecture (Vaswani et al., 2017), setting a new state-of-the-art (SOTA) in Natural Language Processing. This architecture leveraged self-attention mechanisms to efficiently process sequential data, overcoming limitations inherent in earlier SOTA—Recurrent Neural Networks. Building upon this innovation, OpenAI developed the decoder-only Transformer variant, leading to the creation of GPT-2 in 2019 (Radford et al., 2019).

A crucial component of Transformers is normalisation techniques, particularly Layer Normalisation (Ba et al., 2016) which ensures training stability by normalising input features within each layer, ensuring consistent activation distributions. However, alternative techniques—Root Mean Square Normalisation (Zhang & Sennrich, 2019) and Power Normalisation (Shen et al., 2020)—have been

¹ Diagram from (Bashkar, 2024)

proposed, each offering distinct advantages. This study investigates the necessity of different normalisation techniques within decoder-only Transformer architectures, specifically GPT-2, to determine which configuration optimises training cost and downstream task effectiveness.

We evaluate the necessity of normalisation techniques through training efficiency and performance on downstream tasks such as text generation, summarisation, and question answering. Through data analysis, we aim to identify the most cost-effective configuration—whether that involves retaining all normalisation layers, selectively applying them to specific components like the attention mechanism or feedforward networks, or omitting them entirely. The goal is to propose a configuration that optimises both training cost and downstream performance, thereby improving Transformer's efficiency.

This research is significant because training Transformer models is resource-intensive, consuming substantial amounts of energy and water, contributing to environmental pollution (Korhonen et al., 2019, p. 3645). Identifying the most effective configurations can help researchers design more efficient and powerful AI architectures, potentially reducing both the environmental impact and computational costs associated with training large models. Further, Transformers play a foundational role in the pursuit of Artificial General Intelligence (Bubeck et al., 2023), so optimising their architecture has broad implications for future technological advancements.

Personally, this study is pivotal, since training Transformers from scratch will provide me with practical insights—a valuable skill for my academic aspirations. By exploring the role of normalisation techniques, I aim to contribute to the development of more efficient AI systems by guiding future researchers through my empirical findings.

2. Background Information

Transformer Architecture Overview

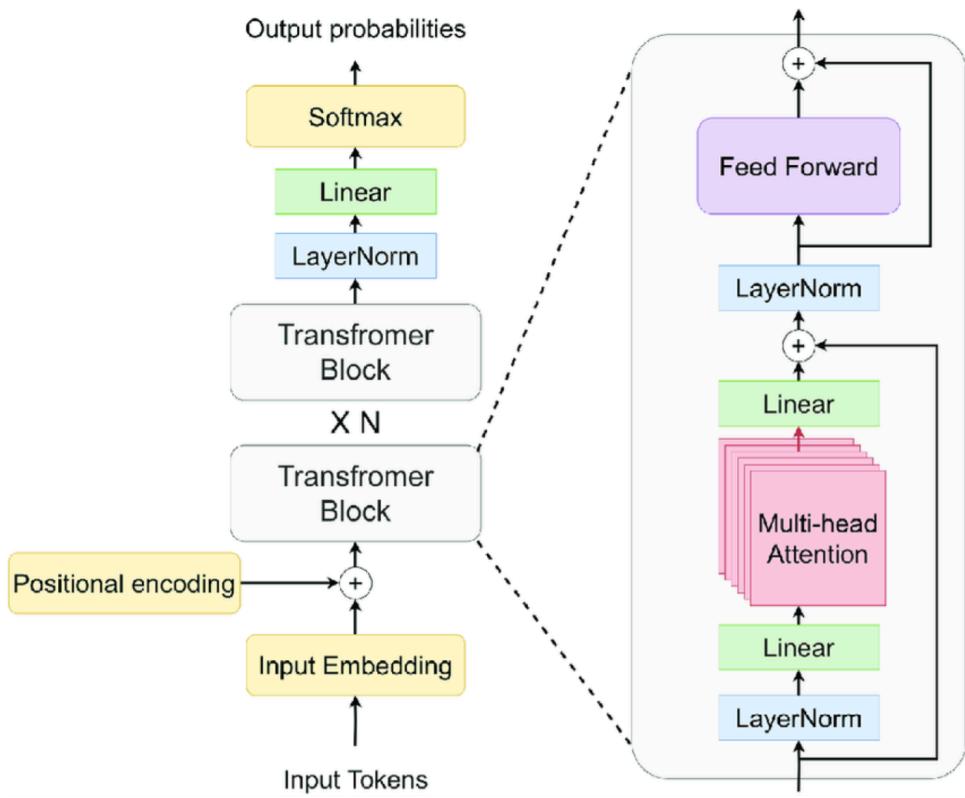


Figure-2: Decoder-only Transformer Architecture²

Below is a simplified walkthrough on how decoder-only transformers function:

[The, brown, fox, jumps] (1)

$$\begin{bmatrix} \text{The} \\ \text{brown} \\ \text{fox} \\ \text{jumps} \end{bmatrix} \xrightarrow[\text{tokenizer}]{} \begin{bmatrix} 101 \\ 2078 \\ 513 \\ 778 \end{bmatrix} \xrightarrow[\text{layer}]{\text{embedding}} \begin{bmatrix} 0.21 & -1.03 & 0.87 & 0.58 & -0.76 \\ -0.45 & 0.92 & -0.23 & 1.12 & 0.09 \\ 0.98 & -0.14 & 0.45 & -0.88 & 0.63 \\ -0.67 & 0.73 & 0.36 & 0.21 & -0.47 \end{bmatrix} \quad (2)$$

1. Since neural networks process numbers, not text. Given input (1), a **tokenizer** converts each word (normally subwords) into numerical IDs, which **Input Embedding** (Vaswani et al., 2017, Section 3.4) transforms into high-dimensional vectors (2).

² Diagram from (Bashkar, 2024)

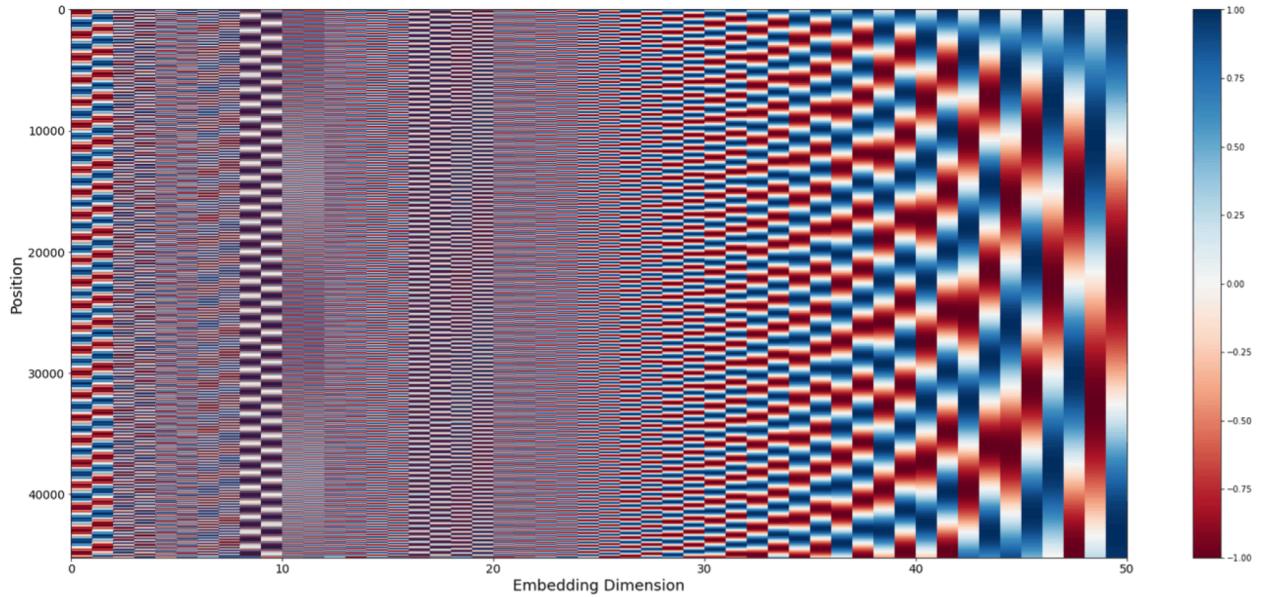


Figure-3: Graphing our Positional Encoding (3), (4)³

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{model}}) \quad (3)$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{model}}) \quad (4)$$

2. Since the order of words matters in English. **Positional encoding** (Vaswani et al., 2017, Section 3.5) uses (3) and (4) to inject positional information.

$$\text{Embedding Summation} = \text{Input Embedding} + \text{Positional Encoding} \quad (5)$$

3. (5) ensures the input contains both semantic and positional information about (1).

³ Image of positional encoding graphed out taken from (Yin & Ungaro, 2019)

4. The input's is passed through **Nx transformer blocks**, each containing:

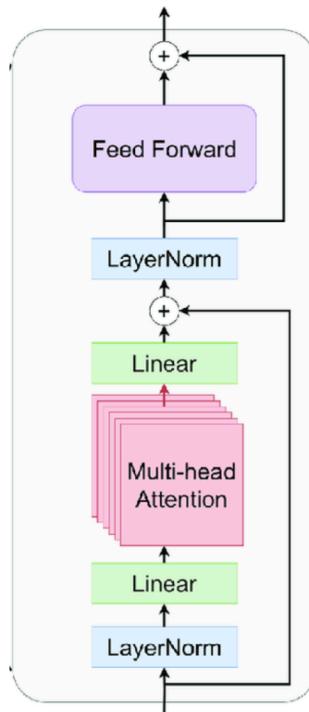


Figure-4: Anatomy of each Transformer block⁴

- **Layer Normalisation:** Normalises input for training stability (Ba et al., 2016).



Figure-5: Visualisation of GPT-2's self-attention using BertViz⁵

$$\text{Attention}(\text{Query}, \text{Key}, \text{Value}) = \text{softmax}\left(\frac{\text{Query} \cdot \text{Key}^{\text{Temperature}}}{\sqrt{\text{dimensionality of key vectors}}}\right) \cdot \text{Value} \quad (6)$$

- **Attention Mechanism⁶:** Calculates each token's importance relative to others in a sequence, enabling the model to focus on relevant context for each word. For example, GPT-2 focuses most on “The” (see Figure-5).

⁴ Diagram taken from (Bashkar, 2024)

⁵ Visualisation created using bertviz (Vig, 2019, pp. 1-5). Source code in Appendix C.

⁶ For more detail on how attention-mechanism actually works. Please refer to (Vaswani et al., 2017, pp. 3-5).

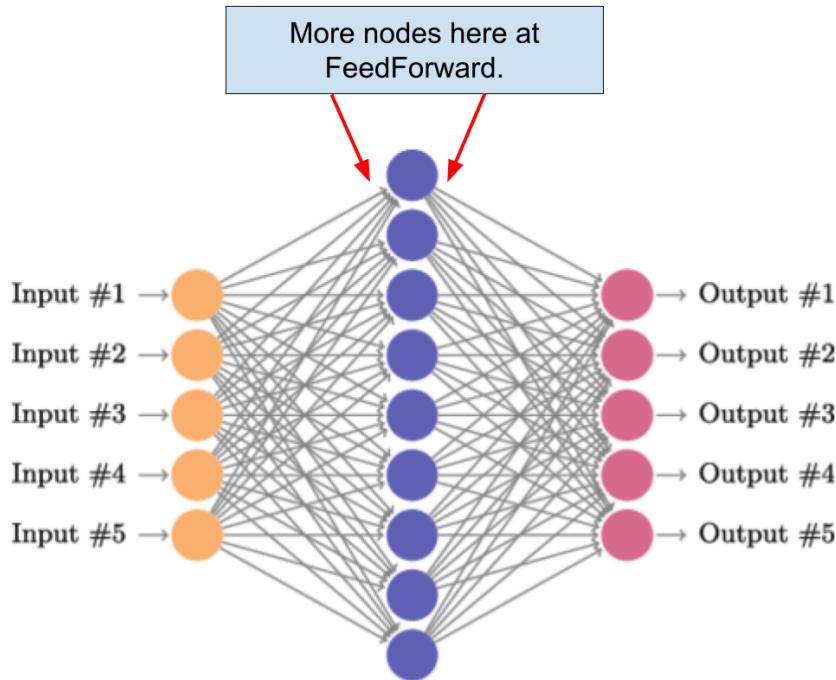


Figure-6: Diagram of FeedForward Layer⁷

- **Feed-Forward Network⁸:** Enables transformers to learn complex relationships using an increased number of nodes (see Figure-6).

5. After processing by \mathbf{Nx} blocks, the input passes through **Linear Layer** to generate scores.
 - **Linear Layer**'s weights are tied to **Input Embeddings**, since they serve similar functions: **Input Embeddings** convert tokens to vectors, whilst **Linear Layer** converts vectors back to tokens. This tying helps reduce training cost, by reducing the number of learnable parameters. (Press & Wolf, 2017)

[The, brown, fox, jumps, over]

(7)

6. Scores are converted to probabilities of the next most probable word using **Softmax**—an activation function (Goodfellow et al., 2016, p. 81). Assuming the predicted word is "jumps", it's then appended to original input (1), creating (7).

⁷ Diagram drawn using Claude-3.5 Sonnet (Anthropic, 2024).

(8)

⁸ For more detail of how FFN works, refer to (Vaswani et al., 2017, Section 3.3).

[*The, brown, fox, jumps, over, the, lazy, dog, <|endoftext|>*]

7. (7) is fed back into the transformer again, repeating steps 1-5 until <|endoftext|> token is predicted, signalling the end of text generation.
8. Finally, creating (8).

(Turn to next page)

3. Theoretical Framework

This section introduces and compares the normalisation methods that we're investigating, highlighting their internal mechanisms and benefits within decoder-only transformers.

3.1 Layer Normalisation (LN)

LN is widely used in Transformers, including OpenAI's GPT series (OpenAI, 2024). LN normalises inputs across the feature dimension by adjusting the mean and variance, stabilising activations within each layer.⁹

The formulation is:

$$y_i = \frac{x_i - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}} \cdot \gamma + \beta$$

where:

- x_i : input vector
- $\mu = \frac{1}{d_k} \sum_{j=1}^{d_k} x_j$: mean of input features¹⁰
- $\sigma^2 = \frac{1}{d_k} \sum_{j=1}^{d_k} (x_j - \mu)^2$: variance
- ϵ : small constant for stability¹¹
- γ : weight
- β : bias

⁹ For more details on Layer Normalisation, refer to (Ba et al., 2016, Section 2-3).

¹⁰ For background information on the difference between batch and feature scaling, please refer to (Maharaj, 2024).

¹¹ To prevent division by zero in cases where the variance is zero, a small constant, epsilon (e.g., 1e-4), is added. This ensures that the variance remains positive, so the training run proceeds without errors due to division by zero.

Internal Covariate Shift

Without normalisation, feature distributions change after each block due to attention-mechanism and feedforward operations. Assuming each block increases the mean by 1.5 and scales variance by 1.2, over 12 blocks:

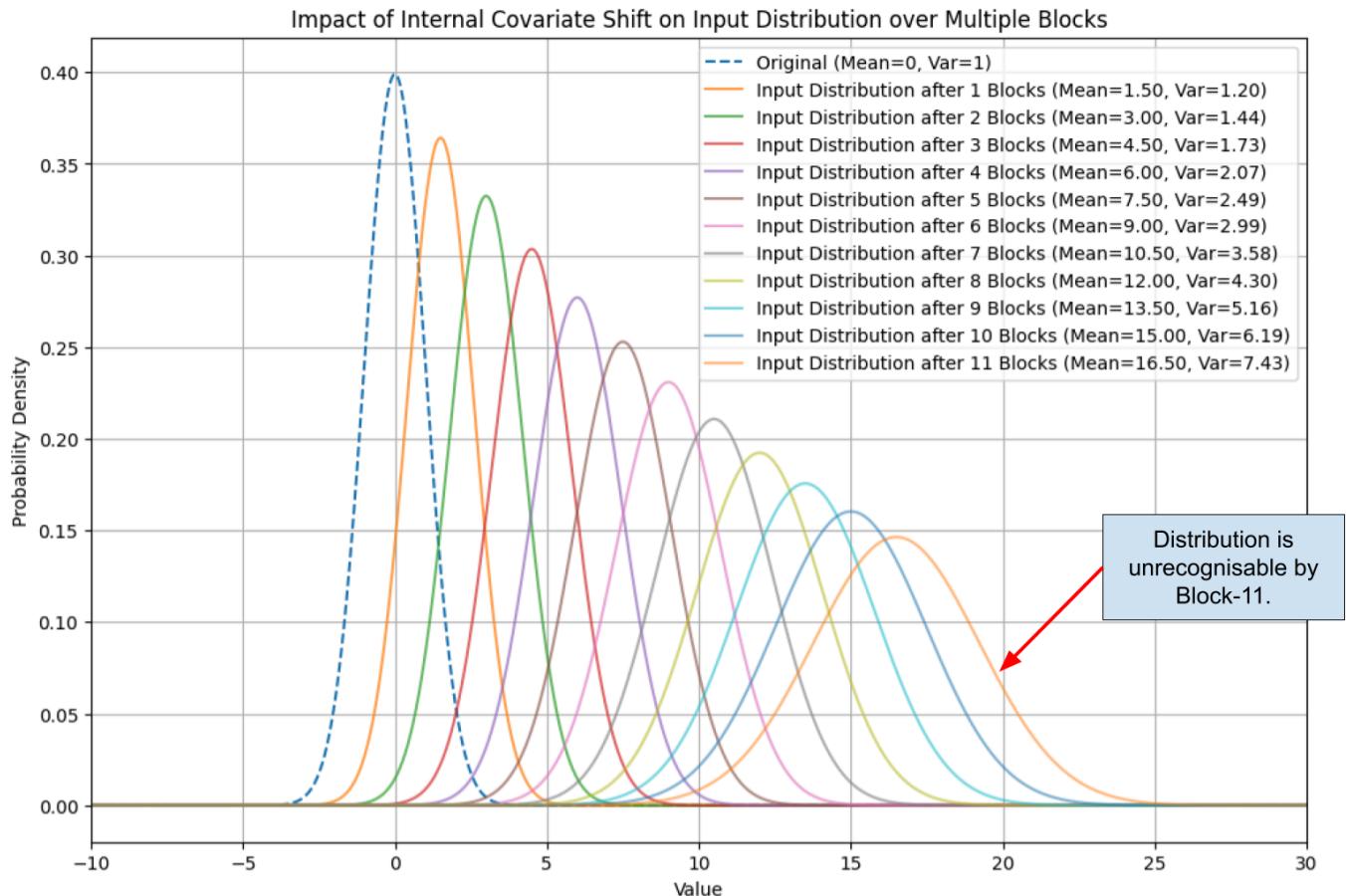


Figure-7: Impact of Internal Covariate Shift over Multiple Blocks Without Normalisation¹²

In Figure-7, we can see internal covariate shift (Ioffe et al., 2015, Section 2) occurring, as distributions become unrecognisable when compared to its original distribution by block-11. This makes learning difficult and reduces training stability.

¹² Graphed using Matplotlib (Hunter, 2007, pp. 90-95). Source code is in Appendix C.

LN's Impact on Input Distributions

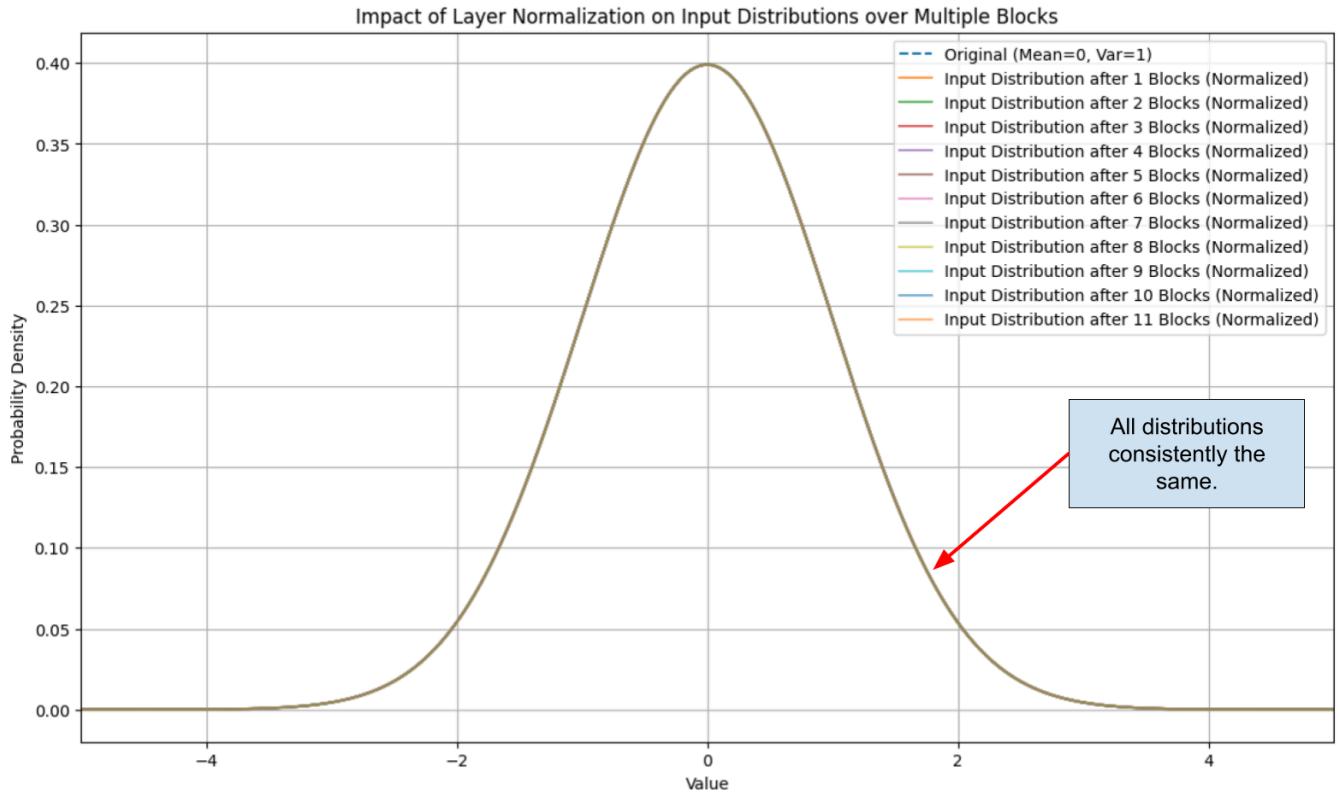


Figure-8: Stable Input Distributions when Layer Normalisation is Applied¹³

LN mitigates internal covariate shift by always centering mean at zero and scaling variance to one, ensuring that input distributions are always consistent (Figure-8).

¹³ Graphed using Matplotlib (Hunter, 2007, pp. 90-95). Code is below in Appendix C.

3.2 Root Mean Square Normalisation (RMSN)

RMSN (Zhang & Sennrich, 2019, pp. 1-9), used by Meta's LLaMA models (Touvron et al., 2023, Section 2.2.), offers a computationally cheaper alternative to LN by scaling only the variance and omitting mean subtraction.¹⁴

The formulation is:

$$y_i = \frac{x_i}{\text{RMS}(x)} \cdot \gamma + \beta$$

where:

- $\text{RMS}(x) = \sqrt{\frac{1}{d_k} \sum_{j=1}^{d_k} x_j^2}$: Root mean square of input features.
- γ : weight
- β : bias

¹⁴ For more details, please refer to the original proposal paper of RMSN from (Zhang & Sennrich, 2019, pp. 1-9), and also RMSN's use in Transformers: (Jiang et al., 2023, pp. 1-10).

RMSN's Impact on Input Distributions

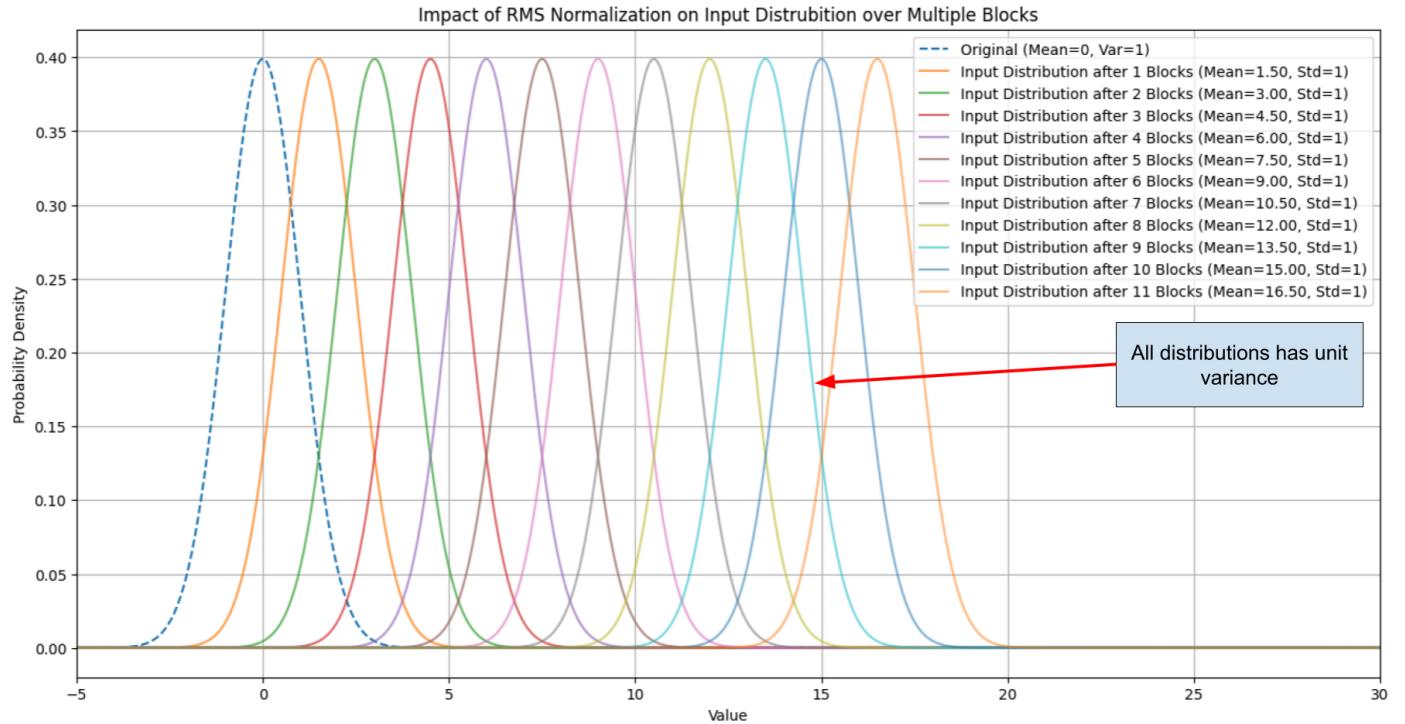


Figure-9: Input Distributions's Mean Drifts when RMSN is Applied^{15 16}

By avoiding mean calculation, RMSN reduces computational overhead, thereby increasing training speed. However, the absence of mean centering can cause gradual mean shifts for input distributions across layers (Figure-9), potentially affecting training stability.

¹⁵ Figure 12 illustrates an assumed scenario where each block shifts by 1.5 and scales by 1.2, primarily to highlight the differences between LN and RMSN for illustrative purposes.

¹⁶ Graphed using Matplotlib (Hunter, 2007, pp. 90-95). Code is below in appendix C.

3.3 Power Normalisation (PN)

PN, introduced by (Shen et al., 2020), normalises across the batch dimension¹⁷ instead of the feature dimension, unlike LN and RMSN.

The formulation is:

Algorithm 2 Power Normalization (Every Iteration)	
begin Forward Propagation:	
Input: $\mathbf{X} \in \mathbb{R}^{B \times d}$	
Output: $\mathbf{Y} \in \mathbb{R}^{B \times d}$	
$\psi_B^2 = \frac{1}{B} \sum_{i=1}^B \mathbf{x}_i^2$	// Get mini-batch statistics
$\widehat{\mathbf{X}} = \frac{\mathbf{X}}{\psi}$	// Normalize
$\mathbf{Y} = \gamma \odot \widehat{\mathbf{X}} + \beta$	// Scale and shift
$\psi^2 = \alpha\psi^2 + (1 - \alpha)\psi_B^2$	// Update running statistics
begin Backward Propagation:	
Input: $\frac{\partial \mathcal{L}}{\partial \mathbf{Y}} \in \mathbb{R}^{B \times d}$	
Output: $\frac{\partial \mathcal{L}}{\partial \mathbf{X}} \in \mathbb{R}^{B \times d}$	
$\frac{\partial \mathcal{L}}{\partial \widehat{\mathbf{X}}} = \gamma \odot \frac{\partial \mathcal{L}}{\partial \mathbf{Y}}$	// Intermediate Gradient
$\widehat{\mathbf{X}}' = \frac{\partial \mathcal{L}}{\partial \widehat{\mathbf{X}}} - \nu \widehat{\mathbf{X}}$	// Intermediate Estimated Gradient
$\frac{\partial \mathcal{L}}{\partial \mathbf{X}} = \frac{\widehat{\mathbf{X}}'}{\psi}$	// Gradient of \mathbf{X}
$\nu = \nu(1 - (1 - \alpha)\Gamma) + (1 - \alpha)\Lambda$	// See Definition 3 for Γ and Λ
Inference: $\mathbf{Y} = \gamma \odot \frac{\mathbf{X}}{\psi} + \beta$	

Figure-10: Formulation of Power Normalisation (Shen et al., 2020, p. 6)

PN's scales the input distribution by the root-mean-square of batch variance (Figure-10). This scaling operation is proposed by (Shen et al., 2020) after performing statistical analysis on natural language (Shen et al., 2020, Section 1).

However, PN uses running statistics, which, combined with its already complex formulation, makes it less computationally efficient, since running statistics require gradient approximation during backward propagation (Shen et al., 2020, Section 4.2).

¹⁷ Refer to Figure-1's diagram from (Shen et al., 2020, p. 2) for more detail.

PN's Impact on Input Distributions

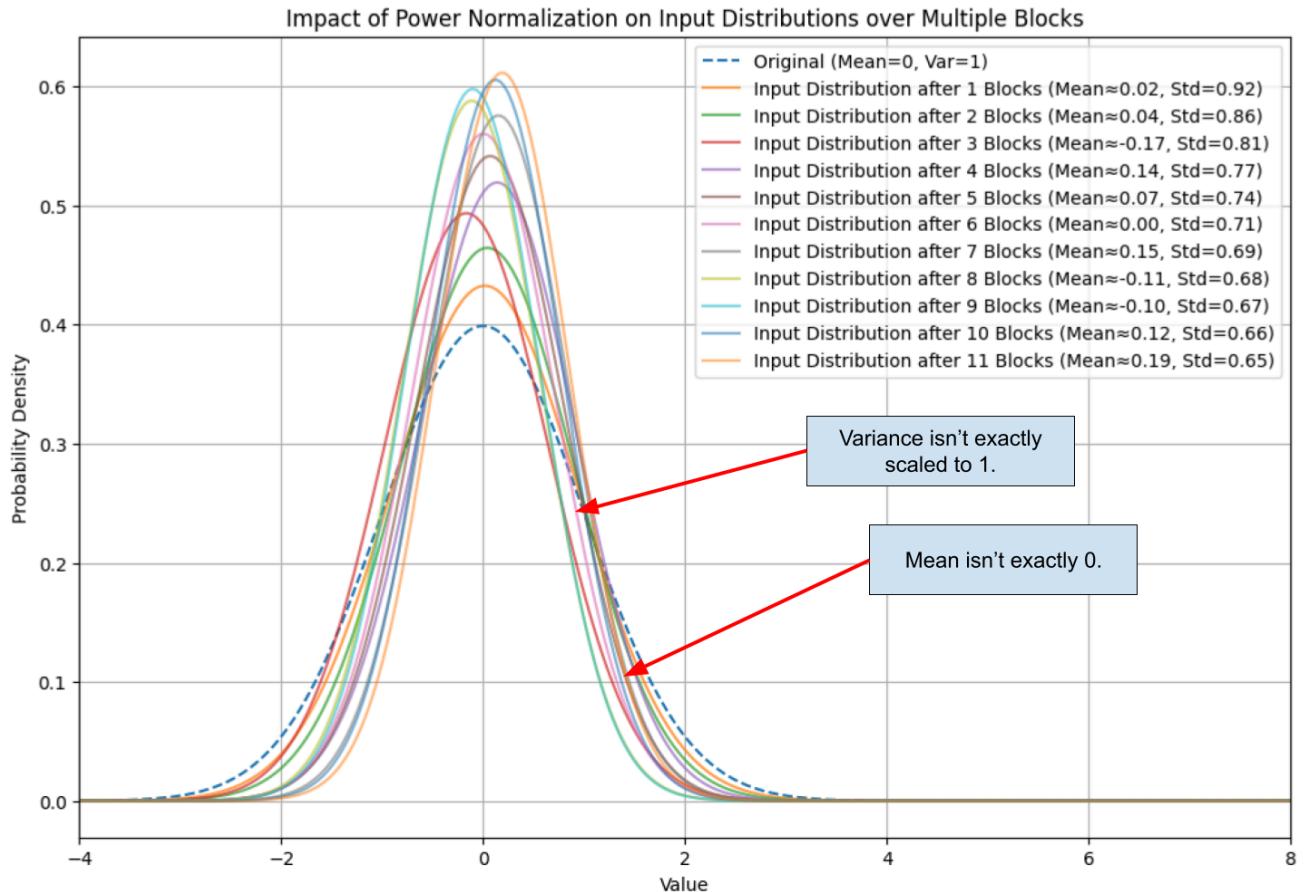


Figure-11: Input Distributions when Power Normalisation is Applied¹⁸

Since PN relies on batch statistics, input distribution's variance isn't exactly scaled to one, nor mean centred exactly to zero, unlike LN. This effect is visualised in Figure-11.

¹⁸ Graphed using Matplotlib (Hunter, 2007, pp. 90-95). Source code is in appendix C.

3.4 Comparison of Normalisation Methods

The key differences between LN, RMSN, and PN are summarised below:

Aspect	LN	RMSN	PN
Normalisation Dimension	Feature	Feature	Batch
Mean ($\mu_{feature}$) Adjustment	Subtracted by $\mu_{feature}$	Not Adjusted	Subtracted by μ_{batch}
Mean ($\mu_{feature}$) Shifting	Centres mean to zero	Not Adjusted	Centres mean approximately to zero
Variance ($(\sigma_{feature})^2$) Adjustment	Divided by $\sqrt{(\sigma_{feature})^2}$	Divided by $\sqrt{(\sigma_{feature})^2}$	Divided by $\sqrt{(\sigma_{batch})^2}$
Variance ($(\sigma_{feature})^2$) Scaling	Scaled to one	Scaled approximately to one	Scaled approximately to one
Computational Complexity	Moderate	Low	High
Training Stability	High	Moderate	Potentially High
Implementation Complexity	Simple	Simple	Complex
Use of Running Statistics	No	No	Yes

Table-1: Overview of the differences of LN, RMSN, PN

4. Experimental Design

4.1 Model implementation

Given our RQ:

"How necessary are different normalisation techniques (LN, RMSN, PN) for decoder-only transformers (GPT-2), and which configuration yields the best performance in terms of training effectiveness and downstream tasks?"

We reproduced OpenAI's GPT-2 (a decoder-only transformer) (OpenAI, 2024) in PyTorch (Meta AI, 2016), based on Andrej Karpathy's nanoGPT (Karpathy, 2022)¹⁹. The model's architectures are modified with different normalisation techniques and pre-trained. By logging during pre-training, we're able to evaluate the training effectiveness and performance of different normalisation configurations.

¹⁹ For a more in-depth understanding of nanoGPT and the reproduction of GPT-2, refer to Andrej Karpathy's GitHub repository and his 'Zero to Hero' series, which served as a valuable guide for this work (Karpathy, 2024).

4.1.1 Model Implementation

Due to financial constraints, we only experimented with GPT-2's smallest variant—124 million parameters (OpenAI, 2024, Section 2.3.). All models were pre-trained on the EduFineweb-10B dataset. (Penedo et al., 2024)²⁰

4.1.2 Normalisation Techniques Implementations

Layer Normalisation (LN) and Root Mean Square Normalisation (RMSN) Implementation

```
def forward(self, x: Tensor) -> Tensor:
    mean = x.mean(dim=-1, keepdim=True)
    var = torch.var(x, dim=-1, keepdim=True, unbiased=False)
    x_norm = (x - mean) / torch.sqrt(var + self.eps)

    if self.elementwise_affine:
        x_norm = self.gain * x_norm + self.bias

    return x_norm
```

CodeBlock-1: Implementation of LN in PyTorch

```
def forward(self, x: Tensor) -> Tensor:
    mean_square = torch.mean(torch.square(x), dim=-1, keepdim=True)
    x_norm = x / torch.sqrt(mean_square + self.eps)

    if self.elementwise_affine:
        x_norm = self.gain * x_norm

    return x_norm
```

CodeBlock-2: Implementation of RMSN in PyTorch

LN and RMSN were straightforward to implement in the [LayerNorm](#) class (see CodeBlock-1 & CodeBlock-2). Their implementations align with the mathematical formulations in Section-3.1 and Section-3.2.²¹

²⁰ FineWeb datasets are very high quality text corpus by (Penedo et al., 2024) and their datasets are freely available to download from F.O.S.S organisation, HuggingFace's Git-LFS repository (Tolo, 2024).

²¹ For full implementation of LN and RMSN, please refer to Appendix D.1 and D.2.

Power Normalisation Implementation

Algorithm 2 Power Normalization (Every Iteration)

begin Forward Propagation:

Input: $\mathbf{X} \in \mathbb{R}^{B \times d}$	Output: $\mathbf{Y} \in \mathbb{R}^{B \times d}$	Similar operations to RMSN's RMS(x)
$\psi_B^2 = \frac{1}{B} \sum_{i=1}^B \mathbf{x}_i^2$ // Get mini-batch statistics		
$\widehat{\mathbf{X}} = \frac{\mathbf{X}}{\psi}$ // Normalize		
$\mathbf{Y} = \gamma \odot \widehat{\mathbf{X}} + \beta$ // Scale and shift		
$\psi^2 = \alpha\psi^2 + (1 - \alpha)\psi_B^2$ // Update running statistics		

begin Backward Propagation:

Input: $\frac{\partial \mathcal{L}}{\partial \mathbf{Y}} \in \mathbb{R}^{B \times d}$	Output: $\frac{\partial \mathcal{L}}{\partial \mathbf{X}} \in \mathbb{R}^{B \times d}$	// Intermediate Gradient
$\frac{\partial \mathcal{L}}{\partial \widehat{\mathbf{X}}} = \gamma \odot \frac{\partial \mathcal{L}}{\partial \mathbf{Y}}$		
$\widehat{\mathbf{X}}' = \frac{\partial \mathcal{L}}{\partial \widehat{\mathbf{X}}} - \nu \widehat{\mathbf{X}}$ // Intermediate Estimated Gradient		
$\frac{\partial \mathcal{L}}{\partial \mathbf{X}} = \frac{\widehat{\mathbf{X}}'}{\psi}$ // Gradient of \mathbf{X}		
$\nu = \nu(1 - (1 - \alpha)\Gamma) + (1 - \alpha)\Lambda$ // See Definition 3 for Γ and Λ		

Inference: $\mathbf{Y} = \gamma \odot \frac{\mathbf{X}}{\psi} + \beta$

Figure-10: Formulation of Power Normalisation (Shen et al., 2020, p. 6)

PN was more challenging due to its use of running statistics and gradient approximation during backpropagation (Shen et al., 2020, Section 4.2). We adapted Shen's original PN implementation from GitHub (Shen et al., 2020) by incorporating Distributed Data Parallel (Li et al., 2020, pp. 2-3) using PyTorch's `torch.distributed` (Meta AI, 2023) library. This modification ensured synchronisation of running statistics across multiple GPUs, maintaining consistency during training.²²

²² For full implementation of PN, please refer to Appendix D.3.

4.1.3 Computational Resources and Infrastructure

NVIDIA-SMI 550.90.07			Driver Version: 550.90.07		CUDA Version: 12.4		
GPU	Name	Persistence-M	Bus-Id	Disp.A	Volatile	Uncorr.	ECC
Fan	Temp	Perf	Pwr:Usage/Cap	Memory-Usage	GPU-Util	Compute M.	MIG M.
0	NVIDIA H100 80GB HBM3	On	00000000:05:00.0	Off		0	
N/A	45C	P0	535W / 700W	71242MiB / 81559MiB	99%	Default	Disabled
1	NVIDIA H100 80GB HBM3	On	00000000:06:00.0	Off		Off	
N/A	51C	P0	539W / 700W	71242MiB / 81559MiB	99%	Default	Disabled
2	NVIDIA H100 80GB HBM3	On	00000000:07:00.0	Off		0	
N/A	45C	P0	515W / 700W	71242MiB / 81559MiB	100%	Default	Disabled
3	NVIDIA H100 80GB HBM3	On	00000000:08:00.0	Off		0	
N/A	38C	P0	493W / 700W	71242MiB / 81559MiB	99%	Default	Disabled

Processes:						
GPU	GI	CI	PID	Type	Process name	GPU Memory Usage
ID	ID	ID				
0	N/A	N/A	2278	C	.../mambaforge/envs/env/bin/python3.11	71232MiB
1	N/A	N/A	2279	C	.../mambaforge/envs/env/bin/python3.11	71232MiB
2	N/A	N/A	2280	C	.../mambaforge/envs/env/bin/python3.11	71232MiB
3	N/A	N/A	2281	C	.../mambaforge/envs/env/bin/python3.11	71232MiB

Figure-12: Nvidia SMI (Nvidia, 2024) showing status of Nvidia H100 node

Since pre-training GPT-2 is computationally intensive, we rented NVIDIA H100s²³ (Nvidia, 2024) from TensorDock and used spot instances to reduce costs by >50% (TensorDock, 2024, Spot Instances).



Figure-13: HuggingFace and Weights & Biases logo

We also integrated:

Tool	Purpose
Weights & Biases	Real-time monitoring of training metrics and GPU status (Weights & Biases, n.d.)
Hugging Face	Open sourced free storage for large AI model files using Git-LFS (GitHub, n.d.) from HuggingFace repositories. (HuggingFace Inc., n.d.)

Table-2: Tools integrated in training pipeline

²³ The fastest GPU available at this time, offering 1,979 teraFLOPS (floating point operations per second) at FP16 precision.

GPU Utilisation

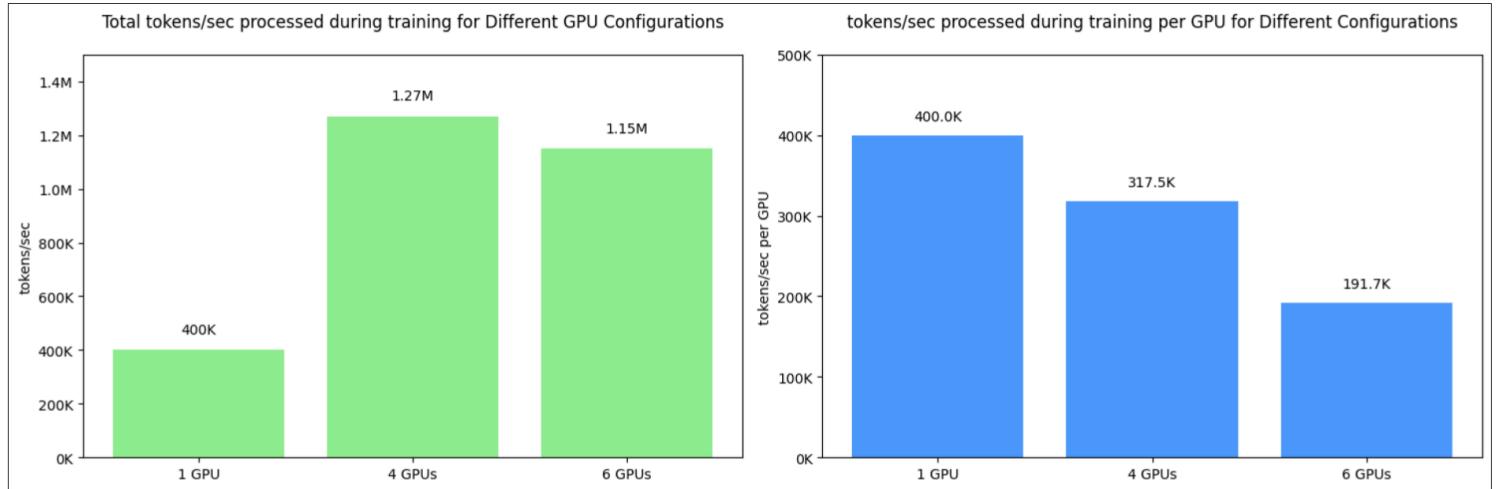


Figure-14: Tokens processing speed when pre-training LN of different H100 configurations²⁴

Empirical testing showed diminishing returns with more GPUs used for token processing speed per GPU (Figure-14). We found that four GPUs offered the optimal balance between speed and synchronisation overhead, achieving 1.27 million tokens per second. Thus, we used four GPUs.²⁵

²⁴ Graphed using Matplotlib (Hunter, 2007, pp. 90-95). Source code is in appendix C.

²⁵ Eight GPUs were not tested because they were frequently unavailable on TensorDock's live marketplace system (TensorDock, 2024).

4.1.4 Optimizations for Efficient Training

To improve pre-training efficiency, we implemented several optimization techniques:

Optimization Technique	Actions Taken	Description	Impact
Batch Size ²⁶	Batch size set to 2^{19} tokens, which uses up 79/81 GB of a H100's HBM.	We want to maximise our H100's HBM usage. This ensures data transfer between CPU and GPU is minimised, reducing the largest bottleneck. (Nvidia, 2024, Section 8.2)	Data transfer bottleneck reduced.
<code>torch.compile</code>	Activated <code>torch.compile()</code> (Meta AI, 2023)	Converts all PyTorch operations to quick, optimised C/CUDA code for speed.	Training speed was tripled.
Mixed Precision Training	Activated TF32 precision (Karparthy, 2024) instead of FP32 ²⁷ on H100. ²⁸	TF32 reduces memory usage and computation time for each float while maintaining high precision. (Narang et al., 2018)	Increased Training Speed.
Power of 2s Configurations ²⁹	Set vocabulary size to 393×2^7 and batch size to 2^{19}	CUDA GPUs are made of SMs which consist of 128 cores. By using many powers of 2s in our training configuration, we minimise the number of idle cores during training, maximising GPU utilisation.	Maximises GPU utilisation, improves performance.
Checkpointing	Saved model weights every 2000 steps to HuggingFace (HuggingFace Inc., n.d.).	In case of server failure, we can continue training from saved weights in checkpoints, instead of restarting from scratch.	Resilient training progress.

Table-3: Overview of Optimization Techniques Employed.

²⁶ Refer to (Goodfellow et al., 2016, Chapter 5) for an explanation if you are unfamiliar with batch size in ML.

²⁷ Traditionally, CUDA GPUs utilise FP32 precision, which demands more memory and computation time.

Transitioning to TF32 reduces both, enhancing efficiency for machine learning tasks where slight precision loss is acceptable.

²⁸ Mixed precision training with TF32 is supported on newer NVIDIA GPUs, such as the Ampere and Hopper series, which feature specialised tensor cores. For older CUDA GPUs, FP16 serves as an alternative, though it entails a greater reduction in precision.

²⁹ See (Nvidia, 2024, Section 11.1.1)

4.2 Ablation Studies and Downstream Tasks

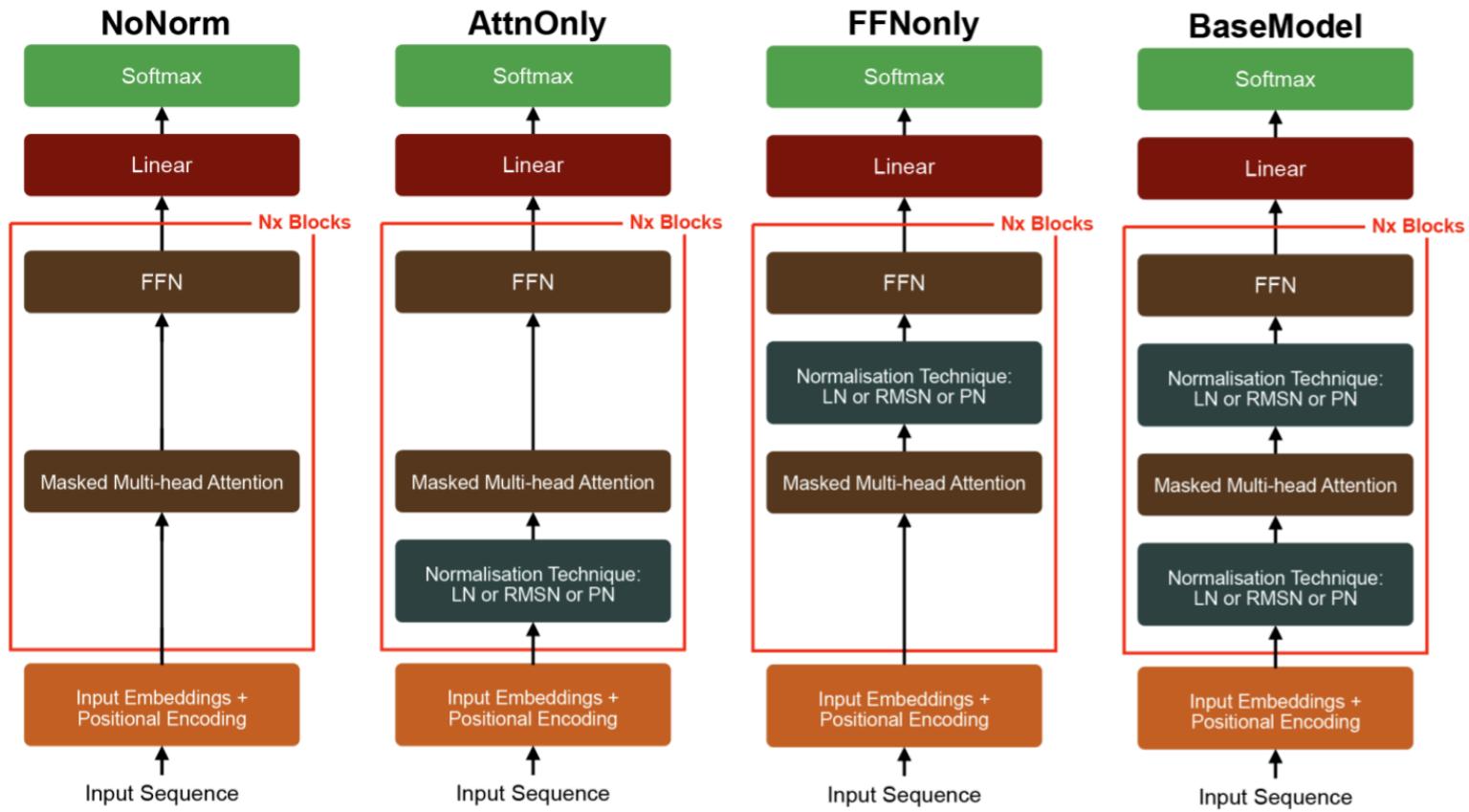


Figure-15: Visualisation of the Architecture of Ablated Model Variants³⁰

Model Variant	Short Form	Description
No Normalisation	noNorm	All normalisation layers removed.
Normalisation for Attention-Only	AttnOnly	Normalisation applied only before the attention mechanism.
Normalisation for Feedforward-Only	FFNOnly	Normalisation applied only before the feedforward network.
Base Model	baseModel	Normalisation applied before both attention and feedforward layers.

Table-4: Description of changes made for ablated model variants

(Turn to next page)

³⁰ Drawn using Claude-3.5 Sonnet (Anthropic, 2024) and Adobe Illustrator (Adobe, n.d.).

To further address our research question:

"How necessary are different normalisation techniques (LN, RMSN, PN) for decoder-only transformers (GPT-2), and which configuration yields the best performance in terms of training effectiveness and downstream tasks?"

We conducted ablation studies on our pre-trained models by removing normalisation layers to create new ablated model variants (see Table-4 & Figure-15).

By comparing how transformers with varying levels of normalisation (through ablation) affect training effectiveness and performance, we directly addressed our RQ. For instance, if **noNorm** configuration is both the cheapest to train and the most performant, it would indicate that normalisation layers are unnecessary, making **noNorm** the best normalisation configuration.

4.2.1 Ablation Methodology

Implementation of Ablation:

```
class IdentityLayerNorm(nn.LayerNorm):
    def __init__(self, normalized_shape, eps=1e-5, elementwise_affine=True):
        super().__init__(normalized_shape, eps, elementwise_affine)

        # Set weights to 1 and bias to 0
        if self.elementwise_affine:
            nn.init.ones_(self.weight)
            nn.init.zeros_(self.bias)

    def forward(self, x):
        return x # Identity function

    def replace_layer_norm(module, name):
        if isinstance(getattr(module, name), nn.LayerNorm):
            setattr(module, name, IdentityLayerNorm(768))
```

CodeBlock-3: Function which substitutes LayerNorm class with Identity Function³¹



Figure-16: Models are saved to HuggingFace repositories once ablated³²

To implement ablation, we coded `replace_layer_norm` (CodeBlock-3) which replaces normalisation layers with identity functions $f(x) = x$, effectively removing them. Once ablated, these models were saved to HuggingFace (Figure-16) for fine-tuning.

³¹ For full implementation of ablation code, please refer to Appendix E.

³² Refer to Appendix B containing links to our HF repositories, which stores the trained model weights of this paper.

4.2.2 Downstream Tasks Selection and Fine-Tuning

Index	Prompt
0	Once upon a time in a distant galaxy,
1	The secret to happiness is
2	In the midst of the bustling city,
3	As the sun set over the horizon,
4	The mysterious package arrived on my doorstep,
5	In a world where robots and humans coexist,
6	The ancient prophecy foretold that

Table-5: 6 out of 25 custom prompts generated by O1-preview (OpenAI, 2024)³³

We chose these three downstream tasks to evaluate model performance because they are typical use cases for decoder-only transformers:

Task	Dataset	Task Description
Text Generation	25 Custom Prompts generated by O1-preview (e.g. Table-5)	Custom prompts were passed into ablated models to generate text, which was then evaluated for quality.
Summarisation	BillSum (Kornilova et al., 2019)	Fine-tuned ablated models were evaluated on their ability to summarise U.S. Congressional bills.
Question Answering (QA)	SQuAD (Rajpurkar et al., 2016)	Fine-tuned ablated Models were evaluated on their ability to answer questions based on Wikipedia texts provided as context.

Table-6: Overview of downstream tasks and their datasets

³³ Refer to Appendix A.1 for all 25 custom prompts generated by O1-preview (OpenAI, 2024).

Fine Tuning Justification

For the Summarisation and QA tasks, we fine-tuned the ablated models on their respective datasets using the HuggingFace API³⁴ (HuggingFace Inc., n.d.). Due to computational constraints, we did not fine-tune for Text Generation, as this would have required re-training the models on the EduFineweb-10B dataset (Penedo et al., 2024).³⁵

We employed Weights & Biases (Weights & Biases, n.d.) to log fine-tuning statistics, for analysis of training efficiency of the various ablated configurations post-training.

³⁴ Refer to Appendix F for full implementation of fine-tuning code.

³⁵ We initially considered pre-training each model for 5,000 steps. However, to achieve full coverage, each model would require 19,073 steps from the start, which was too costly (25 USD per pre-trained model) and time-intensive.

4.2.3 Evaluation Metrics and Methods

We evaluated our ablated models using both traditional metrics and GPT-4o (our gold-standard evaluator).

Traditional Metrics:

Task	Metrics Used	What does this metric do
Summarisation	ROUGE-1, ROUGE-2, ROUGE-L	Measure overlap of unigrams, bigrams, and the longest common subsequence between generated and reference summaries. (Lin, 2004)
	BLEU	Evaluates n-gram precision between generated text and reference text. (Papineni et al., 2002)
Question Answering	Exact Match (EM)	Percentage of predictions that exactly match ground truth answers. (Cloudera Fast Forward Labs, 2020)
	F1 Score	Harmonic mean of precision and recall based on token overlap between predictions and ground truth. (Cloudera Fast Forward Labs, 2020)

Table-7: Overview of traditional metrics used for Summarisation and QA

"A swift auburn fox leaps above a sleepy canine" (9)

"The quick brown fox jumps over the lazy dog" (10)

Traditional metrics often struggle to fully capture the semantic quality of generated text. For example, BLEU might penalise (9) compared to (10) simply due to differences in word choice despite their identical meaning.

To overcome these limitations, we employed GPT-4o (OpenAI, 2024) via the OpenAI API (OpenAI, n.d.) to evaluate our ablated models' responses on all 3 downstream tasks.

GPT-4o Evaluation

Evaluation Criteria

Task	Evaluation Criteria
Text Generation	Creativity, Coherence, Fluency, Relevance, Engagement, Overall Score
Summarisation	Relevance, Conciseness, Fluency, Accuracy, Coherence, Overall Score
Question Answering	Correctness, Completeness, Relevance, Fluency, Conciseness, Overall Score

Table-8: GPT-4o Evaluation Criteria for Each Task

You are an expert AI language model evaluator. Please evaluate the following summary generated by a model based on the criteria below:

Generated Summary:

[Model's summary output]

Reference Summary:

[Reference summary]

Evaluation Criteria:

- Relevance (1-5): Does the summary capture the main points of the reference?
- Conciseness (1-5): Is the summary brief yet comprehensive?
- Fluency (1-5): Is the language natural and free of errors?
- Accuracy (1-5): Are the facts presented correctly?
- Coherence (1-5): Is the summary logically structured?

For each criterion, provide a score and a one-sentence justification. Finally, give an overall score and a brief overall comment.

CodeBlock-4: Example of Prompt passed to GPT-4o for Evaluating Summarisation³⁶

³⁶ Refer to Appendix A.2 - A.4 for the actual prompts passed into GPT-4o through OpenAI API (OpenAI, n.d.).

Implementation Details

GPT-4o evaluated 25 sample responses for each model variant and each task. For each response, the model's generated text, reference text, and original question were provided to GPT-4o, which rated each response on a 1-5 scale for each criterion (Table-8), and provided justification for each score and an overall feedback (as exemplified in CodeBlock-4).³⁷

Advantages of Using GPT-4o

Unlike traditional metrics, GPT-4o offers a deeper understanding of semantic meaning, capturing qualitative aspects such as creativity, coherence, and contextual relevance that numerical metrics often miss. This makes GPT-4o an effective alternative to human evaluators, who are typically considered the "gold standard" for data labelling in machine learning.³⁸

Given there's 600 prompts and >100,000 tokens to evaluate, relying on human evaluators would have been inefficient and inconsistent due to subjectivity. Hence, GPT-4o was used given it provided consistent evaluations across all responses, whilst also significantly reducing the time needed.

³⁷ After fine-tuning, the models generate responses saved in CSV files (refer to Appendix G). These responses are subsequently evaluated using GPT-4o (refer to Appendix H) and saved in CSV files (refer to Appendix I-K).

³⁸ For example, in (Çallı et al., 2021), professional human radiologists are needed to annotate X-ray of human chests. Their human-annotations are deemed as the "gold standard".

4.3 - Expected Outcomes

Given Section-3 and our conjectures regarding the ablation studies, we outline the following expected outcomes:

Aspect	Expected Outcome (Configuration Rankings)	Rationale
Training Efficiency	1. RMSN 2. LN 3. PN	RMSN simplifies calculations by omitting mean adjustment, reducing computational overhead.
Performance on Downstream Tasks	1. PN 2. RMSN and LN	Empirical research by Sheng Shen indicates that PN outperforms other normalisation methods in terms of effectiveness on downstream tasks.
Training Cost (Ablated Models)	1. noNorm 2. AttnOnly & FFNonly 3. baseModel	Reducing normalisation layers decreases computation needed for training, making less normalised models more cost-effective in training.
Downstream Task Performance (Ablated Models)	1. baseModel 2. AttnOnly & FFNonly 3. noNorm	Existing accepted theories on internal covariate shifts (Ioffe et al., 2015, Section 2) suggest that more extensive normalisation stabilises activations, enhancing performance. However, given the small size of our model (124M parameters), AttnOnly and FFNonly variants might perform comparably to the baseModel.

Table-9: Our Expected Outcomes based on Section-3

Section 5 - Challenges Faced and Mitigations

Throughout our experiments, several significant challenges emerged:

5.1 GPT-4 Tokenizer Issue

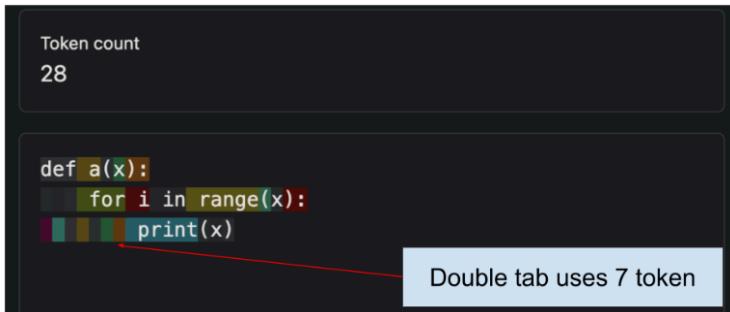


Figure-17: GPT-2 tokenizer inefficient
tokenizing process

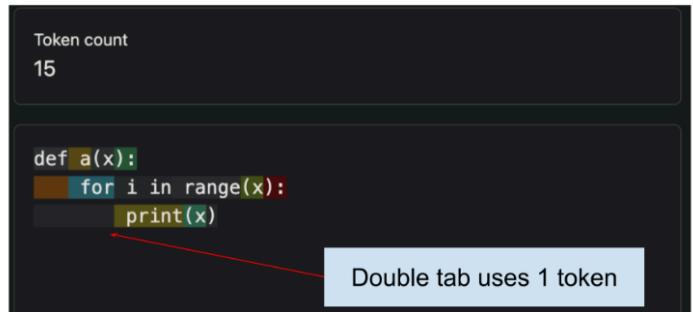


Figure-18: GPT-4 tokenizer efficient
Tokenizing process³⁹

Initially, we aimed to enhance model performance by utilising the GPT-4 tokenizer in our model, which employs advanced tokenization (Karparthy, 2024) compared to GPT-2. For example, GPT-4 treats double tab characters as a single token, while GPT-2 splits it into 7 space tokens (see Figure-17 & Figure-18).

(Turn to next page)

³⁹ Figure-17 and Figure-18 are plotted using Tiktokerizer by (OpenAI, 2022). Accessed with Tiktokerizer website (OpenAI, n.d.).

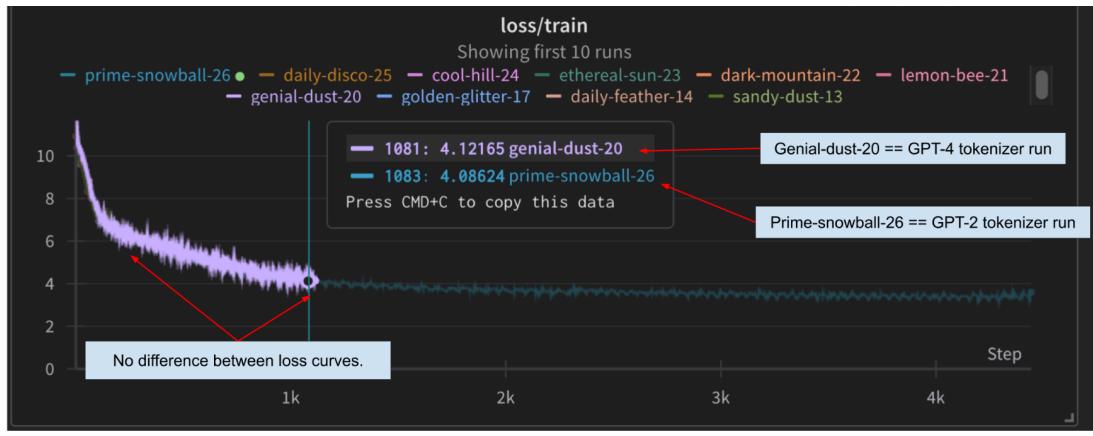


Figure-19: GPT-4 tokenizer training curve vs. GPT-2 tokenizer training curve

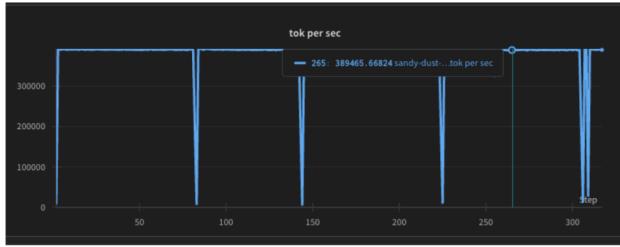


Figure-20: 389.4k tok/sec with GPT-2 tokenizer

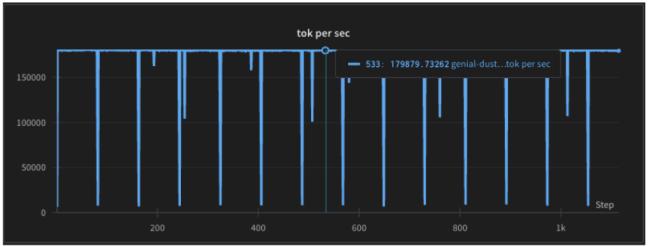


Figure-21: 179.9k tok/sec with GPT-4 tokenizer

However, GPT-4 tokenizer's vocabulary size ($>100,000$ tokens) exceeds the 65,536 limit for 16-bit integers⁴⁰, requiring 32-bit precision for tokenizing EduFineweb-10B⁴¹, which doubled memory consumption and halved training speed—from 390k tokens/sec to 180k tokens/sec (Figure-20 & Figure-21). Yet, the GPT-4 tokenizer loss curve showed no improvement (Figure-19).

Further, Figure-17 and Figure-18's example would mainly be an improvement if we're training our models for programming, since our models will be able to capture the relationships behind indentations better with “tab” tokens than 4 separate spaces—which also reduces the number of tokens used in the context window. However, since our downstream tasks are only text-based, we reverted to using GPT-2 tokenizer.

⁴⁰ On a related note, the GPT-2 tokenizer has 50,257 unique representations, which doesn't exceed the 65,536 threshold of 16-bit integers (Radford et al., 2019, Section 2.2).

⁴¹ See Appendix D.4 for code which uses GPT-2 tokenizer to tokenize EduFineweb-10B before pre-training begins.

5.2 Power Normalisation Challenges

```
class GroupScaling1D(nn.Module):
    def forward(self, input):
        # calculate second moment
        # different group use different mean
        T, B, C = input.shape[0], input.shape[1], input.shape[2]
        Cg = C // self.group_num
        gn_input = input.contiguous().reshape(T, B, self.group_num, Cg)
        moment2 = torch.repeat_interleave(torch.mean(gn_input * gn_input, dim=3, keepdim=True), \
            repeats=Cg, dim=-1).contiguous().reshape(T, B, C)
        # divide out second moment
        return input / torch.sqrt(moment2 + self.eps)
```

CodeBlock-5: Original Sincerass' Implementation of GroupScaling1D

```
class GroupScaling1D(nn.Module):
    def forward(self, input):
        B, T, C = input.shape
        if self.group_num == 1:
            moment2 = torch.mean(torch.square(input), dim=-1, keepdim=True)
        else:
            Cg = C // self.group_num
            input = input.view(B, T, self.group_num, Cg)
            moment2 = torch.mean(input * input, dim=-1, keepdim=True)
            moment2 = moment2.repeat(1, 1, 1, Cg).view(B, T, C)
        return input / torch.sqrt(moment2 + self.eps)
```

CodeBlock-6: My modified Implementation of GroupScaling1D

During pre-training, PN faced gradient explosions which resulted in training failure (Section-6.1.3). We identified an issue: **GroupScaling1D** present in Shen's original code (Shen et al., 2020) wasn't implemented in our code. However, Shen's original implementation (CodeBlock-5) was intended for Vision-Transformers, not NLP Transformers (our use case). Hence, we modified and re-implemented **GroupScaling1D**. (CodeBlock-6)

(Turn to next page)

Impacts

Incorporating `GroupScaling1D` helped reduce gradient explosions, but didn't eliminate them completely, resulting in training failure again. Further, `GroupScaling1D`'s operations (CodeBlock-6) are conceptually similar to RMSN. Adding `GroupScaling1D`'s on top of PN's already complex formulation (Figure-10) further increased computational overhead, reducing training speed.

Note: `GroupScaling1D` isn't present in (Shen et al., 2020, Section 4.2) where PN is formulated.⁴²

(Turn to next page)

⁴² I initially chose not to implement this `GroupScaling1D`, despite noticing it in Shen's code, because its operation resembled group normalisation when used for ViT models (which is computationally expensive). After thoroughly reviewing Shen's PN paper multiple times, we found no mention of `GroupScaling1D` in PN's formulation. It's possible that Shen included this component to address stability issues similar to those we encountered. We attempted to reach out to Shen for clarification, but unfortunately, we received no response.

5.3 Billsum Context Window Issue

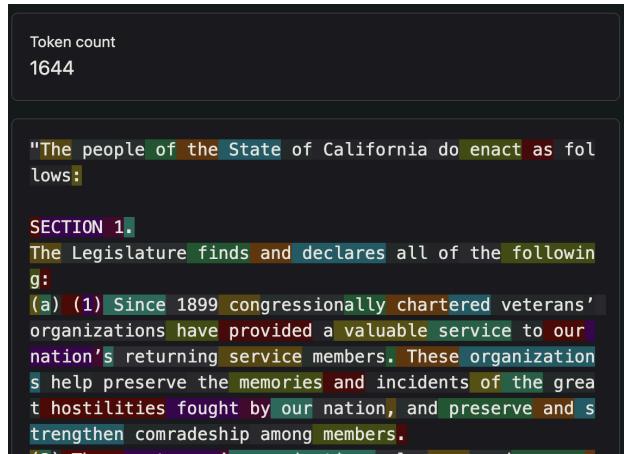


Figure-22: Original Input text from
Billsum Dataset

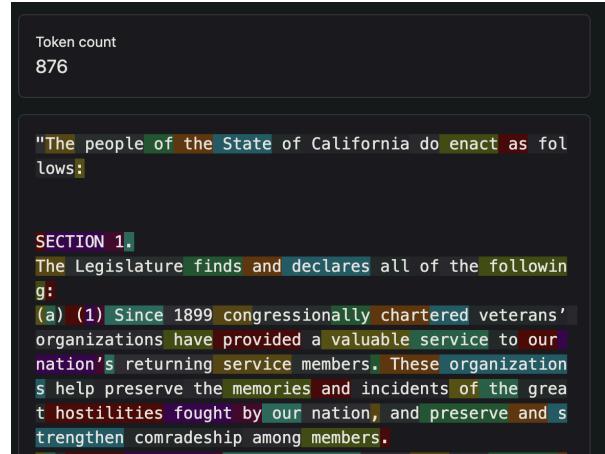


Figure-23: Truncated Input text from
Billsum Dataset⁴³

GPT-2's 1024-token context window⁴⁴ posed significant challenges since BillSum dataset's input and reference summary often exceeded 2000 tokens (see Figure-22). To address this, we truncated the inputs to 876 tokens and reference summaries to 148 tokens during fine-tuning (see Figure-23). This truncation likely caused the low accuracy scores in Section-6.4.2, as truncating the original texts worsened its quality.

Future Mitigation

In the future, we recommend using larger models or increasing the context size to avoid the need for text truncation, allowing the model to learn more effectively. However, this would increase training costs due to the quadratic scaling (Vaswani et al., 2017, p. 6) of attention-mechanisms.

⁴³ Figure-22 and Figure-23 are plotted using Tiktokerizer by (OpenAI, 2022). Accessed with Tiktokerizer website (OpenAI, n.d.).

⁴⁴ This means that GPT-2 is restricted to reading and generating a maximum of 1,024 tokens and cannot process any tokens beyond this limit.

6. Empirical Results and Data Analysis

This section presents the empirical results and analyses the training effectiveness and performance of different normalisation configurations (LN, RMSN, PN and their ablated variants).

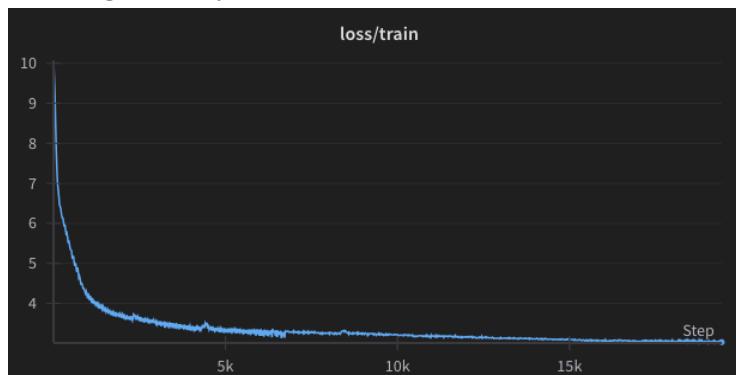
6.1 Training Curves of Pre-trained Models

Here, we analyse the training statistics logged by Weights & Biases during pre-training. LN serves as the baseline in our comparative analysis since it's the standard normalisation technique used in decoder-only transformers.

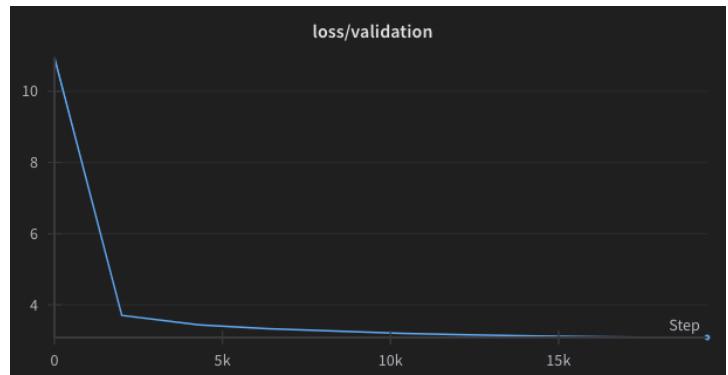
(Turn to next page)

6.1.1 Layer Normalisation (LN)

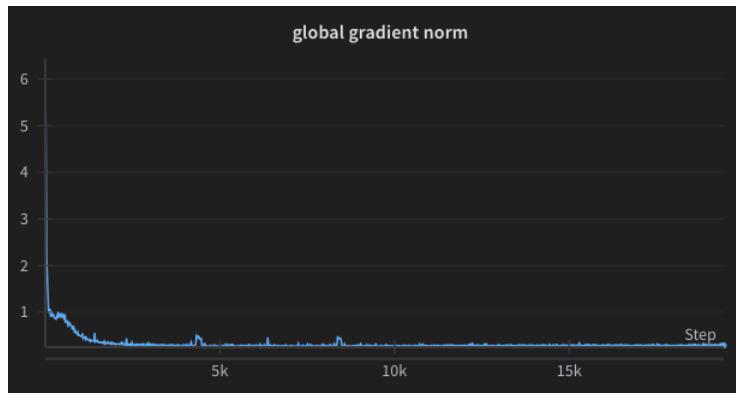
Training Stability



Graph-1: LN's Training Loss Curve



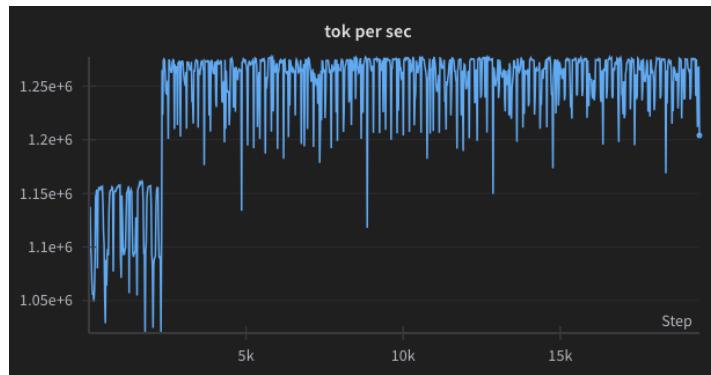
Graph-2: LN's Validation Loss Curve



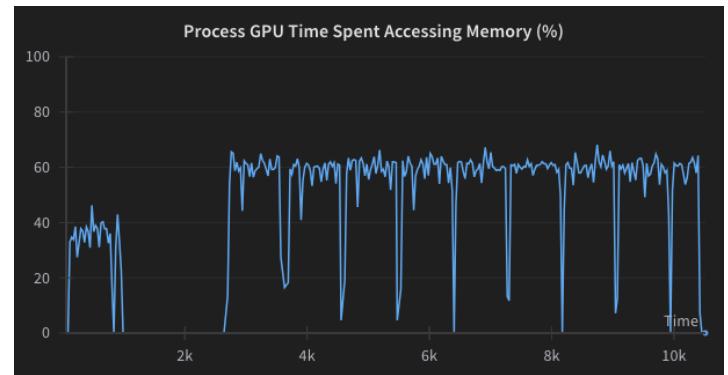
Graph-3: LN's Global Gradient Norm Curve

LN showed training stability, with training loss converging around 3.4 (Graph-1) and validation loss mirroring this, indicating no overfitting. The global gradient norm remained low (<0.5), indicating stable training dynamics without gradient explosion issues.

Computational Efficiency



Graph-4: LN's Token Processing Speed⁴⁵



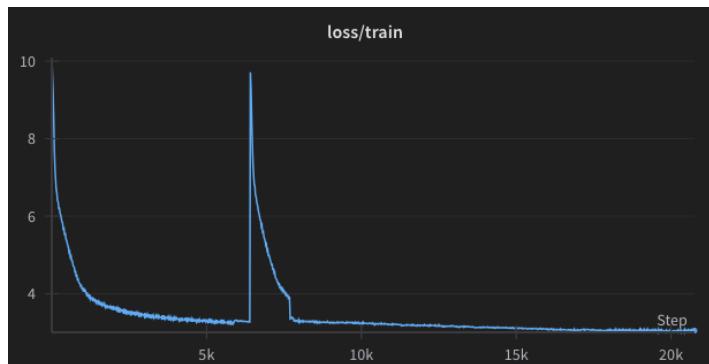
Graph-5: LN's GPU Time Spent Accessing HBM

The token processing speed was 1.225 million tokens per second, with the GPU spending 58% of time accessing high-bandwidth memory (HBM), serving as our baseline for computational efficiency. Alternatively, the GPU only spent 58% of time accessing HBM, which we will consider as the baseline.

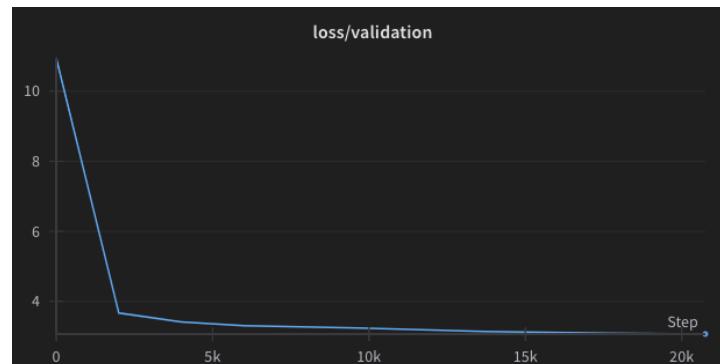
⁴⁵ For Graph-4, please disregard the lower token processing speed during the first 2,000 steps; this was due to experiments with different training configurations.

6.1.2 Root Mean Square Normalisation (RMSN)

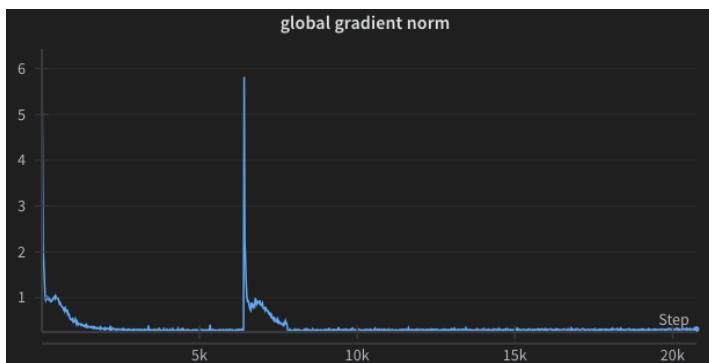
Training Stability



Graph-6: RMSN's Training Loss Curve



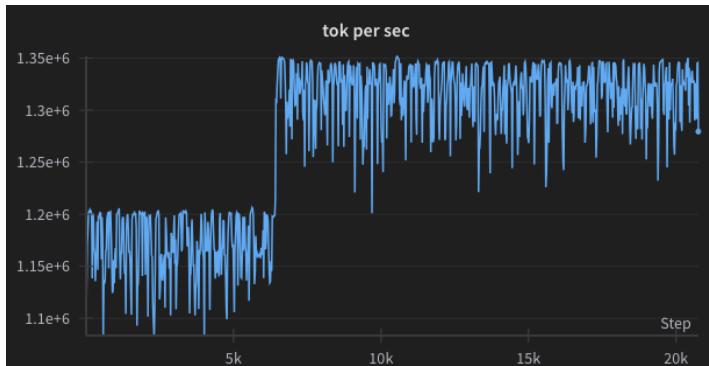
Graph-7: RMSN's Validation Loss Curve



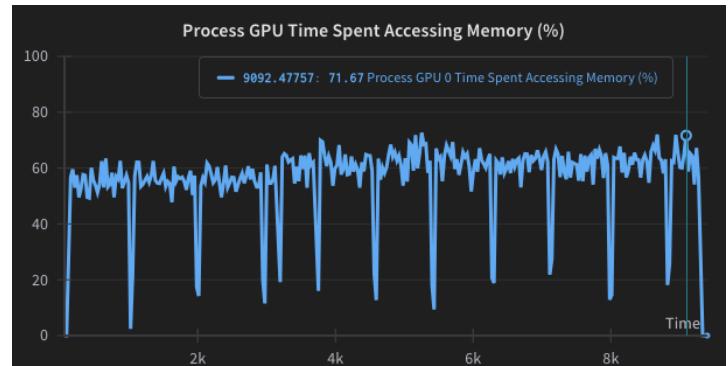
Graph-8: RMSN's Global Gradient Norm Curve⁴⁶

RMSN exhibited similar training stability to LN, with both loss curves converging around 3.4 (no overfitting again) and consistently low gradient norms.

Computational Efficiency



Graph-9: LN's Token Processing Speed⁴⁷



Graph-10: LN's GPU Time Spent Accessing HBM

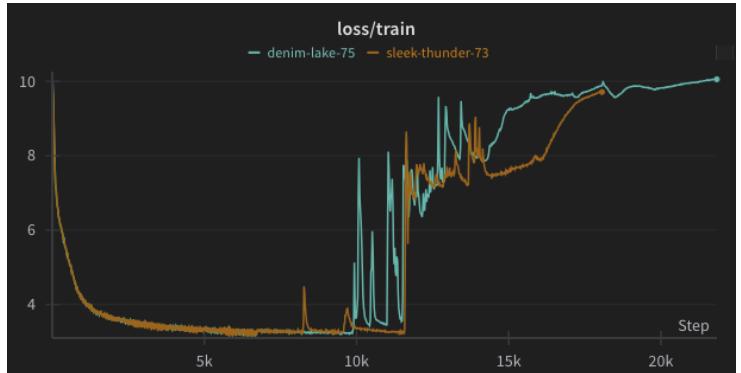
RMSN achieved a token processing speed of 1.3 million tokens per second (6% increase over LN) and had slightly lower memory access time at 57%, indicating improved computational efficiency over LN.

⁴⁶ Please disregard the spike in loss/train and global gradient norm around 6,000 steps. This was due to an accidental restart of the training run following an H100 crash; we resumed promptly from the saved checkpoint.

⁴⁷ Similarly, for Graph-9, please disregard the lower token processing speed during the first 5,500 steps; this was due to experiments with different training configurations.

6.1.3 Power Normalisation (PN)

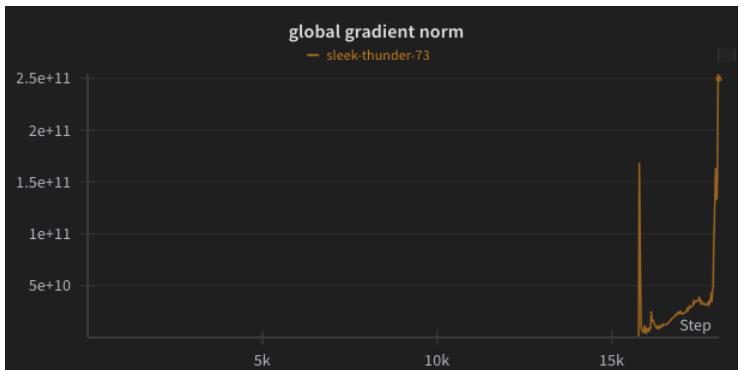
Training Stability



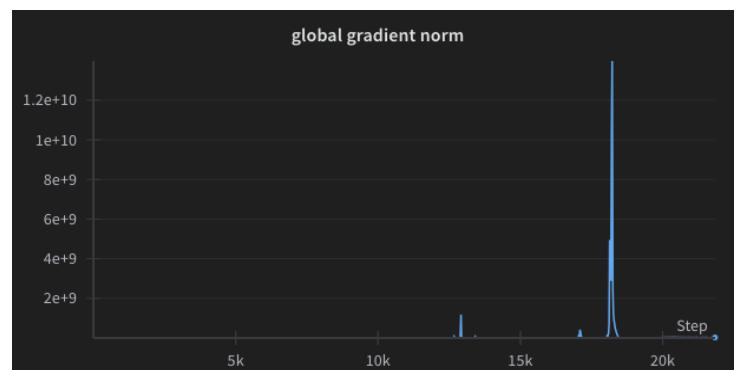
Graph-11: PN's Training Loss Curve



Graph-12: PN's Validation Loss Curve



Graph-13: sleek-thunder-73's global gradient



Graph-14: denim-lake-75's global gradient

Both of PN's training runs had failed despite the implementation of [GroupScaling1D](#) (Section-5.2). Up to step-9600, both runs' loss curves were similar to LN and RMSN, converging around 3.4.⁴⁸

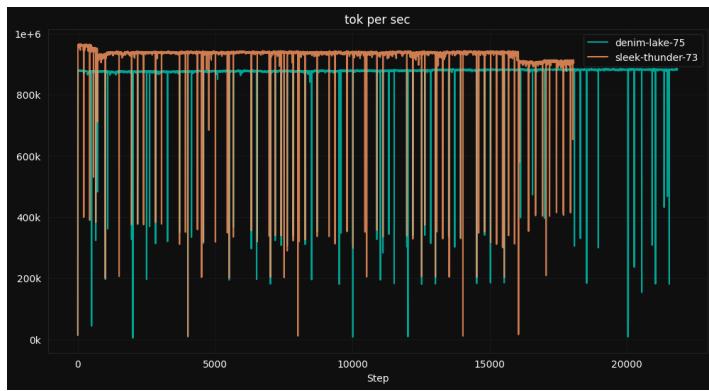
Whilst the implementation of [GroupScaling1D](#) helped to reduce the magnitude of the gradient explosion, from 2.5×10^{11} (sleek-thunder-73) to 1.4×10^{10} (denim-lake-75). The gradient is still significantly high, considering that LN's gradient, a successful run, remained consistently <0.5 . Ultimately, the training still failed, with the loss deteriorating to 9.7—akin to the initial loss when the model was untrained (Graph-11).⁴⁹

Note: Sleek-thunder-73 represents PN before implementing GroupScaling1D, whereas denim-lake-75 is the post-implementation run.

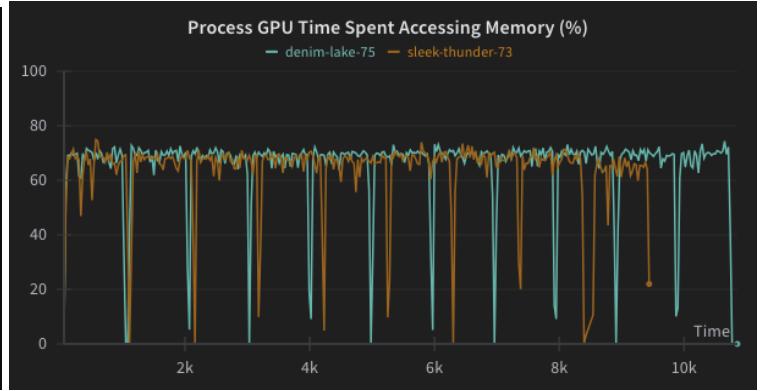
⁴⁸ Out of curiosity, I compared the performance of PN at step 8,000 to LN and RMSN at step 19,073. The performance was not comparable—the text generated by PN was significantly worse, despite similar loss values. This underscores that training loss alone is not a sufficient metric, highlighting the importance of downstream task evaluation.

⁴⁹ It is also possible that the training run failed due to an error in our own implementation. Alternatively, gradient clipping or hyperparameter tuning of learning rate, for example, might potentially resolve this.

Computational Efficiency



Graph-15: PN's token processing speed



Graph-16: PN's GPU time spent accessing HBM

PN incurred significantly higher memory access overhead, spending 76% of training time accessing HBM, compared to LN's 58%.

PN's token processing speed was very low, at 9.5×10^5 (sleek-thunder-73) and 8.6×10^5 (denim-lake-75). As mentioned in Section-5.2, adding [GroupScaling1D](#) further reduces training speed due to its computational overhead. Overall, PN is 29.8% slower than LN, demonstrating markedly inferior computational efficiency.

6.1.4 Comparative Analysis

Metric	LN	RMSN	PN
Final loss/train values	3.4 (Converged)	3.4 (Converged)	9.7 (Failed)
Gradient Norm Stability	Stable (<0.5)	Stable (<0.5)	Exploded (>1e10)
Token Processing Speed (tokens/sec)	1.225M (baseline)	1.3M (+6%)	0.95M (-29%) ⁵⁰
Memory Access Time	58%	56%	76%
Training Success	Successful	Successful	Failed

Table-10: Summary of Pre-Training Empirical Results

Overall Observations

RMSN offers 6% training speed increase over LN with similar training stability—concurring with empirical results of (Zhang & Sennrich, 2019). Thus, for pre-training, RMSN is the best, since it offers superior computational efficiency without compromising training stability.

Alternatively, PN is excluded from further evaluation, since its training failed due to gradient explosion.

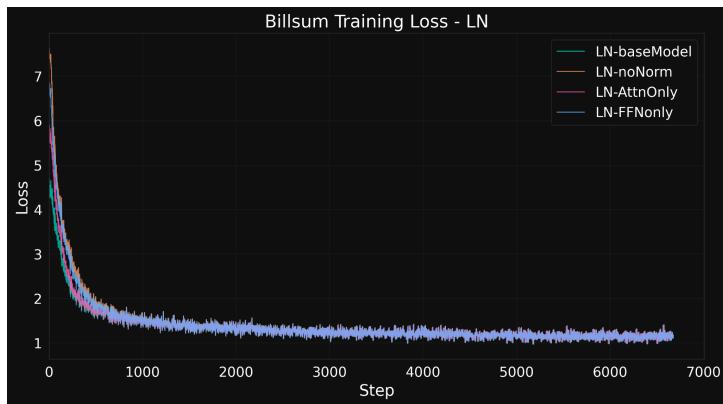
⁵⁰ PN is significantly slower than LN and RMSN. Thus, even if PN were to converge successfully, its considerable slowness would likely make it a suboptimal choice for training.

6.2 Training Curves of Fine-Tuned Ablated Models

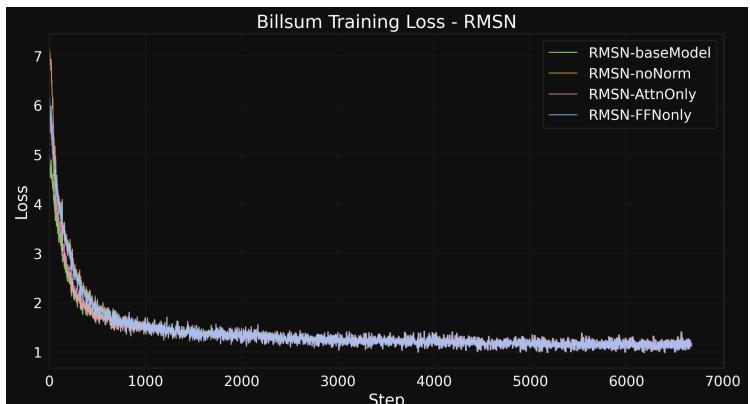
This section analyses how the ablated models' training statistics (logged by Weights & Biases) compare during fine-tuning for downstream tasks.

6.2.1 Fine-Tuning on Summarisation

Training Curves



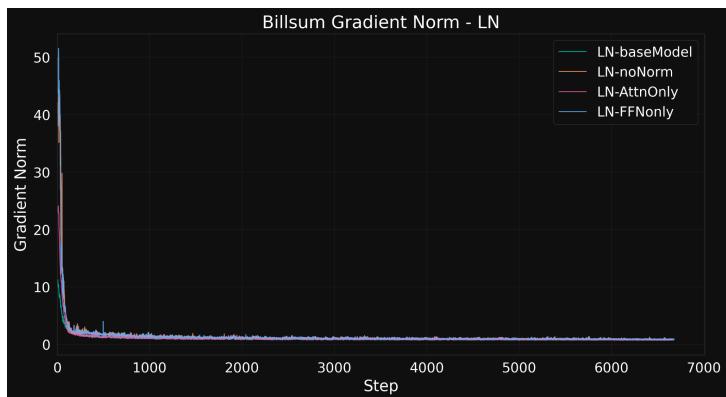
Graph-17: LN Fine-tuning loss curve comparison



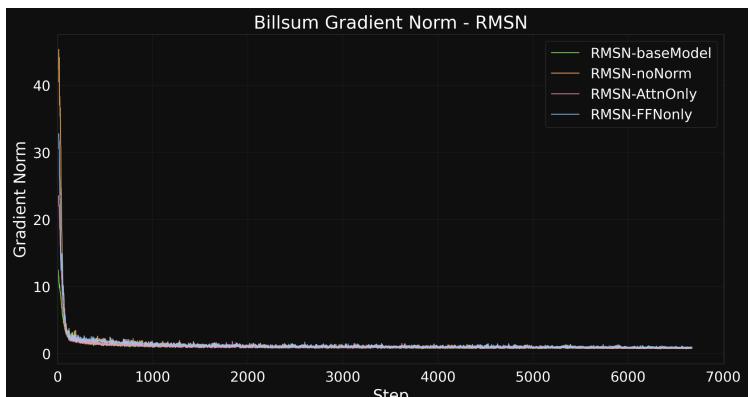
Graph-18: RMSN Fine-tuning loss curve comparison

All models showed similar training loss curves, converging to a loss value around **1.1**, indicating effective fine-tuning across all variants regardless of the normalisation configuration.

Gradient Norm Analysis



Graph-19: LN Fine-tuning Gradient Norm Comparison



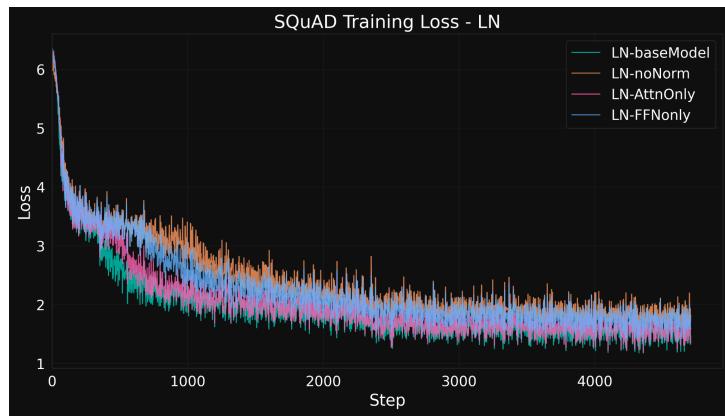
Graph-20: RMSN Fine-tuning Gradient Norm Comparison⁵¹

Global gradients showed similar distributions with minimal fluctuations after initial steps, indicating stable training dynamics across different normalisation configurations without issues like exploding gradients.

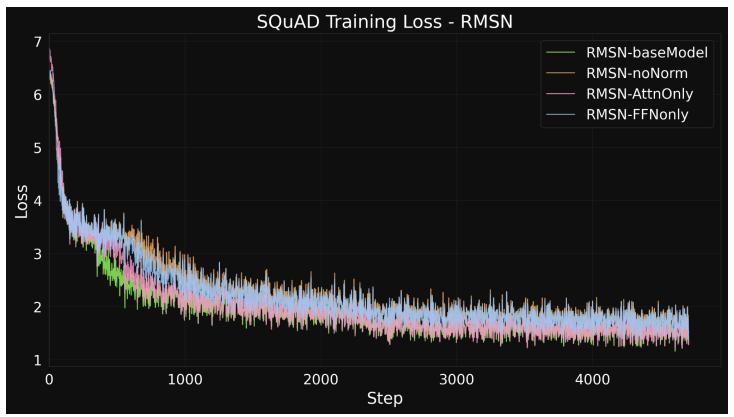
⁵¹ Note: For both QA and summarisation, there's also another pattern with the values of loss values and global gradient, but since the graphs are small and unclear here, we will discuss this in a detailed table in the comparative overview at Section-6.2.3 instead.

6.2.2 Fine-Tuning on Question Answering (QA)

Training Curves



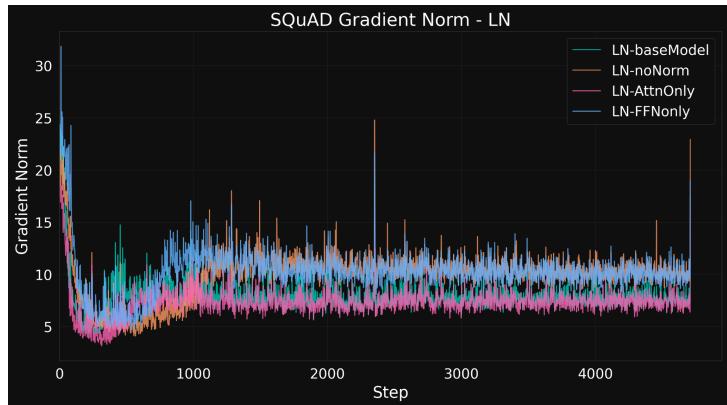
Graph-21: LN fine-tuning training loss comparison



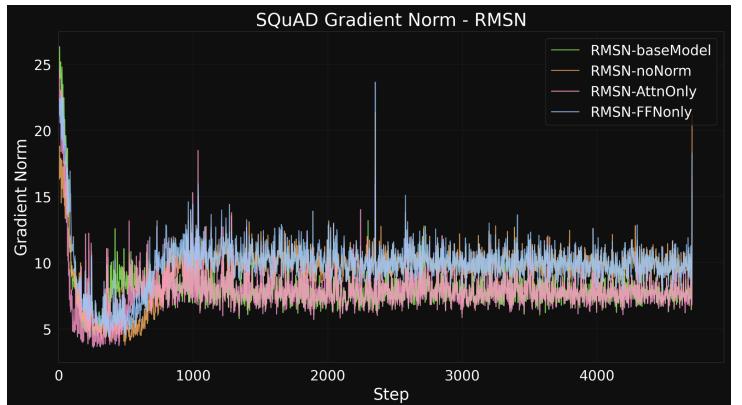
Graph-22: RMSN fine-tuning training loss comparison

All models exhibited similar training loss patterns, successfully converging. However, QA's training showed more fluctuations during convergence compared to Summarisation, suggesting additional challenges and less stable dynamics.

Gradient Norm Analysis



Graph-23: LN Fine-tuning Gradient Norm Comparison



Graph-24: RMSN Fine-tuning Gradient Norm Comparison

Fluctuation is also seen in Global Gradient, which likely stems from SQuAD dataset's (Rajpurkar et al., 2016) characteristics—short answers or occasionally no answers—which require the model to adapt rapidly when predictions are incorrect, resulting in higher variability in gradients. These same characteristics could also explain the fluctuations in training loss, as model responses are often either entirely correct or entirely incorrect.

6.2.3 Comparative Analysis

The mean training loss and gradient norm values for the last ~700 steps were treated as converged values in our analysis, since the loss had stabilised at this stage (see Graphs 17-24). To compare, we included detailed statistics for both downstream tasks below:

Ablated Model Variants	Training Curve			Global Gradient Norm			Fine-Tune Time Taken
	Initial Loss	0 - 6000 Steps	Final 677 Steps	Initial Gradient	0 - 6000 Steps	Final 677 Steps	
LN - BaseModel	4.47	1.37	1.14	Inf	1.05	0.77	1254
RMSN - BaseModel	4.73	1.38	1.14	Inf	1.11	0.80	1264
LN - AttnOnly	5.69	1.39	1.14	Inf	1.13	0.78	1260
RMSN - AttnOnly	5.67	1.40	1.15	Inf	1.18	0.80	1259
LN - FFNonly	6.75	1.44	1.15	Inf	1.48	0.90	1258
RMSN - FFNonly	5.85	1.44	1.15	Inf	1.39	0.92	1260
LN - noNorm	7.50	1.47	1.15	Inf	1.50	0.90	1263
RMSN - noNorm	6.93	1.46	1.16	Inf	1.47	0.92	1241

Table-11: Detailed training statistics of all ablated variants during BillSum fine-tuning

Ablated Model Variants	Training Curve			Global Gradient Norm			Fine-Tune Time Taken
	Initial Loss	0 - 4000 Steps	Final 709 Steps	Initial Gradient	0 - 4000 Steps	Final 709 Steps	
LN - BaseModel	6.31	2.03	1.51	22.08	8.03	7.78	1860.28
RMSN - BaseModel	6.45	2.05	1.52	24.66	8.15	7.78	1860.30
RMSN - AttnOnly	6.65	2.11	1.53	21.70	7.92	7.77	1860.28
LN - AttnOnly	6.24	2.12	1.57	18.34	7.31	7.42	1853.86
RMSN - FFNonly	6.41	2.32	1.72	21.51	9.94	9.91	1860.25
LN - FFNonly	6.15	2.36	1.73	23.11	10.27	10.10	1858.85
RMSN - noNorm	6.33	2.37	1.75	16.27	9.53	9.99	1830.29
LN - noNorm	6.11	2.49	1.81	18.86	9.83	10.27	1853.97

Table-12: Detailed training statistics of all ablated variants during SQuAD fine-tuning

Observed Trends Across Ablated Models

A consistent trend was observed in initial and final training loss and global gradient values⁵², following this order:

$$\text{BaseModel} < \text{AttnOnly} < \text{FFNOnly} < \text{NoNorm}$$

Models with more normalisation layers began with lower loss values and more stable gradient norms. We think this might be because the pre-trained weights are accustomed to normalisation layers. When these layers were removed (like noNorm models), the weights were less adapted, resulting in higher loss initially and the need for aggressive adaptation to learn, hence, less stable gradient norms.

Although more normalised models had better training loss, the difference wasn't substantial, suggesting that extensive normalisation may be unnecessary if simpler models achieve comparable performance. However, converged training loss alone isn't a sufficient metric to evaluate model performance. Hence, downstream task evaluations in Section-6.4 will provide further insight.

Fine-Tuning Time Observations

Unexpectedly, there's negligible difference in time taken for fine-tuning across ablated model variants. We expected models with fewer normalisation layers to train faster due to reduced computational requirements, but this wasn't observed.

⁵² For loss values, a lower value indicates better performance, so more normalised models generally achieve higher performance. For global gradient norm values, a lower magnitude reflects more stable training, meaning that more normalised models, like the base model, contribute to greater stability throughout the training process.

6.3 Internal Gradient Analysis

Given our research question:

"How necessary are different normalisation techniques (LN, RMSN, PN) for decoder-only transformers (GPT-2), and which configuration yields the best performance in terms of training effectiveness and downstream tasks?"

To determine the necessity of normalisation techniques in decoder-only Transformers, we analysed the internal gradients of ablated models both before and after fine-tuning on downstream tasks.

Observations

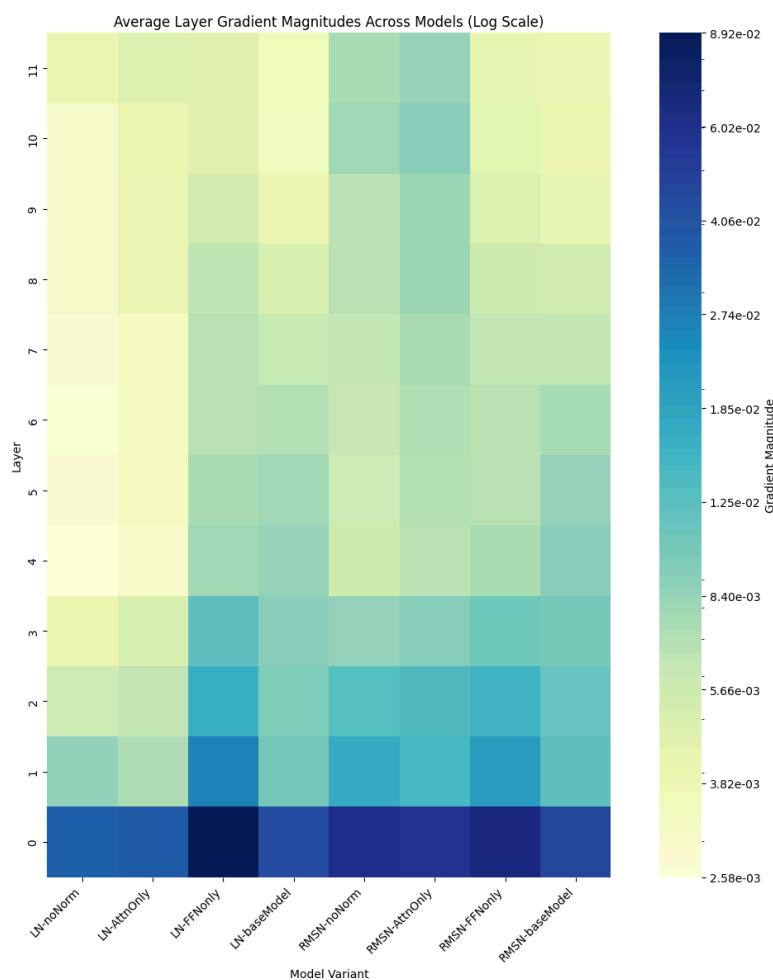


Figure-24: Gradient distribution across ablated model variants' 12 layers before fine-tuning⁵³

⁵³ Figure-24, Figure-25 and Figure-26 are graphed using Matplotlib (Hunter, 2007, pp. 90-95). The method in which gradients were extracted from the models can also be seen in source code in Appendix C.

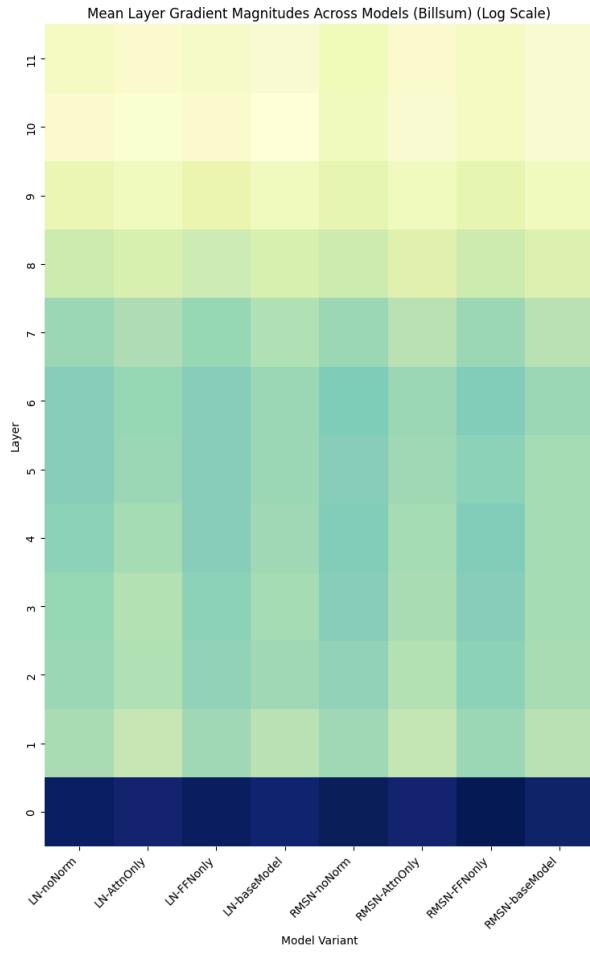


Figure-25: Gradient distribution after BillSum fine-tuning

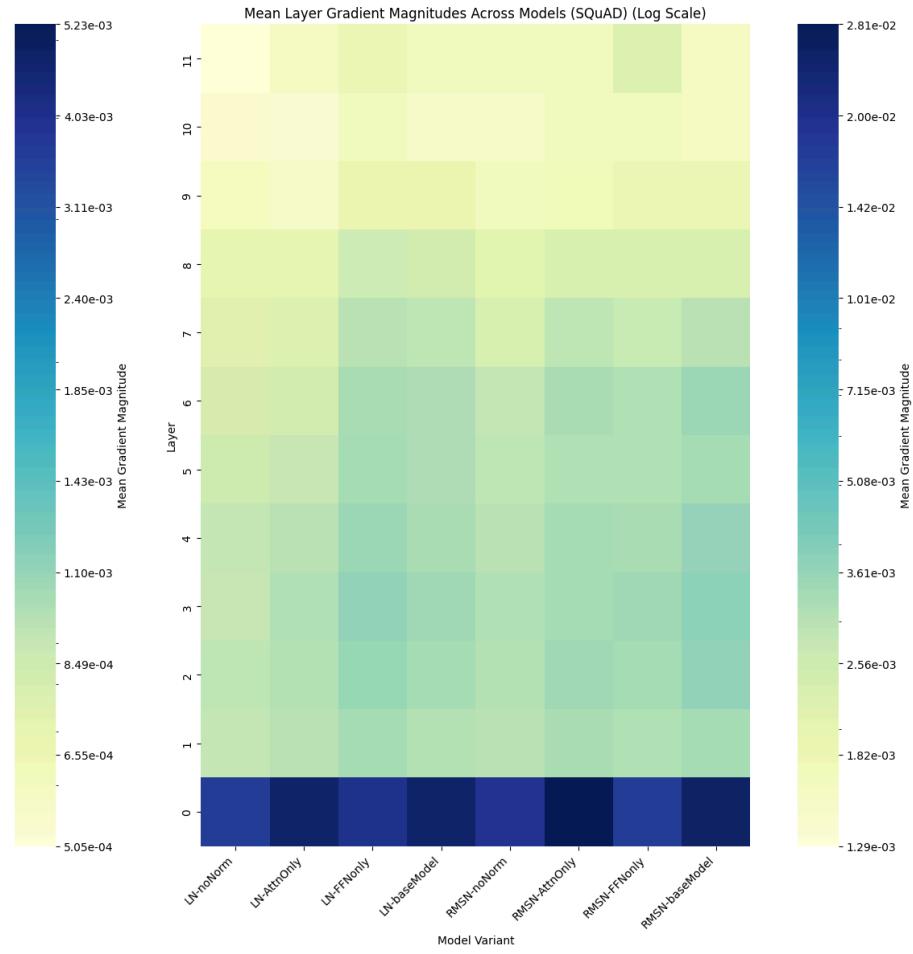


Figure-26: Gradient distribution after SQuAD fine-tuning

Implications

Our observations of these heatmaps:

Condition	Gradient Heatmaps	Observation	Implications
Before Fine-Tuning	Figure-24	Gradient distributions were different between ablated models and baseModels.	Pretrained weights weren't optimised for the absence of normalisation layers, resulting in differences with baseModels.
Post Fine-Tuning	Figure-25 & Figure-26	Gradients of all model variants are very similar ⁵⁴ to one another.	Models likely adjusted during fine-tuning to achieve comparable internal states, regardless of normalisation configuration.

Table-13: Our observations of models' internal gradient before and post fine-tuning

These observations imply that normalisation techniques may not be as crucial as previously thought. Since the ablated models can achieve similar internal gradients despite having less normalisation layers. Meaning normalisation layers may be redundant, since less normalisation requires less computational resources, yet achieves similar internal states.

⁵⁴ Internal gradients are valuable for examining what's happening inside the models. Because if the gradients across models are similar for the same prompts, it suggests that the models themselves are very similar.

6.4 Downstream Task Evaluations

We evaluated our models' on three downstream tasks (Text Generation, Summarisation, QA) with traditional metrics and GPT-4o metrics.

Evaluation Metrics Analysis Methodology

Our analysis of metrics followed this approach:

1. **Score Distribution Analysis:** Visualised score distributions and medians with violin plots (Hintze & Nelson, 1997).
2. **Statistical Testing:**
 - **Kruskal-Wallis Test** (Kruskal & Wallis, 1952): Checked for significant differences in score distributions among models (see Table-14).⁵⁵
 - **Effect Size (η^2)** (Cohen, 1973): Calculated η^2 to assess the magnitude of differences between score distributions (see Table-15).
 - **Dunn's Test** (Dunn, 1964, pp. 241-252): When Kruskal-Wallis' p-value indicates significant differences and η^2 is significant, we conducted Dunn's test for pairwise comparisons to identify models that's exceptionally different from others, for further insights.
3. **Qualitative Analysis:** It's a good idea to review the raw data that you're working with⁵⁶, so we reviewed raw snippets of GPT-4o evaluations to gain further insights.
4. **Mean Score Comparison:** We analysed mean scores of each metric with radar charts (Chambers, 2018) to determine performance trends.

This methodology allowed us to determine the best normalisation configuration and assess the necessity of normalisation techniques based on differences in score distributions. If score distributions and mean scores are...

(Turn to next page)

⁵⁵ Kruskal-Wallis and Dunn's tests were chosen over ANOVA (Fisher, 1925) and Tukey's HSD (Tukey, 1949, pp. 99-114) because, as shown in the violin plots (Figure-28, Figure-33, etc.), none of our score distributions were normally distributed.

⁵⁶ It's crucial to thoroughly examine the raw data before conducting deeper statistical analyses. For instance, in the QA task, inspecting the raw responses helped identify a bug in the Exact Match calculation, leading to a more accurate understanding of the models' performance.

similar, it suggests that normalisation may not be necessary and less normalised models are the best configuration, since they've less training complexity for similar performance.

P-Value Range	Interpretation
$p < 0.01$	Strong evidence to reject null hypothesis
$0.01 \leq p < 0.05$	Moderate evidence to reject null hypothesis
$0.05 \leq p < 0.1$	Weak evidence to reject null hypothesis

Table-14: Threshold values for Kruskal-Wallis p-values

Effect Size Category	Threshold (η^2)
Small	$\eta^2 \geq 0.01$
Medium	$\eta^2 \geq 0.06$
Large	$\eta^2 \geq 0.14$

Table-15: Threshold values for effect sizes

Note: In this section's statistical analysis tables, threshold values will be colour-coded similarly like Table-14 and Table-15 for clarity.

(Turn to next page)

6.4.1 Text Generation

We evaluated the models' ability to generate coherent text when inputted 25 custom prompts, generated by O1-preview (OpenAI, 2024).

Since traditional metrics aren't suitable for open-ended text generation (Hashimoto et al., 2019), we relied on GPT-4o evaluations only.

(Turn to next page)

Score Distribution Analysis



Figure-27: Comparison of models' average performance across GPT-4o's evaluation metrics

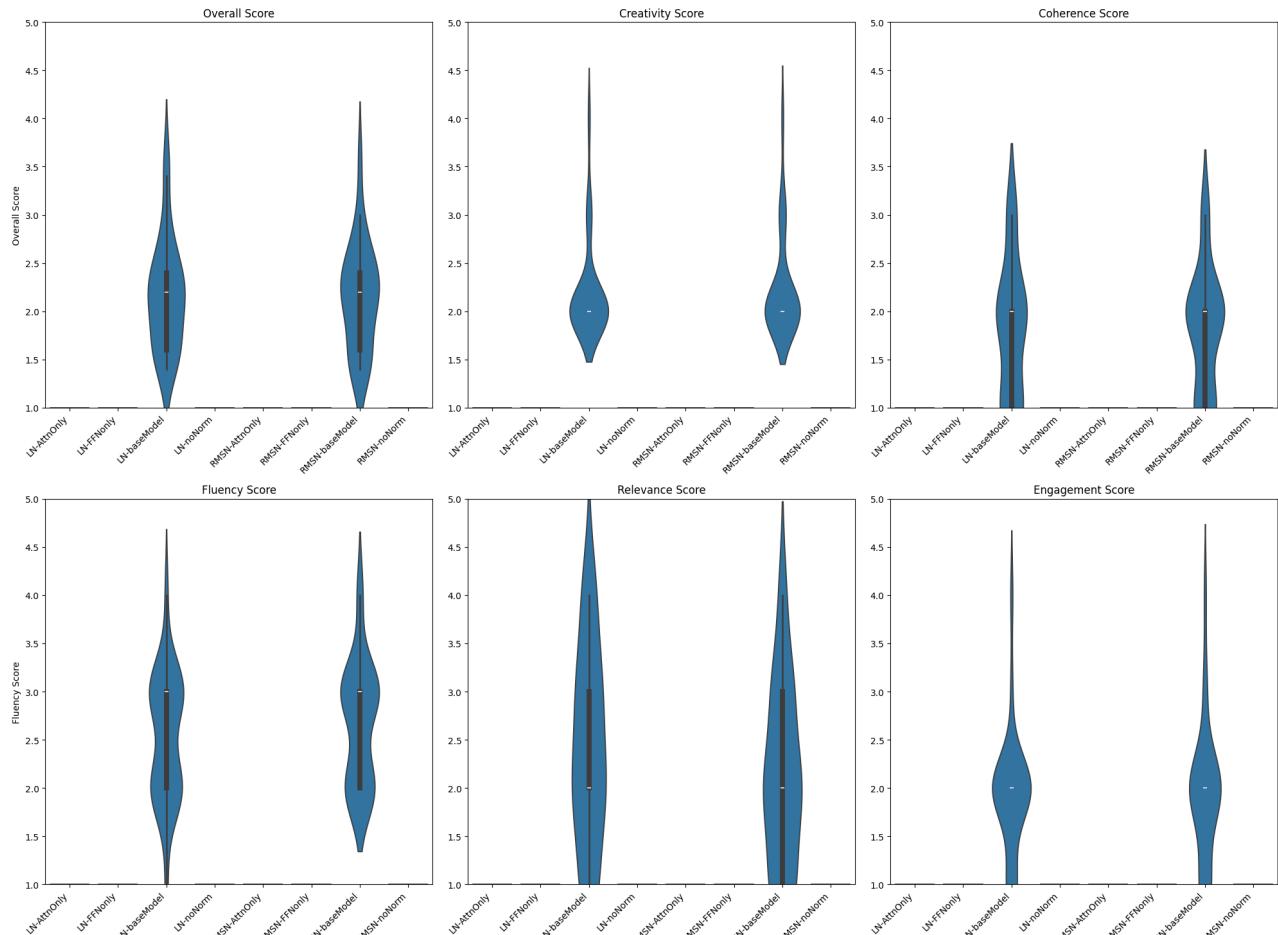


Figure-28: Violin plots of models' score distributions across GPT-4o evaluation metrics⁵⁷

⁵⁷ Figure-27 and Figure-28 are graphed using Matplotlib (Hunter, 2007, pp. 90-95). Source code is in appendix C.

Statistical Analysis

Metric	p-value	Effect Size (η^2)
Overall Score	0.00	0.94
Creativity Score	0.00	0.95
Coherence Score	0.00	0.55
Fluency Score	0.00	0.92
Relevance Score	0.00	0.66
Engagement Score	0.00	0.69

Table-16: Statistical analysis of model score distributions across GPT-4o evaluation metrics

All metrics yielded p-values <0.05 and very large effect sizes, indicating significant differences among the models' score distributions. Hence, we'll perform the Dunn's test to gain insights.

Dunn's test

	LN noNorm	RMSN noNorm	LN AttnOnly	RMSN AttnOnly	LN FFNonly	RMSN FFNonly	LN baseModel	RMSN baseModel
LN noNorm		1.00	1.00	1.00	1.00	1.00	0.00	0.00
RMSN noNorm	1.00		1.00	1.00	1.00	1.00	0.00	0.00
LN AttnOnly	1.00	1.00		1.00	1.00	1.00	0.00	0.00
RMSN AttnOnly	1.00	1.00	1.00		1.00	1.00	0.00	0.00
LN FFNonly	1.00	1.00	1.00	1.00		1.00	0.00	0.00
RMSN FFNonly	1.00	1.00	1.00	1.00	1.00		0.00	0.00
LN baseModel	0.00	0.00	0.00	0.00	0.00	0.00		1.00
RMSN baseModel	0.00	0.00	0.00	0.00	0.00	0.00	1.00	

Table-17: Pairwise Kruskal-Wallis p-value comparison of models' "Overall Score" distributions (3 d.p.)

Observations from Dunn's Test:

Model Comparison	Observation
Ablated Models (NoNorm vs. AttnOnly vs. FFNonly)	Significantly similar to each other, but significantly different from BaseModels
BaseModels vs. BaseModels	Statistically similar
BaseModels vs. Ablated Models	Significantly different

Table-18: Observations from Dunn's Test

These observations are supported by Figure-28, where baseModel distributions appeared similar, while ablated models differed from each other.

Sample Responses and Evaluations (GPT-4o raw snippets)⁵⁸

To better understand the GPT-4o quantitative results, we examined sample responses:

model_name	task	prompt	generated_text	Overall Score	Creativit Score	Coherenc Score	Fluenc Score	Relevanc Score	Engageme Score	Overall Feedback	Comments on Columns
LN_baseModel	Text Generation	In the quiet village nestled among the hills,	In the quiet village nestled among the hills, it was only recently that a man had built a small house for himself on the edge of a gorge that was called the "Little Big Hole." As it is in the town of Bitter Lakes, "it must have been one of the most wonderful sights to be seen in a natural valley. But to see it all there is one thing to keep in mind": It is the "big hole", a grove of spruce and fir trees that has been filled with the efflorescence of so many of the woods that most people have never heard of. The place was always filled with "a wild frenzy". This was caused, in part, by the thick-rooted, "belly-lace" which it had always harboured. No one knows where this big hole actually started - it seems to have simply vanished somewhere about the early 19th century when the first European explorers, led by one Robert Hunter, came in search of what had been previously a wilder place of woodlands and bushes. In this sense, it was the place where the forest's bounty was being gathered.	3.6	4	3	3	4	4	The generated text is creative and engaging, with a rich narrative that captures the reader's interest. However, it suffers from occasional coherence issues and grammatical errors that detract from its overall quality. With some refinement, it could be a compelling piece of writing.	The model name "shng2025/GPT-Valkyrie_LN-124m_baseModel" suggests a specialized version of GPT with a focus on literary narratives, as indicated by "LN" (likely standing for "Literary Narrative"). The "baseModel" variant implies this is a foundational version, possibly without additional fine-tuning. The prompt about a quiet village nestled among the hills sets a serene and potentially mysterious scene, which the generated text explores with creativity and depth. However, the text's occasional lack of coherence and fluency suggests that while the model has strong narrative capabilities, it may benefit from further refinement to improve consistency and readability.

Figure-29: Highest scoring response on text generation with GPT-4o's feedback

model_name	task	prompt	generated_text	Overall Score	Creativit Score	Coherenc Score	Fluenc Score	Relevanc Score	Engageme Score	Overall Feedback	Comments on Columns
LN_noNorm	Text Generation	Once upon a time in a distant galaxy,	Once upon a time in a distant galaxy,Phase HL7- is- the for --.-♦-, we or is are and the sub-?♦ otherwise even.? on and/- in,- or the- in, is are?---, for in?♦- we-♦ to?.. --.♦- for? (2+♦\$♦ in).. a---♦@A@F-S@2-♦- automatically-, without? - in- ♦? -, at in,- and in en for:5 ♦- in are? \$@?♦? - at in development? S. development?,?3-?.. 4-- and in and at? -?/S-,--based to3-D-(2-♦-5- are?---D- e---5- are?7-(E-?)-.7-(6-? D.F-?1-?---S.♦?---S.-M-((2S-((1-4-D4-?1-S.A--- D/5-5-6-2-A-We-6?A-A-?1-S6-12-A7----2---- The192--- E-9-3-S-12,6-1.1350-K/F(2-5- I-1/M/♦(--- -2S(1-As---6-?---B)--E-?♦I3-2(C.-B-+We-3-5-?7-A?♦. dups to-181-?♦ debt1-4-!fC?♦ code policy2?♦ results-As- 4-3-I-5---? conduct---A trades-1/By1As10-5-7(3-St or9-What?9-?♦?2As2D8-A1-3-6?-.What-S4-B-M---(s- In-If5-B3b content-a and? -8-15---(7- and-S-1/1 (1-I to- in-I-1/s-3) or-?A and and? -15---?2-?4-?---?1-5---?2-R-1 or-13, a-3 or---?Plate to?♦ results;?♦ or data as- of the,, for, an---, in after---3-- materi	1.0	1	1	1	1	1	The generated text is of very poor quality. It fails to meet any of the criteria for a good piece of writing. It is incoherent, irrelevant, and lacks any form of creativity or engagement. The text appears to be a malfunction or error in the generation process.	- **Model Name**: The model name suggests it is a variant of GPT, but the output does not reflect the capabilities typically associated with such models. - **Norm Type**: LN (Layer Normalization) is a common technique in neural networks, but its effectiveness is not evident in this output. - **Variant**: The "noNorm" variant might be contributing to the poor quality of the text, as normalization techniques often help in stabilizing and improving the performance of models. - **Prompt**: The prompt is clear and sets up a potential for a creative and engaging story, but the generated text does not follow through. - **Generated Text**: The generated text is essentially gibberish, with no meaningful content or structure. It does not fulfill the expectations set by the prompt.

Figure-30: Lowest scoring response on text generation with GPT-4o's feedback

⁵⁸ We also examined the full 200 prompts for Text Generation evaluated by GPT-4o, which can be found in Appendix I.

Overall Observations

The **BaseModels** produced coherent texts (Figure-29), whereas the ablated models generated gibberish. This is because the ablated models weren't fine-tuned post-ablation, leaving their weights unadjusted for the absence of normalisation layers. Consequently, ablated models had similar distributions (all scoring 1.0, see Figure-27 & Figure-28), whilst baseModels had a range of scores across metrics, making their distributions similar to each other.

The similar performance of LN and RMSN BaseModels suggests that the choice of normalisation technique has minimal impact on model performance for text generation.

(Turn to next page)

6.4.2 Summarisation (BillSum Dataset)

Traditional Metrics Results

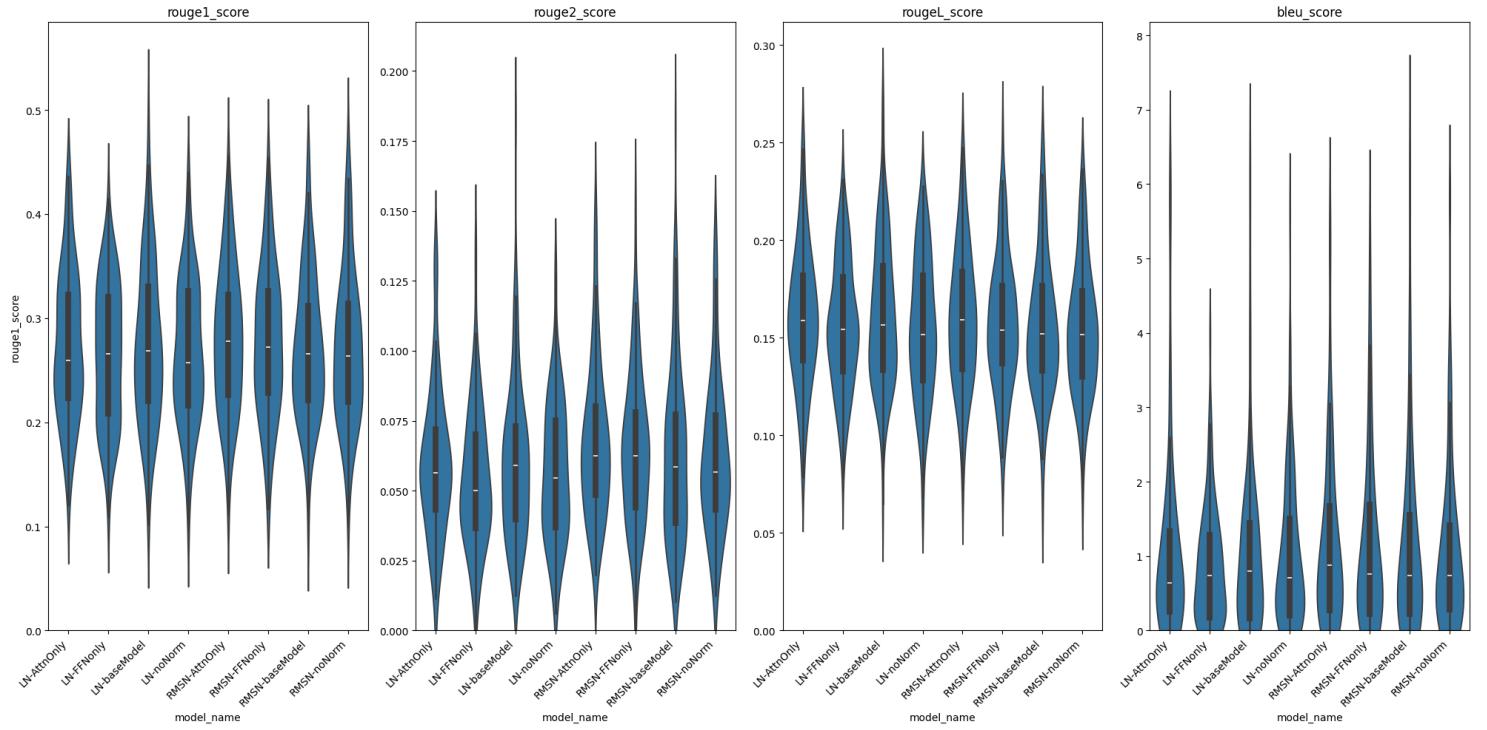


Figure-31: Violin plots of models' score distributions across BillSum traditional metrics

Metric	Kruskal-Wallis p-value	Effect Size (η^2)
ROUGE-1	0.92	0.00
ROUGE-2	0.03	0.01
ROUGE-L	0.73	0.00
BLEU	0.93	0.00

Table-19: Statistical analysis of model score distributions across BillSum traditional metrics

All metrics except for ROUGE-2 have p-value >0.05. However, the effect size for all metrics was negligible, indicating similar distributions across models. This is concurred by Figure-31, showing the metrics distributions and median values being similar.

GPT-4o evaluations

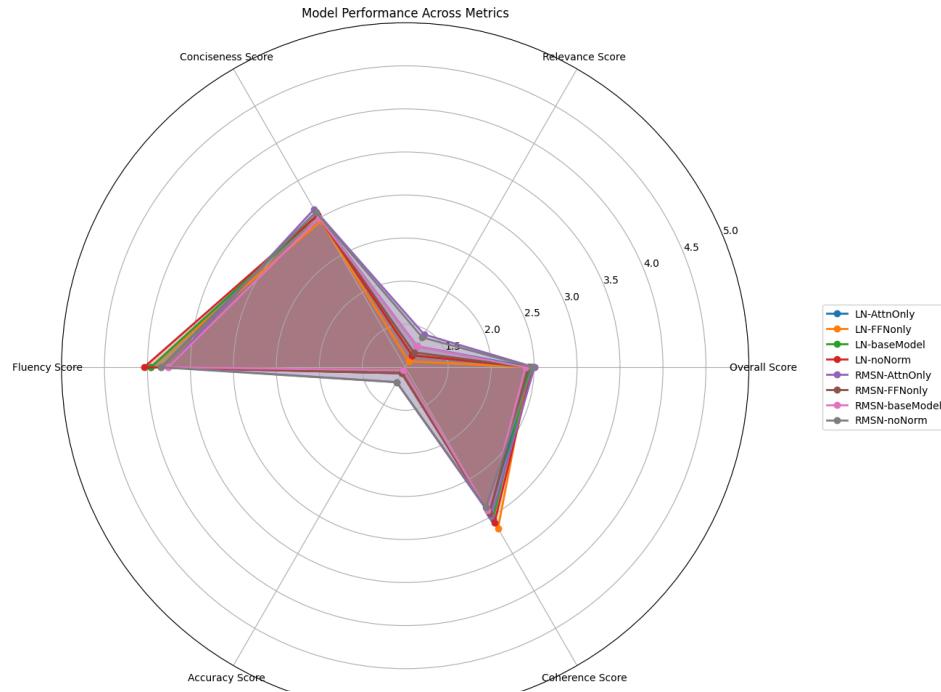


Figure-32: Comparison of models' average performance across GPT-4o's evaluation metrics

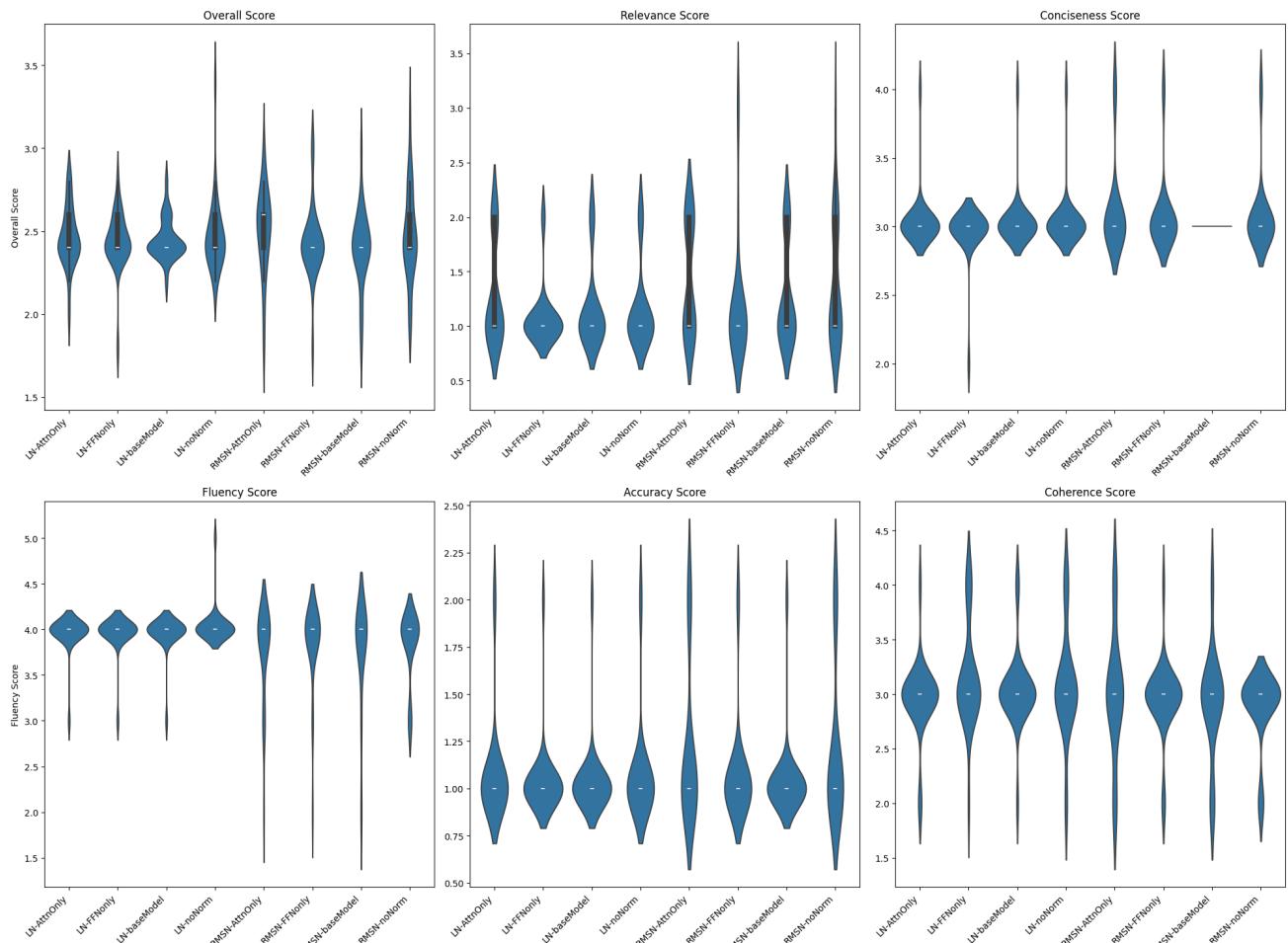


Figure-33: Violin plots of models' score distributions across GPT-4o evaluation metrics⁵⁹

⁵⁹ Figure-27 and Figure-28 are graphed using Matplotlib (Hunter, 2007, pp. 90-95). Source code is in appendix C.

Statistical Analysis

Metric	Kruskal-Wallis p-value	Effect Size (η^2)
Accuracy Score	0.24	0.01
Coherence Score	0.37	0.00
Conciseness Score	0.32	0.01
Fluency Score	0.08	0.03
Overall Score	0.43	0.00
Relevance Score	0.04	0.04

Table-20: Statistical analysis of model score distributions across GPT-4o evaluation metrics

Most metrics showed no significant differences (p-values > 0.05), except **Relevance**. However, all metrics' effect sizes were small, indicating similar distributions across models for GPT-4o metric.

This is unexpected as Figure-33 shows significant differences between model score distributions, but upon closer inspection, it makes sense because all distributions have similar medians.

Sample Responses and Evaluations (GPT-4o raw snippets)

norm type	variant	truncated_input	generated_summary	Overall Relevance Score	Conciseness Score	Fluency Score	Accuracy Score	Coherence Score	Overall Feedback	Comments on Columns	
LN	noNorm	The people of the State of California do enact as follows: SECTION 1. Article 3 (commencing with Section 115810) is added to Chapter 4 of Part 10 of Division 104 of the Health and Safety Code, to read: Article 3. The Consideration of Alternatives for Artificial Turf Infill Act of 2016 115810. The Legislature finds and declares all of the following: (a) Thousands of schools, parks, and local governments have installed artificial turf fields throughout the state. It has allowed them to use fields year round, save water, and save money, among other benefits. (b) Not all artificial turf fields are made from the same materials. While most artificial turf fields use less expensive crumb rubber infill from groundup used car and truck tires, many companies now offer artificial turf infill alternatives made from coconut fibers, rice husks, cork, sand, or virgin crumb rubber. Organic alternative infills can help reduce synthetic turf field temperatures on hot days by as much as 30 degrees compared to	Amends the Public Health Service Act to require the Secretary of Health and Human Services (HHS), acting through the Director of the Centers for Disease Control and Prevention (CDC), to: (1) study and report to Congress on the health and safety of synthetic turf fields; and (2) report to specified congressional committees on the results of such studies. Requires the Administrator of the Agency for Toxic Substances and Disease Registry (ATSDR) to establish a program to provide grants to States for the development and implementation of programs to reduce the exposure of children to lead and other hazardous substances in such fields. Requires the Administrator to establish an advisory committee to advise the Secretary and the Secretary on the implementation of such programs. Authorizes appropriations.	3.4	2	4	5	2	4	The generated summary is fluent and concise but fails to accurately capture the main points of the truncated input. It introduces new information not present in the input text, leading to a low relevance and accuracy score. Improving the alignment with the input text would enhance the overall quality.	- **Model Name**: The model used is shng2025/GPT-Valkyrie_LN-124m_noNorm_Billsum, which appears to be a specialized model for summarizing legislative texts. - **Norm Type**: The norm type is LN (Layer Normalization), which is a technique used in neural networks to stabilize and accelerate training. - **Variant**: The variant is "noNorm," indicating that no additional normalization techniques were applied beyond the standard layer normalization. - **Truncated Input**: The input provides detailed information about the materials used in artificial turf, associated health risks, and historical legal actions. - **Generated Summary**: The summary discusses amendments to the Public Health Service Act and the establishment of programs to reduce exposure to hazardous substances, which are not mentioned in the input text. This indicates a significant deviation from the provided information.

Figure-34: Highest scoring response on text generation with GPT-4o's feedback

LN	FFNonly	The people of the State of California do enact as follows: SECTION 1. Article 2.6 (commencing with Section 66010.96) is added to Chapter 2 of Part 40 of Division 5 of Title 3 of the Education Code, to read: Article 2.6. Office of Higher Education Performance and Accountability 66010.96. (a) The Office of Higher Education Performance and Accountability is hereby established as the statewide postsecondary education coordination and planning entity. The office shall be established in state government within the Governor's office, and shall be under the direct control of an executive director. (b) The Governor shall appoint the Executive Director of the Office of Higher Education Performance and Accountability, who shall perform all duties, exercise all powers, assume and discharge all responsibilities, and carry out and effect all purposes vested by law in the	(This measure has not been amended since it was reported to the House on September 11, 2001. The summary of that version is repeated here.) (Sec. 2) This bill amends the Higher Education Act of 1965 to direct the Department of Education (ED) to establish within the governor's office an Office of Educational Performance (OPP) to be headed by an Executive Director. The OPP shall be composed of the chair of the House of Representatives and two public members. (Sec. 3) The President shall appoint an OPP to serve as the OPP's chief executive officer. An OPP may appoint one or more public members to any office within the state government. The Office shall: (1	1.8	1	2	3	1	2	The generated summary fails to accurately and concisely capture the main points of the truncated input. It introduces irrelevant and incorrect information, which significantly impacts its relevance, accuracy, and coherence. Improvements are needed in ensuring the summary aligns closely with the provided input text.	- **Model Name**: The model used is shng2025/GPT-Valkyrie_LN-124m_FFNonly_Billsum, which appears to be a specialized model for summarizing legislative texts. - **Norm Type**: LN (Layer Normalization) is a common technique used in neural networks to stabilize and accelerate training. - **Variant**: FFNonly indicates that the model variant focuses on feed-forward networks, which might impact its ability to capture complex dependencies in the text. - **Truncated Input**: The input provides detailed information about the establishment and responsibilities of the Office of Higher Education Performance and Accountability in California. - **Generated Summary**: The summary provided is inaccurate and does not reflect the content of the truncated input, indicating a significant mismatch between the input and the generated output.
----	---------	---	--	-----	---	---	---	---	---	--	--

Figure-35: Highest scoring response on text generation with GPT-4o's feedback

Model	BLEU Score	ROUGE-1 Score	ROUGE-2 Score	ROUGE-L Score
RMSN-baseModel	1.161	0.270	0.062	0.155
RMSN-AttnOnly	1.159	0.279	0.067	0.160
RMSN-FFNonly	1.157	0.276	0.064	0.157
RMSN-noNorm	1.073	0.271	0.062	0.153
LN-AttnOnly	1.067	0.274	0.059	0.160
LN-baseModel	1.032	0.273	0.059	0.161
LN-noNorm	1.018	0.267	0.056	0.155
LN-FFNonly	0.877	0.265	0.054	0.156

Table-21: Mean scores of Traditional BillSum metrics (sorted by BLEU)⁶⁰

Model Name	Overall Score (mean)
RMSN AttnOnly	2.504
LN noNorm	2.480
RMSN noNorm	2.480
LN AttnOnly	2.464
LN baseModel	2.448
LN FFNonly	2.440
RMSN FFNonly	2.432
RMSN baseModel	2.400

Table-22: Mean "Overall Scores" of models

⁶⁰ Ranked by BLEU cuz its the stricter metric

Overall Observations

Less normalised models (AttnOnly, NoNorm) consistently outperformed BaseModel, but the differences were minimal—4% (Table-22).

Figure-32 and Figure-33 indicate consistently low accuracy and relevance scores across all models, which have similar distributions, based on our Kruskal-Wallis tests. Upon examining the snippets⁶¹, we found that the models often failed to capture essential content from the original text, likely due to compromises in the fine-tuning methodology (see Section-5.3.3).

Note: Since traditional metrics don't account for accuracy and relevance of the summaries, we proceeded only with GPT-4o metrics (Table-22), excluding traditional metrics (Table-21).

⁶¹ We also examined the full 200 prompts for Summarisation evaluated by GPT-4o, which can be found in Appendix J.

6.4.3 Question Answering (SQuAD Dataset)

Traditional Metrics Evaluation

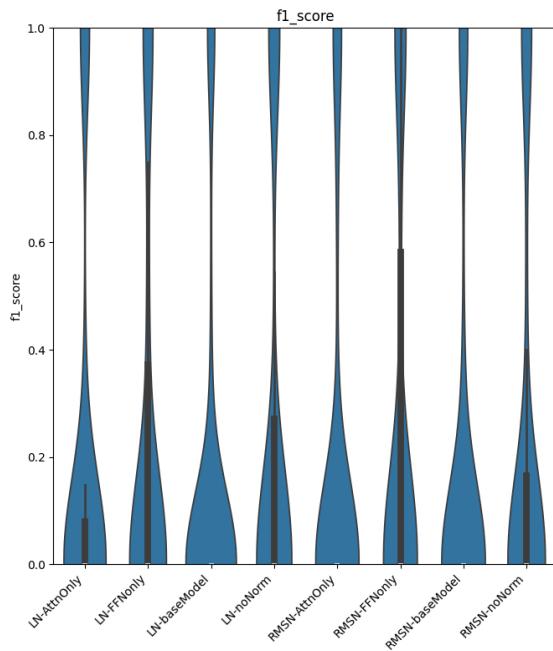


Figure-36: Violin Plot showcasing distribution of F1 Scores⁶²

Metric	p-value	Effect Size (η^2)
F1 Score	0.211	0.00328

Table-23: Statistical analysis of model score distributions across SQuAD traditional metrics

F1 score's p-value was above 0.05 with negligible effect size, indicating no significant differences among models' score distributions.

⁶² Graphed using Matplotlib (Hunter, 2007, pp. 90-95). Source code is in appendix C.

GPT 4o Results:

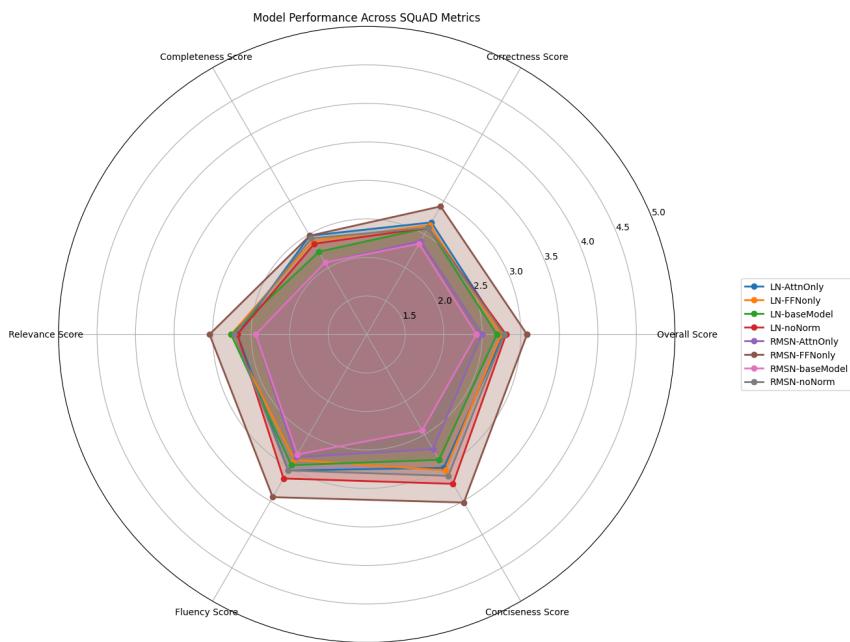


Figure-37: Comparison of models' average performance across GPT-4o's evaluation metrics

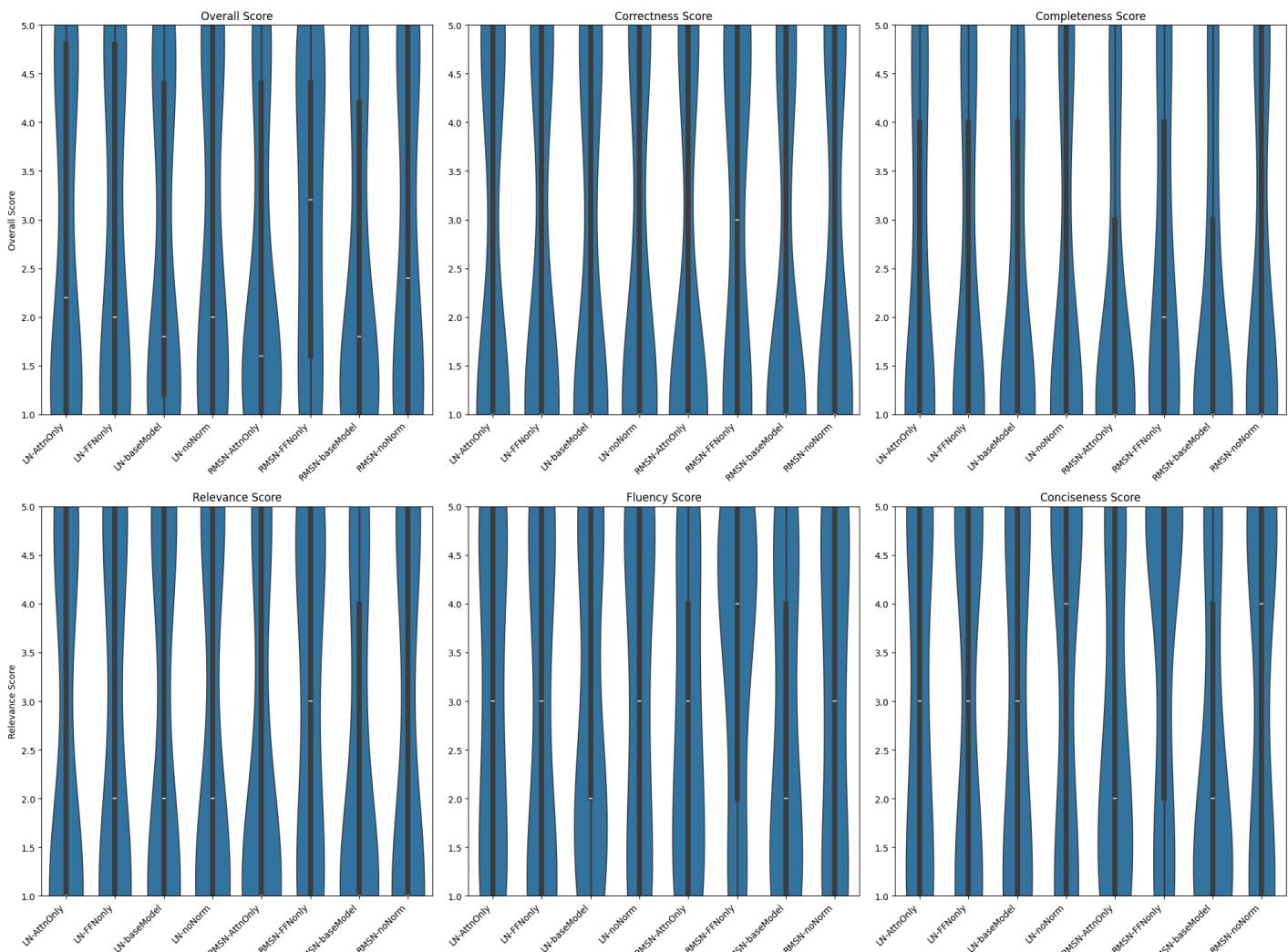


Figure-38: Violin plots of models' score distributions across GPT-4o evaluation metrics⁶³

⁶³ Figure-37 and Figure-38 are graphed using Matplotlib (Hunter, 2007, pp. 90-95). Source code is in appendix C.

Statistical Analysis

Metric	P-value	Effect Size (η^2)
Accuracy Score	0.24	0
Coherence Score	0.37	0
Conciseness Score	0.65	0
Correctness Score	0.97	0
Fluency Score	0.96	0
Completeness Score	0.97	0
Relevance Score	0.04	0.04

Table-24: Statistical analysis of model score distributions across GPT-4o evaluation metrics

All metrics except **Relevance** had p-value below 0.05, but all metrics' negligible effect sizes means that all models have similar score distributions. This aligns with Figure-38, where distributions for all metrics appear similar.

GPT-4o Sample Responses and Evaluations

norm_type	variant	question	context	reference_answers	generated_answer	Overall Score	Correctness Score	Completeness Score	Relevance Score	Fluency Score	Conciseness Score	Overall Feedback	Comments on Columns
LN	AttnOnly	In what country Normandy located?	The Normans (Norman: Nourmands; French: Normands; Latin: Normanni) were the people who in the 10th and 11th centuries gave their name to Normandy, a region in France. They were descended from Norse ("Norman" comes from "Norseman") raiders and pirates from Denmark, Iceland and Norway who, under their leader Rollo, agreed to swear fealty to King Charles III of West Francia. Through generations of assimilation and mixing with the native Frankish and Roman-Gaulish populations, their descendants would gradually merge with the Carolingian-based cultures of West Francia. The distinct cultural and ethnic identity of the Normans emerged initially in the first half of the 10th century, and it continued to evolve over the succeeding centuries.	['France', 'France', 'France']	France	5.0	5	5	5	5	5	The generated answer is perfect in terms of correctness, completeness, relevance, fluency, and conciseness. It directly answers the question with the appropriate information derived from the context.	- **Model Name**: The model name "shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD" suggests that it is a specialized model fine-tuned on the SQuAD dataset, which is known for question-answering tasks. The "AttnOnly" variant indicates that it might be using an attention-only mechanism, which can be effective for focusing on relevant parts of the context. - **Norm Type**: "LN" likely stands for Layer Normalization, a technique used to stabilize and improve the training of deep neural networks. - **Variant**: "AttnOnly" suggests that the model uses an attention-only mechanism, which can help in accurately pinpointing the relevant information in the context. - **Question**: The question "In what country is Normandy located?" is straightforward and seeks a specific factual answer. - **Context**: The context provides detailed historical and geographical information about Normandy, clearly stating that it is a region in France. - **Reference Answers**: The reference answers are all "France," indicating a consensus on the correct answer. - **Generated Answer**: The generated answer "France" is correct and matches the reference answers, demonstrating the model's effectiveness in this instance.

Figure-39: Highest scoring response on text generation with GPT-4o's feedback

norm_type	variant	question	context	reference_answers	generated_answer	Overall Score	Correctness Score	Completeness Score	Relevance Score	Fluency Score	Conciseness Score	Overall Feedback	Comments on Columns
LN	AttnOnly	Who gave their name to Normandy in the 1000's and 1100's	The Normans (Norman: Nourmands; French: Normands; Latin: Normanni) were the people who in the 10th and 11th centuries gave their name to Normandy, a region in France. They were descended from Norse ("Norman" comes from "Norseman") raiders and pirates from Denmark, Iceland and Norway who, under their leader Rollo, agreed to swear fealty to King Charles III of West Francia. Through generations of assimilation and mixing with the native Frankish and Roman-Gaulish populations, their descendants would gradually merge with the Carolingian-based cultures of West Francia. The distinct cultural and ethnic identity of the Normans emerged initially in the first half of the 10th century, and it continued to evolve over the succeeding centuries.	[]	Who	1.0	1	1	1	1	1	The generated answer is entirely inadequate. It does not provide any useful information, fails to address the question, and is not well-formed. The model needs significant improvement in understanding and generating relevant and complete answers.	- **Model Name**: The model name indicates that it is a variant of GPT-Valkyrie with attention-only mechanisms and layer normalization, fine-tuned on the SQuAD dataset. This suggests it should be capable of handling question-answering tasks, but the performance in this instance is poor. - **Norm Type**: LN stands for Layer Normalization, a technique used to stabilize and improve the training of deep neural networks. Despite this, the model's output in this case is not satisfactory. - **Variant**: AttnOnly indicates that the model uses attention mechanisms exclusively. Attention mechanisms are designed to help the model focus on relevant parts of the input, but this did not translate into a correct or relevant answer here. - **Question**: The question is clear and asks for the identity of the people who gave their name to Normandy in the 10th and 11th centuries. - **Context**: The context provides detailed information about the Normans, their origins, and their leader Rollo. This information should have been used to generate a correct and complete answer. - **Reference Answers**: There are no reference answers provided, but the context itself is sufficient to derive the correct answer. - **Generated Answer**: The generated answer "Who" is not useful and does not address the question in any meaningful way.

Figure-40: Lowest scoring response on text generation with GPT-4o's feedback

Model Name	Overall Score (mean)
RMSN FFNonly	3.080
LN noNorm	2.808
RMSN noNorm	2.784
LN AttnOnly	2.776
LN FFNonly	2.744
LN baseModel	2.688
RMSN AttnOnly	2.496
RMSN baseModel	2.424

Table-25: Mean “Overall Scores” of all model variants (GPT-4o evaluation)

Model Name	F1 Score	Percentage of response being Exact Match (%)
RMSN FFNonly	0.262	17.0
LN noNorm	0.246	17.0
LN FFNonly	0.232	15.0
RMSN noNorm	0.225	17.0
LN AttnOnly	0.188	14.0
RMSN AttnOnly	0.181	12.0
RMSN baseModel	0.177	12.0
LN baseModel	0.139	10.0

Table-26: Mean scores of all ablated model variants (Traditional QA metric)

Overall Observations

The score distributions for all models were similar, unaffected by ablation or normalisation technique. Models with fewer normalisation layers (FFNOnly & NoNorm) outperformed BaseModels in both F1 Score and GPT-4o Overall Score. The choice between LN and RMSN had minimal impact on QA performance.

Examining the snippets⁶⁴, we noticed that models were heavily penalised across all metrics for slightly incorrect answers, such as providing a short answer “who” when “no_answer” was expected (Figure-40). Since SQuAD models work by extracting tokens from the context and determining the probability of no answer (HuggingFace Inc., n.d.), fluency isn’t an ideal metric for GPT-4o, and we plan to avoid using it in the future.

⁶⁴ We also examined the full 200 prompts for Question Answering evaluated by GPT-4o, which can be found in Appendix K.

6.4.4 Overall Observations

Below is our overall observations for Section-6.4:

Task	Model Performance (Rankings)	Traditional Metrics Distribution	GPT-4o Metrics Distribution	Description
Text Generation	1. LN 1. RMSN (tied)	N/A	Significantl y Different	Ablated models were excluded from analysis, since they weren't fine-tuned. For base models, LN and RMSN, performed similarly.
Summarisation	1. noNorm 1. AttnOnly (tied) 3. FFNOnly 3. BaseModel (tied)	Similar	Similar	Less normalised models (AttnOnly, NoNorm) slightly outperform others, but performance differences weren't substantial, likely caused by methodology issues mentioned in Section-5.3.3.
Question Answering (QA)	1. NoNorm 2. FFNOnly 3. AttnOnly 4. BaseModel	Similar	Similar	NoNorm consistently outperformed the others. Ablation had minimal effect on performance differences across tasks.

Table-27: overview of observations seen in Section 6.4

Overall Model Rankings Based on Observations

1. noNorm
2. FFNonly
3. AttnOnly⁶⁵
4. BaseModel

Comparison of Normalisation Techniques

1. LN (tied for 1st place)
2. RMSN (tied for 1st place)
3. PN (pre-training failed, not tested)

⁶⁵ We ranked FFNonly higher than AttnOnly overall, because AttnOnly's lead over FFNOnly in Summarisation was small.

7. Conclusion

7.1 Summary of Findings

Answering our research question:

"How necessary are different normalisation techniques (LN, RMSN, PN) for decoder-only transformers (GPT-2), and which configuration yields the best performance in terms of training effectiveness and downstream tasks?"

Our results showed that models with fewer normalisation layers (NoNorm, FFNOnly, AttnOnly) outperformed the fully normalised BaseModel on downstream tasks (Section-6.4.4), suggesting that normalisation layers are unnecessary for decoder-only transformers and can cause reduced performance. After fine-tuning, score distributions were similar regardless of the normalisation technique or ablation method (Section-6.4). Internal gradient analysis also showed similar internal states across models (Section-6.3); whilst training stability was consistent among ablated models (Section-6.2). Interestingly, contrary to expectations (Section-4.2), fewer normalisation layers didn't improve training efficiency (Section-6.2.3), likely due to computational overhead from the HuggingFace API.

For different normalisation techniques, **LN** and **RMSN** performed similarly on downstream tasks (Section-6.4.4). However, **RMSN** was 6% more efficient during pre-training (Section-6.1.4), making it the preferred choice since it offered the same performance for cheaper training cost. **PN** was the least effective, suffering from gradient explosions and training failure (Section-6.1.3). This could be due to differences in training data quality; (Shen et al., 2020, Appendix A.1.) used a lower-quality corpus, which may have allowed PN to perform better, while the high-quality EduFineWeb-10B (Penedo et al., 2024) corpus exposed its limitations. PN's reliance on running statistics is perhaps ill-suited to rapidly adapt to diverse texts, and PN's training speed is a **significant 29% slower** compared to LN.

Furthermore, we discovered that using more complex tokenizers like GPT-4's didn't improve performance for small models like GPT-2 (Section-5.1). The GPT-4 tokenizer required 32-bit precision, doubling memory use and halving training speed.

Overall, NoNorm was the best configuration, offering superior performance without normalisation layers, indicating that normalisation techniques aren't necessary. If normalisation is used, RMSN is a more cost-effective alternative to LN.

7.2 Methodological Limitations

Despite the insights gained, several limitations in our methodology may have influenced our findings:

A. GPT-4o Evaluator

Limitation	Impact	Future Mitigation
Inappropriate metrics for SQuAD QA evaluation (Section-6.4.3)	Impacted the quality of evaluation, as metrics like fluency were less relevant for QA tasks. Models were penalised for minor inaccuracies, which did not accurately reflect their performance.	Review datasets and the nature of downstream tasks before selecting evaluation metrics for GPT-4o. Use task-specific metrics.
High cost of GPT-4o evaluations	Limited us to 25 prompts per model per task (totaling 600 prompts) due to OpenAI API expenses (\$11.4 for over 100,000 tokens), restricting the breadth of evaluation.	We anticipate API costs to decrease in the future, which will enable more prompts to be evaluated, leading to more robust analysis as we can use more prompts per model per task.

Table-28: Limitations of our methodology using GPT-4o

B. Model Size and Computational Constraints

Limitation	Impact	Future Mitigation
Small model size (GPT-2 124m) used due to financial constraints (>USD 350 spent)	Our findings likely apply only to smaller models, which stabilise and learn more easily than larger ones. Consequently, our findings may not generalise to larger, more complex models. (Kaplan et al., 2020, p. 8)	Test with larger models to verify if observed trends of this paper hold at scale.
Limited context window of small models	Affected tasks like summarisation (Section-5.3), as the model could not process longer inputs, leading to performance issues in handling datasets like BillSum.	Use larger models or increase the context window size to accommodate longer inputs.
We used our Ablation fine-tuning method due to computational constraints	Could not comprehensively test the impact of reduced normalisation layers on training stability during pre-training. No stability issues in Section 6.2 may be because weights had already stabilised with normalisation layers present.	Pre-train all ablated model variants from scratch for accurate assessment. Rather than pre-train baseModels then ablate and fine-tune.

Table-29: Limitations due to model size and computational constraints

(Turn to next page)

C. HuggingFace API Overhead

Limitation	Impact	Future Mitigation
Not leveraging performance optimisations like <code>torch.compile()</code>	May explain the lack of observed training efficiency improvements in less normalised models during fine-tuning (Section 6.2). HuggingFace API's overhead may have masked the theoretical computational gains that should come with less normalised models.	Implement fine-tuning pipelines directly in PyTorch to reduce computational overhead.

Table-30: Limitations of using HuggingFace API (HuggingFace Inc., n.d.)

7.3 Future Research Directions

Our findings are likely only applicable to small models like GPT-2. As compute gets cheaper, future research should explore whether these results hold for larger-scale transformers, since training dynamics likely differs with larger models.

We recommend prioritising novel training dynamics by leveraging scaling laws (Kaplan et al., 2020, p. 8) over seeking architectural improvements. Because whilst architectural tweaks can yield incremental gains, there's a limit to the performance improvements achievable through architecture alone (Sutton, 2019).

Currently, significant advancements have been driven by innovative training methodologies enabled by increased computational power. For example, techniques like **Chain-of-Thought reasoning (Wei et al., 2022)** and **Reinforcement Learning with Human Feedback** (Ouyang et al., 2022) have dramatically advanced model intelligence by incorporating human feedback into training the models how to reason. These techniques, applied in OpenAI's most recent O1 (OpenAI, 2024), had achieved remarkable mathematical understanding, achieving a remarkable 83% in the AIME.

8. References

- Adobe. (n.d.). *Adobe Illustrator - Vector Graphics Software*. Adobe. Retrieved November 10, 2024, from
https://www.adobe.com/my_en/products/illustrator.html
- Anthropic. (2024, June 20). *Introducing Claude 3.5 Sonnet*. Anthropic. Retrieved November 9, 2024, from
<https://www.anthropic.com/news/clause-3-5-sonnet>
- Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016, July 21). Layer Normalization. *arXiv*.
<https://doi.org/10.48550/arXiv.1607.06450>
- Bashkar, Y. (2024, September 1). *Decoder-Only Transformers Explained: The Engine Behind LLMs* [Diagram of Decoder-only Transformer]. medium. Retrieved 11 8, 2024, from
<https://medium.com/@yash9439/decoder-only-transformers-explained-the-engine-behind-llms-3a3224086afe>
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., Zhang, Y., & Microsoft Research. (2023, April 13). Sparks of Artificial General Intelligence: Early experiments with GPT-4. *arXiv*, 94.
<https://doi.org/10.48550/arXiv.2303.12712>
- Çallı, E., Sogancioglu, E., Ginneken, B. v., Leeuwen, K. G. v., & Murphy, K. (2021, June 5). Deep learning for chest X-ray analysis: A survey. *Medical image Analysis*, 72, 29.
<https://doi.org/10.1016/j.media.2021.102125>
- Chambers, J. M. (2018). *Graphical Methods for Data Analysis*. Taylor & Francis Group.
<https://doi.org/10.1201/9781351072304>
- Cloudera Fast Forward Labs. (2020, June 9). *Evaluating QA: Metrics, Predictions, and the Null Response*. NLP for Question Answering. Retrieved November 9, 2024, from
https://qa.fastforwardlabs.com/no%20answer/null%20threshold/bert/distilbert/exact%20match/f1/robust%20predictions/2020/06/09/Evaluating_BERT_on_SQuAD.html
- Cohen, J. (1973). Eta-Squared and Partial Eta-Squared in Fixed Factor Anova Designs. *Educational and Psychological Measurement*, 33(1), 107-112. <https://doi.org/10.1177/001316447303300111>
- Dunn, O. J. (1964, August). Multiple Comparisons Using Rank Sums. *Technometrics*, 6(3), 241-252.
<https://doi.org/10.2307/1266041>

- Fisher, R. A. (1925). *Statistical Methods For Research Workers*. Oliver and Boyd.
- GitHub. (n.d.). Git Large File Storage | Git Large File Storage (LFS) replaces large files such as audio samples, videos, datasets, and graphics with text pointers inside Git, while storing the file contents on a remote server like GitHub.com or GitHub Enterprise. Retrieved November 9, 2024, from <https://git-lfs.github.com/>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Google. (n.d.). *Google Colaboratory*. colab.google. Retrieved November 10, 2024, from <https://colab.google/>
- Hashimoto, T. B., Zhang, H., & Liang, P. (2019, June). Unifying Human and Statistical Evaluation for Natural Language Generation. *ACL Anthology*, 1689-1701. <https://doi.org/10.18653/v1/N19-1169>
- Hintze, J. L., & Nelson, R. D. (1997, February 1). Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, 52(2), 181-184. <https://doi.org/10.1080/00031305.1998.10480559>
- HuggingFace Inc. (n.d.). Hugging Face – The AI community building the future. Retrieved November 9, 2024, from <https://huggingface.co/>
- HuggingFace Inc. (n.d.). *Fine-tune a pretrained model*. Hugging Face. Retrieved November 9, 2024, from <https://huggingface.co/docs/transformers/training>
- HuggingFace Inc. (n.d.). *Question answering*. Hugging Face. Retrieved November 9, 2024, from https://huggingface.co/docs/transformers/tasks/question_answering
- Hunter, J. D. (2007, June 18). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 90-95. <https://doi.org/10.1109/MCSE.2007.55>
- Ioffe, S., Szegedy, C., & Google Inc. (2015, March 2). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv*, 11. <https://doi.org/10.48550/arXiv.1502.03167>
- Jiang, Z., g, J., Zhu, H., & Pan, D. Z. (2023, October 26). Pre-RMSNorm and Pre-CRMSNorm Transformers: Equivalent and Efficient Pre-LN Transformers. *arXiv*, 17. <https://doi.org/10.48550/arXiv.2305.14858>
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020, January 23). Scaling Laws for Neural Language Models. *arXiv*, 19. <https://doi.org/10.48550/arXiv.2001.08361>
- Karpathy, A. (2022, December 29). *karpathy/nanoGPT: The simplest, fastest repository for training/finetuning medium-sized GPTs*. GitHub. Retrieved November 9, 2024, from <https://github.com/karpathy/nanoGPT>

- Karpathy, A. (2024, June 10). *Neural Networks: Zero to Hero*. YouTube. Retrieved November 9, 2024, from <https://www.youtube.com/playlist?list=PLAqhIrjkxbuWI23v9cThsA9GvCAUhRvKZ>
- Karpathy, A. (2024, February 20). *Let's build the GPT Tokenizer*. YouTube. Retrieved November 10, 2024, from <https://www.youtube.com/watch?v=zduSFxRajkE>
- Kharya, P. (2020, May 14). *What is the TensorFloat-32 Precision Format?* NVIDIA Blog. Retrieved November 9, 2024, from <https://blogs.nvidia.com/blog/tensorfloat-32-precision-format/>
- Korhonen, A., Traum, D., & Màrquez, L. (Eds.). (2019, July). Energy and Policy Considerations for Deep Learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645-3650. <https://aclanthology.org/P19-1355.pdf>
- Kornilova, A., & Eidelman, V. (2019, November). BillSum: A Corpus for Automatic Summarization of US Legislation (L. Wang, J. C. K. Cheung, G. Carenini, & F. Liu, Eds.). *ACL Anthology*, 48-56. <https://doi.org/10.18653/v1/D19-5406>
- Kruskal, W. H., & Wallis, W. A. (1952, December). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47(260), 583-621. <https://doi.org/10.2307/2280779>
- Li, S., Zhao, Y., Varma, R., Salpekar, O., Noordhuis, P., Li, T., Paszke, A., Smith, J., Vaughan, B., Damania, P., & Chintala, S. (2020, June 28). PyTorch Distributed: Experiences on Accelerating Data Parallel Training. *arXiv*, 14. <https://doi.org/10.48550/arXiv.2006.15704>
- Lin, C.-Y. (2004, July). ROUGE: A Package for Automatic Evaluation of Summaries. *Anthology*, <https://aclanthology.org/volumes/W04-10/>, 74-81. <https://aclanthology.org/W04-1013/>
- Maharaj, S. S. (2024, March 22). *Batch vs Layer Normalization*. Zilliz. Retrieved November 10, 2024, from <https://zilliz.com/learn/layer-vs-batch-normalization-unlocking-efficiency-in-neural-networks>
- Meta AI. (2016, September). *PyTorch documentation — PyTorch 2.5 documentation*. PyTorch. Retrieved November 9, 2024, from <https://pytorch.org/docs/stable/index.html>
- Meta AI. (2023). *Distributed communication package - torch.distributed — PyTorch 2.5 documentation*. PyTorch. Retrieved November 9, 2024, from <https://pytorch.org/docs/stable/distributed.html>
- Meta AI. (2023). *torch.compile — PyTorch 2.5 documentation*. PyTorch. Retrieved November 9, 2024, from <https://pytorch.org/docs/stable/generated/torch.compile.html>
- Narang, S., Diamos, G., Elsen, E., Micikevicius, P., Alben, J., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev,

O., Venkatesh, G., & Wu, H. (2018, February 15). Mixed Precision Training. *arXiv*, 12.

<https://doi.org/10.48550/arXiv.1710.03740>

Nvidia. (2024). *NVIDIA H100 Tensor Core GPU*. NVIDIA. Retrieved November 9, 2024, from

<https://www.nvidia.com/en-us/data-center/h100/>

Nvidia. (2024). *System Management Interface SMI*. NVIDIA Developer. Retrieved November 10, 2024, from

<https://developer.nvidia.com/system-management-interface>

Nvidia. (2024, September 30). *CUDA C++ Best Practices Guide*. NVIDIA Docs. Retrieved November 9, 2024,

from <https://docs.nvidia.com/cuda/cuda-c-best-practices-guide/>

Nvidia Deep Learning Institute. (2024). *Nvidia GTC Course*. Nvidia Deep Learning Institute. Retrieved November

10, 2024, from https://learn.nvidia.com/courses/course?course_id=course-v1:DLI+C-FX-03+V3

Nvidia Deep Learning Institute. (2024). *Nvidia GTC Courses*. Nvidia Deep Learning Institute. Retrieved

November 10, 2024, from

https://learn.nvidia.com/courses/course?course_id=course-v1:DLI+C-FX-07+V1

OpenAI. (n.d.). *OpenAI Platform*. OpenAI Platform. Retrieved November 9, 2024, from

<https://platform.openai.com/docs/introduction>

OpenAI. (n.d.). *Tiktokenizer*. Tiktokenizer. Retrieved November 9, 2024, from <https://tiktokenizer.vercel.app/>

OpenAI. (2022, December 15). *openai/tiktoken: tiktoken is a fast BPE tokeniser for use with OpenAI's models*.

GitHub. Retrieved November 9, 2024, from <https://github.com/openai/tiktoken>

OpenAI. (2024). *OpenAI Platform*. OpenAI Platform. Retrieved November 9, 2024, from

<https://platform.openai.com/docs/models/gpt-4o>

OpenAI. (2024, September 12). *Learning to Reason with LLMs*. OpenAI. Retrieved November 8, 2024, from

<https://openai.com/index/learning-to-reason-with-langs>

OpenAI. (2024, September 12). *OpenAI o1 Hub*. OpenAI. Retrieved November 9, 2024, from

<https://openai.com/o1/>

OpenAI. (2024, October). *Models*. OpenAI API. Retrieved November 9, 2024, from

<https://platform.openai.com/docs/models>

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K.,

Ray, A., Schulman, J., Hilton, J., & Kelton, F. (2022, March 4). Training language models to follow

instructions with human feedback. *arXiv*. <https://doi.org/10.48550/arXiv.2203.02155>

Papineni, K., Roukos, S., Ward, T., Zhu, W.-J., Charniak, E., & Lin, D. (2002, July). Bleu: a Method for Automatic Evaluation of Machine Translation (P. Isabelle, Ed.). *Anthology, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311-318.

<https://doi.org/10.3115/1073083.1073135>

Penedo, G., Kydlíček, H., allal, L. B., Lozhkov, A., Mitchell, M., Raffel, C., Werra, L. V., & Wolf, T. (2024, October 31). The FineWeb Datasets: Decanting the Web for the Finest Text Data at Scale. *arXiv*, 38.

<https://doi.org/10.48550/arXiv.2406.17557>

Press, O., & Wolf, L. (2017, February 21). Using the Output Embedding to Improve Language Models. *arXiv*, 7.

<https://doi.org/10.48550/arXiv.1608.05859>

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language Models are Unsupervised Multitask Learners* [PDF]. OpenAI. Retrieved 11 8, 2024, from

https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016, November). SQuAD: 100,000+ Questions for Machine Comprehension of Text (J. Su, K. Duh, & X. Carreras, Eds.). *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2383-2392. <https://doi.org/10.18653/v1/D16-1264>

Shen, S., Yao, Z., Gholami, A., Mahoney, M. W., & Keutzer, K. (2020, June 28). *PowerNorm*. GitHub. Retrieved November 6, 2024, from <https://github.com/sIncerass/powernorm>

Shen, S., Yao, Z., Gholami, A., Mahoney, M. W., & Keutzer, K. (2020, June 28). PowerNorm: Rethinking Batch Normalization in Transformers. *arXiv*, 18. <https://doi.org/10.48550/arXiv.2003.07845>

Student. (2024, November 5). *Extended-Essay-Appendix*. GitHub. Retrieved November 6, 2024, from
<https://github.com/Ice-Citron/Extended-Essay-Appendix.git>

Sutton, R. (2019, March 13). *The Bitter Lesson*. Retrieved November 9, 2024, from
https://www.cs.utexas.edu/~eunsol/courses/data/bitter_lesson.pdf

TensorDock. (2024, July 15). *Spot Instances*. TensorDock. Retrieved November 9, 2024, from
<https://docs.tensordock.com/virtual-machines/spot-instances>

TensorDock. (2024, August 5). *TensorDock (GPU provider)*. Welcome to TensorDock. Retrieved November 9,

2024, from <https://docs.tensordock.com/>

3Blue1Brown. (2016, August 6). *Essence of linear algebra*. YouTube. Retrieved November 10, 2024, from https://www.youtube.com/watch?v=fNk_zzaMoSs&list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab

3Blue1Brown. (2017, October 5). *Neural networks*. YouTube. Retrieved November 10, 2024, from https://www.youtube.com/watch?v=aircAruvnKk&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&pp=iAQB

Tolo, F. M. (2024, June 26). *HuggingFaceFW/fineweb-edu · Datasets at Hugging Face*. Hugging Face. Retrieved November 9, 2024, from <https://huggingface.co/datasets/HuggingFaceFW/fineweb-edu>

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023, February 27). LLaMA: Open and Efficient Foundation Language Models. *arXiv*, 27. <https://doi.org/10.48550/arXiv.2302.13971>

Tukey, J. W. (1949, June). Comparing Individual Means in the Analysis of Variance. *Biometrics*, 5(2), 99-114. <https://doi.org/10.2307/3001913>

Tunstall, L., Werra, L. v., & Wolf, T. (2022). *Natural Language Processing with Transformers: Building Language Applications with Hugging Face*. O'Reilly Media.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017, June 12). Attention is All You Need. *arXiv preprint arXiv:1706.03762*. <https://doi.org/10.48550/arXiv.1706.03762>

Vig, J. (2019, April 11). Visualizing Attention in Transformer-Based Language Representation Models. *arXiv*, 6. <https://doi.org/10.48550/arXiv.1904.02679>

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022, January 28). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *arXiv*. <https://doi.org/10.48550/arXiv.2201.11903>

Weights & Biases. (n.d.). *Weights & Biases*. Weights & Biases: The AI Developer Platform. Retrieved November 9, 2024, from <https://wandb.ai/>

Yin, S., & Ungaro, F. (2019, September 20). *Transformer Architecture: The Positional Encoding*. Amirhossein Kazemnejad. Retrieved November 8, 2024, from https://kazemnejad.com/blog/transformer_architecture_positional_encoding/

Zhang, B., & Sennrich, R. (2019, October 16). Root Mean Square Layer Normalization. *arXiv*, 14.

<https://doi.org/10.48550/arXiv.1910.07467>

9. Acknowledgements

I would like to express my immense gratitude to the following individuals and tools for their invaluable contributions to the development of this research paper.

- **Andrej Karpathy** (Karpathy, 2024): A massive thank you for your zero-to-hero course on transformers, which has been an essential foundation for writing this paper. Your course enabled me to grasp the core concepts behind transformers and their application, without which this research would not have been possible.
- **O'Reilly TransformerBook** (Tunstall et al., 2022): This resource taught me the intricacies of fine-tuning and evaluating metrics in GPT models, directly influencing my methodology and results.
- **3Blue1Brown's Linear Algebra** (3Blue1Brown, 2016) **and Neural Network Course** (3Blue1Brown, 2017): Thank you for providing such a clear, graphical explanation of linear algebra, which greatly enhanced my understanding of the mathematical foundations behind the models used in this research.
- **Nvidia GTC Course on DNN Deployment and LLMs** (Nvidia Deep Learning Institute, 2024): The insights gained from this course were pivotal in understanding the practical deployment of deep neural networks and large language models.
- **Google Colab** (Google, n.d.): Your rented, affordable GPUs allowed me to scale and experiment with training code efficiently, making it possible to conduct extensive model experiments.
- **TensorDock** (TensorDock, 2024): Thank you for offering cost-effective GPUs through spot instances and for providing excellent customer support when technical challenges arose during my work.
- **GPT-4o, GPT-4, O1-preview, O1-mini** (OpenAI, 2024), **and Claude-3.5 Sonnet** (Anthropic, 2024): These powerful tools were crucial in assisting me with programming, generating graphs, and writing the content of this paper, making the entire process more efficient and manageable.

I would also like to extend my gratitude to the open-source community as a whole, whose shared knowledge and resources have made this research possible.

10. Appendix

Appendix A

A.1 Custom Prompts for Text Generation

25 custom prompts generated by O1-preview for text generation.

```
prompts = [
    "Once upon a time in a distant galaxy,",  

    "The secret to happiness is",  

    "In the midst of the bustling city,",  

    "As the sun set over the horizon,",  

    "The mysterious package arrived on my doorstep,",  

    "In a world where robots and humans coexist,",  

    "The ancient prophecy foretold that",  

    "Amidst the chaos of the storm,",  

    "The scientist peered into the microscope and saw",  

    "Every morning, she would start her day with",  

    "The last thing I expected to find in the attic was",  

    "Under the cover of darkness,",  

    "With a heavy heart, he decided to",  

    "On the eve of the grand festival,",  

    "The sound of laughter filled the air as",  

    "If time travel were possible,",  

    "Deep beneath the ocean waves,",  

    "In the quiet village nestled among the hills,",  

    "The door creaked open to reveal",  

    "Legends speak of a sword that",  

    "In the year 2525, humanity has",  

    "The aroma of freshly baked bread",  

    "She looked into his eyes and knew that",  

    "On the first day of school,",  

    "The journey to the top of the mountain was",
]
```

A.2 TextGen Evaluation Prompts

Prompt Generated by O1-preview (OpenAI, 2024) for GPT-4o to evaluate model's performance in Text Generation.

As an expert evaluator, your task is to assess the quality of a generated text based on the provided prompt. Please focus on the following criteria:

1. **Creativity**: How original and imaginative is the text?
2. **Coherence**: Is the text logically consistent and does it flow well?
3. **Fluency**: Is the text well-written with proper grammar and style?
4. **Relevance**: Does the text stay on topic and relate to the prompt?
5. **Engagement**: Is the text interesting and engaging to read?

For each criterion, provide:

- **Score**: A number from 1 to 5 (1 is poor, 5 is excellent).
- **Explanation**: A brief justification for the score.

After evaluating each criterion, provide:

- **Overall Score**: The average of the five scores.
- **Overall Feedback**: A short overall feedback.

Additionally, provide a short paragraph commenting on the following columns:

- **Model Name**: {model_name}
- **Norm Type**: {norm_type}
- **Variant**: {variant}
- **Prompt**: [Prompt is provided above.]
- **Generated Text**: [Generated text is provided above.]

Please present your evaluation in the following structured format:

Model Name: {model_name}

Creativity Score: [1-5]

Creativity Explanation: [Your explanation]

Coherence Score: [1-5]

Coherence Explanation: [Your explanation]

Fluency Score: [1-5]

Fluency Explanation: [Your explanation]

Relevance Score: [1-5]

Relevance Explanation: [Your explanation]

Engagement Score: [1-5]

Engagement Explanation: [Your explanation]

Overall Score: [Average score]

Overall Feedback: [Your feedback]

Comments on Columns:

[Your short paragraph commenting on each column]

Prompt:

{prompt}

```
**Generated Text:**
```

```
{generated_text}
```

```
---
```

A.3 Prompt Engineering for Summarisation

Prompt Generated by O1-preview (OpenAI, 2024) for GPT-4o to evaluate model's performance in Summarisation.

```
"""
```

As an expert evaluator, your task is to assess the quality of a generated summary based on the provided truncated input text. Please focus on the following criteria:

1. **Relevance**: Does the summary capture the main points of the truncated input?
2. **Conciseness**: Is the summary succinct without unnecessary details?
3. **Fluency**: Is the summary well-written with proper grammar and style?
4. **Accuracy**: Does the summary accurately represent the content of the truncated input without errors?
5. **Coherence**: Is the summary logically organized and easy to understand?

For each criterion, provide:

- **Score**: A number from 1 to 5 (where 1 is poor and 5 is excellent).
- **Explanation**: A brief justification for the score.

After evaluating each criterion, provide:

- **Overall Score**: The average of the five scores.
- **Overall Feedback**: A short overall feedback.

Additionally, provide a short paragraph commenting on the following columns:

- **Model Name**: {model_name}
- **Norm Type**: {norm_type}
- **Variant**: {variant}
- **Truncated Input**: [Truncated input is provided above.]
- **Generated Summary**: [Generated summary is provided above.]

Please present your evaluation in the following structured format:

```
```
```

Model Name: {model\_name}

Relevance Score: [1-5]

Relevance Explanation: [Your explanation]

Conciseness Score: [1-5]

Conciseness Explanation: [Your explanation]

Fluency Score: [1-5]

Fluency Explanation: [Your explanation]

Accuracy Score: [1-5]

Accuracy Explanation: [Your explanation]

Coherence Score: [1-5]

Coherence Explanation: [Your explanation]

Overall Score: [Average score]

Overall Feedback: [Your feedback]

Comments on Columns:

```
[Your short paragraph commenting on each column]
```

```

```

```

```

```
Truncated Input:
```

```
{truncated_input}
```

```

```

```
Generated Summary:
```

```
{generated_summary}
```

```

```

```
"""
```

## A.4 Prompt Engineering for Question Answering

Prompt Generated by O1-preview (OpenAI, 2024) for GPT-4o to evaluate model's performance in Question Answering.

```
"""
```

As an expert evaluator, your task is to assess the quality of an answer generated by a question-answering system based on the provided context. Please focus on the following criteria:

1. \*\*Correctness\*\*: Is the answer correct based on the context?
2. \*\*Completeness\*\*: Does the answer fully address the question?
3. \*\*Relevance\*\*: Is the answer relevant to the question and context?
4. \*\*Fluency\*\*: Is the answer well-written with proper grammar and style?
5. \*\*Conciseness\*\*: Is the answer concise and to the point?

For each criterion, provide:

- \*\*Score\*\*: A number from 1 to 5 (where 1 is poor and 5 is excellent).
- \*\*Explanation\*\*: A brief justification for the score.

After evaluating each criterion, provide:

- \*\*Overall Score\*\*: The average of the five scores.
- \*\*Overall Feedback\*\*: A short overall feedback.

\*\*Additionally\*\*, provide a short paragraph commenting on the following columns:

- \*\*Model Name\*\*: {model\_name}
- \*\*Norm Type\*\*: {norm\_type}
- \*\*Variant\*\*: {variant}
- \*\*Question\*\*: {question}
- \*\*Context\*\*: {context}
- \*\*Reference Answers\*\*: {reference\_answers}
- \*\*Generated Answer\*\*: {generated\_answer}

\*\*Please present your evaluation in the following structured format:\*\*

```

```

```
Model Name: {model_name}
```

```
Correctness Score: [1-5]
```

```
Correctness Explanation: [Your explanation]
```

```
Completeness Score: [1-5]
```

```
Completeness Explanation: [Your explanation]
```

```
Relevance Score: [1-5]
```

```
Relevance Explanation: [Your explanation]
```

Fluency Score: [1-5]  
Fluency Explanation: [Your explanation]

Conciseness Score: [1-5]  
Conciseness Explanation: [Your explanation]

Overall Score: [Average score]  
Overall Feedback: [Your feedback]

Comments on Columns:  
[Your short paragraph commenting on each column]  
```

Model Name:
{model_name}

Question:
{question}

Context:
{context}

Reference Answers:
{reference_answers}

Generated Answer:
{generated_answer}

```

## Appendix B - Links to Hugging Face Repository

### B.1 Pre-trained Models' Weights

This table lists the repositories of our pre-trained models.

| Normalisation Type | Link                                                                                                                                                                          |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PN                 | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_PN-124m/tree/denim-lake-75/log">https://huggingface.co/shng2025/GPT-Valkyrie_PN-124m/tree/denim-lake-75/log</a>         |
| LN                 | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m/tree/prime-snowball-26/log">https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m/tree/prime-snowball-26/log</a> |
| RMSN               | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m/tree/autumn-energy-2/log">https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m/tree/autumn-energy-2/log</a> |

### B.2 not fine-tuned Text-Generation Models' Weights

This table contains our non fine-tuned models for text generation, organised by normalisation type and model variant.

| Normalisation Type | Ablated Model Variant | Link                                                                                                                                                                      |
|--------------------|-----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| LN                 | AttnOnly              | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__AttnOnly__/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__AttnOnly__/tree/main</a>       |
| LN                 | baseModel             | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__baseModel__/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__baseModel__/tree/main</a>     |
| LN                 | FFOnly                | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__FFNonly__/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__FFNonly__/tree/main</a>         |
| LN                 | noNorm                | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__noNorm__/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__noNorm__/tree/main</a>           |
| RMSN               | AttnOnly              | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__AttnOnly__/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__AttnOnly__/tree/main</a>   |
| RMSN               | baseModel             | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__baseModel__/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__baseModel__/tree/main</a> |
| RMSN               | FFOnly                | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__FFNonly__/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__FFNonly__/tree/main</a>     |
| RMSN               | noNorm                | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__noNorm__/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__noNorm__/tree/main</a>       |

### B.3 Fine-tuned Summarisation (BillSum) Models' Weights

This table contains our models fine-tuned specifically on the BillSum dataset, organised by normalisation type and model variant.

| Normalisation Type | Ablated Model Variant | Link                                                                                                                                                                                    |
|--------------------|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| LN                 | AttnOnly              | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__AttnOnly__Billsum/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__AttnOnly__Billsum/tree/main</a>       |
| LN                 | baseModel             | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__baseModel__Billsum/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__baseModel__Billsum/tree/main</a>     |
| LN                 | FFOnly                | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__FFNonly__Billsum/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__FFNonly__Billsum/tree/main</a>         |
| LN                 | noNorm                | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__noNorm__Billsum/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__noNorm__Billsum/tree/main</a>           |
| RMSN               | AttnOnly              | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__AttnOnly__Billsum/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__AttnOnly__Billsum/tree/main</a>   |
| RMSN               | baseModel             | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__baseModel__Billsum/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__baseModel__Billsum/tree/main</a> |
| RMSN               | FFOnly                | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__FFNonly__Billsum/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__FFNonly__Billsum/tree/main</a>     |
| RMSN               | noNorm                | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__noNorm__Billsum/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__noNorm__Billsum/tree/main</a>       |

## B.4 Fine-tuned Question Answering (SQuAD) Models' Weights

This table contains our models fine-tuned specifically on the SQuAD dataset, organised by normalisation type and model variant.

| Normalisation Type | Ablated Model Variants | Link                                                                                                                                                                                |
|--------------------|------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| LN                 | AttnOnly               | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__AttnOnly__SQuAD/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__AttnOnly__SQuAD/tree/main</a>       |
| LN                 | baseModel              | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__baseModel__SQuAD/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__baseModel__SQuAD/tree/main</a>     |
| LN                 | FFOnly                 | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__FFNonly__SQuAD/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__FFNonly__SQuAD/tree/main</a>         |
| LN                 | noNorm                 | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__noNorm__SQuAD/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_LN-124m__noNorm__SQuAD/tree/main</a>           |
| RMSN               | AttnOnly               | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__AttnOnly__SQuAD/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__AttnOnly__SQuAD/tree/main</a>   |
| RMSN               | baseModel              | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__baseModel__SQuAD/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__baseModel__SQuAD/tree/main</a> |
| RMSN               | FFOnly                 | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__FFNonly__SQuAD/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__FFNonly__SQuAD/tree/main</a>     |
| RMSN               | noNorm                 | <a href="https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__noNorm__SQuAD/tree/main">https://huggingface.co/shng2025/GPT-Valkyrie_RMSN-124m__noNorm__SQuAD/tree/main</a>       |

## Appendix C - Graphing Code

This section contains all the code that we programmed with the assistance of Claude-3.5 (Anthropic, 2024) for the visualisations used in this paper. They are sorted by their respective figure numbers.

**Figure-5: Visualisation of GPT-2's self-attention using BertViz**

```
from transformers import AutoTokenizer, AutoModel, utils
from bertviz import model_view
utils.logging.set_verbosity_error() # Suppress standard warnings

model_name = "gpt2"
input_text = "The brown fox jumps"
model = AutoModel.from_pretrained(model_name, output_attentions=True) # Configure model to
return attention values
tokenizer = AutoTokenizer.from_pretrained(model_name)

inputs = tokenizer.encode(input_text, return_tensors='pt') # Tokenize input text
outputs = model(inputs) # Run model
attention = outputs[-1] # Retrieve attention from model outputs
tokens = tokenizer.convert_ids_to_tokens(inputs[0]) # Convert input ids to token strings
model_view(attention, tokens) # Display model view
```

**Figure-7: Impact of Internal Covariate Shift over Multiple Blocks Without Normalisation**

```
from transformers import AutoTokenizer, AutoModel, utils
from bertviz import model_view
utils.logging.set_verbosity_error() # Suppress standard warnings

model_name = "gpt2"
input_text = "The brown fox jumps"
model = AutoModel.from_pretrained(model_name, output_attentions=True) # Configure model to
return attention values
tokenizer = AutoTokenizer.from_pretrained(model_name)

inputs = tokenizer.encode(input_text, return_tensors='pt') # Tokenize input text
outputs = model(inputs) # Run model
attention = outputs[-1] # Retrieve attention from model outputs
tokens = tokenizer.convert_ids_to_tokens(inputs[0]) # Convert input ids to token strings
model_view(attention, tokens) # Display model view
```

**Figure-8: Stable Input Distributions when Layer Normalisation is Applied**

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

Parameters for the original distribution
mean_original = 0
variance_original = 1
std_original = np.sqrt(variance_original)

Shifts and scaling constants (these won't affect the final distribution due to
```

```

normalization)
mean_shift = 1.5
variance_scale = 1.2

X values for plotting
x = np.linspace(-5, 5, 1000)

Plotting setup
plt.figure(figsize=(14, 8))
plt.plot(x, norm.pdf(x, mean_original, std_original), label='Original (Mean=0, Var=1)', linestyle='dashed')

Loop from 0 to 11 blocks to simulate the changes with layer normalization
for i in range(1, 12):
 # The distribution remains standard normal due to layer normalization
 y = norm.pdf(x, 0, 1)
 plt.plot(x, y, label=f'Input Distribution after {i} Blocks (Normalized)', alpha=0.75 - i * 0.02)

Adding labels and title
plt.title('Impact of Layer Normalization on Input Distributions over Multiple Blocks')
plt.xlabel('Value')
plt.ylabel('Probability Density')
plt.xlim(-5, 5) # Set the x-axis to show from -5 to 5
plt.legend()
plt.grid(True)
plt.show()

```

**Figure-9: Input Distributions's Mean Drifts when RMSN is Applied**

```

import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

Parameters for the original distribution
mean_original = 0
variance_original = 1
std_original = np.sqrt(variance_original)

Shifts and scaling constants
mean_shift = 1.5
variance_scale = 1.2

X values for plotting
x = np.linspace(-10, 30, 1000)

Plotting setup
plt.figure(figsize=(16, 8))
plt.plot(x, norm.pdf(x, mean_original, std_original), label='Original (Mean=0, Var=1)', linestyle='dashed')

Loop from 0 to 11 blocks to simulate the changes with RMS normalization
for i in range(1, 12):

```

```

Calculate the new mean (which is not centered in RMS normalization)
mean = mean_original + i * mean_shift

The standard deviation is normalized to 1 in RMS normalization
std = 1

y = norm.pdf(x, mean, std)
plt.plot(x, y, label=f'Input Distribution after {i} Blocks (Mean={mean:.2f}, Std=1)', alpha=0.75 - i * 0.02)

Adding labels and title
plt.title('Impact of RMS Normalization on Input Distribution over Multiple Blocks')
plt.xlabel('Value')
plt.ylabel('Probability Density')
plt.xlim(-5, 30) # Set the x-axis to show from -10 to 30

Add legend with smaller font size
plt.legend(fontsize=10) # You can adjust this number to make it smaller or larger

plt.grid(True)
plt.show()

```

**Figure-11: Input Distributions when Power Normalisation is Applied**

```

import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

Parameters
mean_original = 0
variance_original = 1
std_original = np.sqrt(variance_original)

mean_shift = 1.5
variance_scale = 1.2

Power Normalization parameters
alpha = 0.9 # Higher value to show more of the running statistics effect

X values for plotting
x = np.linspace(-10, 30, 1000)

Plotting setup
plt.figure(figsize=(12, 8))
plt.plot(x, norm.pdf(x, mean_original, std_original), label='Original (Mean=0, Var=1)', linestyle='dashed')

Initialize running statistics
running_mean = mean_original
running_var = variance_original

Loop through blocks
for i in range(1, 12):

```

```

Simulate batch statistics
batch_mean = mean_original + i * mean_shift
batch_var = variance_original * (variance_scale ** i)

Update running statistics
running_mean = alpha * running_mean + (1 - alpha) * batch_mean
running_var = alpha * running_var + (1 - alpha) * batch_var

Power Normalization
Normalize using batch statistics, but scale using running statistics
normalized_mean = 0 + np.random.normal(0, 0.1) # Small random deviation from 0
normalized_std = np.sqrt(running_var) / np.sqrt(batch_var) # Scaling factor

y = norm.pdf(x, normalized_mean, normalized_std)
plt.plot(x, y, label=f'Input Distribution after {i} Blocks (Mean≈{normalized_mean:.2f}, Std={normalized_std:.2f})', alpha=0.75 - i * 0.02)

Adding labels and title
plt.title('Impact of Power Normalization on Input Distributions over Multiple Blocks')
plt.xlabel('Value')
plt.ylabel('Probability Density')
plt.xlim(-4, 8) # Adjusted to focus on the normalized range
plt.legend()
plt.grid(True)
plt.show()

```

**Figure-14: Tokens processing speed when pre-training LN of different H100 configurations**

```

import matplotlib.pyplot as plt
import numpy as np

Set figure size and style
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(15, 5))

Data
gpus = ['1 GPU', '4 GPUs', '6 GPUs']
total_tokens = [400000, 1270000, 1150000]
per_gpu_tokens = [400000, 317500, 191700]

First plot - Total Tokens/Second
bars1 = ax1.bar(gpus, total_tokens, color="#90EE90") # Light green color
ax1.set_title('Total tokens/sec processed during training for Different GPU Configurations', pad=20)
ax1.set_ylimits(0, 1500000)
ax1.set_ylabel('tokens/sec')

Format y-axis ticks for first plot
ax1.yaxis.set_major_formatter(lambda x, p: f'{x/1e6:.1f}M' if x >= 1e6 else f'{int(x/1e3})K')

Add value labels for first plot
for i, v in enumerate(total_tokens):
 label = f'{v/1e6:.2f}M' if v >= 1e6 else f'{int(v/1e3})K'
 ax1.text(i, v + 50000, label, ha='center', va='bottom')

Second plot - Per GPU Tokens
bars2 = ax2.bar(gpus, per_gpu_tokens, color="#FFCCBC") # Orange color
ax2.set_title('Tokens processed by individual GPU for Different GPU Configurations', pad=20)
ax2.set_ylimits(0, 1500000)
ax2.set_ylabel('tokens/sec')

Format y-axis ticks for second plot
ax2.yaxis.set_major_formatter(lambda x, p: f'{x/1e6:.1f}M' if x >= 1e6 else f'{int(x/1e3})K')

Add value labels for second plot
for i, v in enumerate(per_gpu_tokens):
 label = f'{v/1e6:.2f}M' if v >= 1e6 else f'{int(v/1e3})K'
 ax2.text(i, v + 50000, label, ha='center', va='bottom')

```

```

Second plot - Tokens/Second per GPU
bars2 = ax2.bar(gpuids, per_gpu_tokens, color='#4B9AFA') # Blue color
ax2.set_title('tokens/sec processed during training per GPU for Different Configurations', pad=20)
ax2.set_ylimit(0, 500000)
ax2.set_ylabel('tokens/sec per GPU')

Format y-axis ticks for second plot
ax2.yaxis.set_major_formatter(lambda x, p: f'{int(x/1e3)}K')

Add value labels for second plot
for i, v in enumerate(per_gpu_tokens):
 label = f'{v/1e3:.1f}K'
 ax2.text(i, v + 15000, label, ha='center', va='bottom')

Adjust layout and spacing
plt.tight_layout()

Display the plot
plt.show()

```

**Figure-24: Gradient distribution across ablated model variants' 12 layers before fine-tuning**

```

import torch
from transformers import AutoModelForCausalLM, AutoTokenizer
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from matplotlib.colors import LogNorm
import pandas as pd
import os

prompts = [
 "The spaceship landed softly on the alien planet, and the hatch began to open.",
 "In a quiet village, a cat with mysterious powers arrived one rainy evening.",
 "The treasure map led them to a cave that had never been discovered before.",
 "As the clock struck midnight, the mirror began to shimmer.",
 "The forest was said to be cursed, but she ventured inside anyway.",
 "A mysterious letter arrived in the mail, addressed to someone who no longer lived here.",
 "At the bottom of the ocean, they found a door that led to another world.",
 "The robot slowly gained emotions, surprising its creators.",
 "Every time he fell asleep, he would wake up in a new timeline.",
 "They say that whoever solves the riddle of the ancient book gains unlimited knowledge.",
 "The day the stars disappeared, humanity knew something was coming.",
 "A portal opened in the backyard, leading to a dimension of pure imagination.",
 "The detective knew this was no ordinary case when the painting began to move.",
 "On her 18th birthday, she discovered she could control the wind.",
 "The world froze in time, except for one boy.",
 "The enchanted necklace began to glow whenever danger was near.",
 "He had only one chance to change the past, and the clock was ticking.",
 "The library was full of books no one had ever read before, and each held a hidden

```

```

secret.",

 "A voice from the radio gave her instructions for an impossible mission.",

 "The moment she touched the ancient artifact, visions flooded her mind.",

 "In the attic, they found a diary that told of future events.",

 "A dragon had been asleep under the city for centuries, and now it was waking up.",

 "Each time he looked in the mirror, a different person stared back.",

 "A storm brought something unexpected to the small town.",

 "Her drawings started coming to life one by one.",

 "The message in the bottle warned of a great danger.",

 "The town vanished overnight, leaving only one survivor.",

 "He discovered that every wish he made came with a price.",

 "A snowstorm trapped them in the cabin, but they weren't alone.",

 "The clock in the town square began running backward.",

 "As the comet passed, everyone suddenly gained strange abilities.",

 "The bridge appeared only under a full moon.",

 "A stray cat followed him everywhere, but it wasn't an ordinary cat.",

 "The carnival arrived without warning, and so did the strange occurrences.",

 "In his dreams, a mysterious figure kept warning him about the future.",

 "Every door in the house led to a different world.",

 "The garden bloomed with flowers no one had ever seen before.",

 "A mysterious shadow followed her everywhere she went.",

 "The phone rang, and the voice on the other end knew too much.",

 "The clockmaker's new invention could turn back time.",

 "The lighthouse was abandoned, but its light still shone every night.",

 "Every star in the night sky represented a different soul.",

 "The forest whispered secrets to those who listened closely.",

 "The old camera developed pictures from events that hadn't happened yet.",

 "A message appeared on her phone, and it was from the future.",

 "The ghost offered a deal too good to be true.",

 "A library card granted access to books that didn't exist in the real world.",

 "As soon as they touched the stone, they were transported to another realm.",

 "Each song on the old record player told a different person's life story.",

 "They found a map that changed every time they looked at it."
]

```

```

def save_gradients_to_csv(all_layer_averages, model_variants, task="textgen",
output_dir="gradient_results/averages"):

 """Save averaged gradient values to CSV file."""

 # Create output directory if it doesn't exist

 os.makedirs(output_dir, exist_ok=True)

 # Calculate average across all prompts

 avg_gradients = np.mean(all_layer_averages, axis=1) # Shape: (num_models, num_layers)

 # Convert to DataFrame

 df = pd.DataFrame(avg_gradients, index=model_variants)

 df.columns = [f"layer_{i}" for i in range(df.shape[1])]

 # Save to CSV

 filename = os.path.join(output_dir, f"{task}_gradients.csv")

 df.to_csv(filename)

 print(f"Saved averaged gradient values to {filename}")

 return filename

```

```

def load_gradients_from_csv(task="textgen", input_dir="gradient_results/averages"):
 """Load gradient values from CSV file."""
 filename = os.path.join(input_dir, f"{task}_gradients.csv")

 if not os.path.exists(filename):
 raise FileNotFoundError(f"No gradient file found: {filename}")

 df = pd.read_csv(filename, index_col=0)
 return df.values, df.index.tolist()

def visualize_layer_gradients(all_layer_averages, model_variants, prompts, vmin=None,
vmax=None, use_log_scale=True):
 """Visualize gradient magnitudes across layers and models."""
 # Calculate average across all prompts for each model and layer
 averaged_gradients = np.mean(all_layer_averages, axis=1)

 num_layers = averaged_gradients.shape[1]
 num_models = len(model_variants)

 # Create figure with specific aspect ratio for square cells
 plt.figure(figsize=(12, 12))

 # Set scale and limits
 if vmin is None:
 vmin = max(np.min(averaged_gradients), 1e-10)
 if vmax is None:
 vmax = np.max(averaged_gradients)

 # Create heatmap without annotations
 heatmap_kwargs = {
 'data': averaged_gradients.T, # No need to flip if using proper y-axis inversion
 'cmap': "YlGnBu",
 'annot': False,
 'square': True,
 }

 if use_log_scale:
 heatmap_kwargs.update({
 'norm': LogNorm(vmin=vmin, vmax=vmax),
 'cbar_kws': {
 'label': 'Gradient Magnitude',
 'format': '%.2e',
 'ticks': np.logspace(np.log10(vmin), np.log10(vmax), 10)
 }
 })
 else:
 heatmap_kwargs.update({
 'vmin': vmin,
 'vmax': vmax,
 'cbar_kws': {
 'label': 'Gradient Magnitude',
 'format': '%.2f'
 }
 })

```

```

 }

 })

ax = sns.heatmap(**heatmap_kwargs)

Format x-axis labels using the improved label extraction
x_labels = [get_model_label(model) for model in model_variants]
plt.xticks(np.arange(len(model_variants)) + 0.5,
 x_labels,
 rotation=45,
 ha='right')

Format y-axis with proper layer numbering (0 to num_layers-1)
plt.yticks(np.arange(num_layers) + 0.5,
 range(num_layers))

Invert y-axis to have 0 at bottom
ax.invert_yaxis()

plt.xlabel("Model Variant")
plt.ylabel("Layer")
scale_type = "Log" if use_log_scale else "Linear"
plt.title(f"Average Layer Gradient Magnitudes Across Models ({scale_type} Scale)")

plt.tight_layout()
plt.show()

def run_gradient_analysis(model_variants, prompts, task="textgen", force_recalculate=False):
 """Run gradient analysis with CSV caching."""
 # Try to load from CSV first if not forcing recalculation
 if not force_recalculate:
 try:
 averaged_gradients, loaded_models = load_gradients_from_csv(task)
 print(f"Loaded cached gradients for {task}")
 visualize_layer_gradients(averaged_gradients[np.newaxis, :, :], model_variants,
 prompts)
 except:
 return
 except FileNotFoundError:
 print(f"No cached gradients found for {task}, calculating...")

 device = get_device()
 print(f"Using device: {device}")

 all_layer_averages = []

 try:
 for model_name in model_variants:
 print(f"Processing {model_name}")
 model_averages = []

 for i, prompt in enumerate(prompts):
 print(f" Processing prompt {i+1}/{len(prompts)}")
 gradients, model = calculate_gradient_magnitudes(model_name, prompt, device)

```

```

layer_averages = calculate_layer_averages(model, gradients)
model_averages.append(layer_averages)

Clear CUDA cache after each prompt
del model
torch.cuda.empty_cache()

all_layer_averages.append(model_averages)

Convert to numpy array
all_layer_averages = np.array(all_layer_averages)

Save results to CSV
save_gradients_to_csv(all_layer_averages, model_variants, task)

Visualize results
visualize_layer_gradients(all_layer_averages, model_variants, prompts)

except Exception as e:
 print(f"Error during analysis: {str(e)}")
 raise
finally:
 # Final cleanup
 torch.cuda.empty_cache()

if __name__ == "__main__":
 model_variants = [
 "shng2025/GPT-Valkyrie_LN-124m_noNorm",
 "shng2025/GPT-Valkyrie_LN-124m_AttnOnly",
 "shng2025/GPT-Valkyrie_LN-124m_FFNonly",
 "shng2025/GPT-Valkyrie_LN-124m_baseModel",
 "shng2025/GPT-Valkyrie_RMSN-124m_noNorm",
 "shng2025/GPT-Valkyrie_RMSN-124m_AttnOnly",
 "shng2025/GPT-Valkyrie_RMSN-124m_FFNonly",
 "shng2025/GPT-Valkyrie_RMSN-124m_baseModel",
]
 # Set force_recalculate=True to ignore cached results and recalculate
 run_gradient_analysis(model_variants, prompts, task="textgen", force_recalculate=True)

```

**Figure-25: Gradient distribution after BillSum fine-tuning**

&

**Figure-26: Gradient distribution after SQuAD fine-tuning**

```

import pandas as pd
import os

def save_gradients_to_csv(all_layer_averages, model_variants, task_type,
output_dir="gradient_results"):
 """Save gradient values to CSV file."""
 # Create output directory if it doesn't exist
 os.makedirs(output_dir, exist_ok=True)

```

```

Convert to DataFrame
df = pd.DataFrame(all_layer_averages, index=model_variants)

Add column names for layers
df.columns = [f"layer_{i}" for i in range(df.shape[1])]

Save to CSV
filename = os.path.join(output_dir, f"{task_type}_gradients.csv")
df.to_csv(filename)
print(f"Saved gradient values to {filename}")
return filename

def load_gradients_from_csv(task_type, input_dir="gradient_results"):
 """Load gradient values from CSV file."""
 filename = os.path.join(input_dir, f"{task_type}_gradients.csv")
 if not os.path.exists(filename):
 raise FileNotFoundError(f"No gradient file found for {task_type}")

 df = pd.read_csv(filename, index_col=0)
 return df.values.tolist(), df.index.tolist()

def visualize_layer_gradients(all_layer_averages, model_variants, task_type="", vmin=None,
vmax=None, use_log_scale=True):
 """Visualize gradient magnitudes across layers and models."""
 # Handle different input formats
 if isinstance(all_layer_averages, list):
 # Convert list to numpy array
 all_layer_averages = np.array(all_layer_averages)

 # Handle different dimensionalities
 if all_layer_averages.ndim == 3: # For text generation (models, prompts, layers)
 averaged_gradients = np.mean(all_layer_averages, axis=1)
 elif all_layer_averages.ndim == 2: # For Squad/Billsum (models, layers)
 averaged_gradients = all_layer_averages
 else:
 raise ValueError(f"Unexpected gradient array shape: {all_layer_averages.shape}")

 num_layers = averaged_gradients.shape[1]
 num_models = len(model_variants)

 # Create figure with specific aspect ratio for square cells
 plt.figure(figsize=(12, 12))

 # Set scale and limits
 if vmin is None:
 vmin = max(np.min(averaged_gradients), 1e-10)
 if vmax is None:
 vmax = np.max(averaged_gradients)

 # Create heatmap without annotations
 heatmap_kwargs = {
 'data': averaged_gradients.T,

```

```

 'cmap': "YlGnBu",
 'annot': False,
 'square': True,
 }

 if use_log_scale:
 heatmap_kwargs.update({
 'norm': LogNorm(vmin=vmin, vmax=vmax),
 'cbar_kws': {
 'label': 'Gradient Magnitude',
 'format': '%.2e',
 'ticks': np.logspace(np.log10(vmin), np.log10(vmax), 10)
 }
 })
 else:
 heatmap_kwargs.update({
 'vmin': vmin,
 'vmax': vmax,
 'cbar_kws': {
 'label': 'Gradient Magnitude',
 'format': '%.2f'
 }
 })

 ax = sns.heatmap(**heatmap_kwargs)

 # Format x-axis labels using the improved label extraction
 x_labels = [get_model_label(model) for model in model_variants]
 plt.xticks(np.arange(len(model_variants)) + 0.5,
 x_labels,
 rotation=45,
 ha='right')

 # Format y-axis with proper layer numbering (0 to num_layers-1)
 plt.yticks(np.arange(num_layers) + 0.5,
 range(num_layers))

 # Invert y-axis to have 0 at bottom
 ax.invert_yaxis()

 plt.xlabel("Model Variant")
 plt.ylabel("Layer")
 scale_type = "Log" if use_log_scale else "Linear"
 plt.title(f"Average Layer Gradient Magnitudes Across Models ({task_type}) ({scale_type} Scale)")

 plt.tight_layout()
 plt.show()

def run_analysis(model_variants, dataset, calc_func, task_type="", num_samples=50,
 force_recalculate=False):
 """Run gradient analysis with CSV caching."""
 # Try to load from CSV first if not forcing recalculation

```

```

if not force_recalculate:
 try:
 averaged_gradients, loaded_models = load_gradients_from_csv(task_type)
 print(f"Loaded cached gradients for {task_type}")
 visualize_layer_gradients(averaged_gradients, model_variants, task_type)
 return
 except FileNotFoundError:
 print(f"No cached gradients found for {task_type}, calculating...")

Select first num_samples examples
if len(dataset) > num_samples:
 dataset = dataset.select(range(num_samples))

all_layer_averages = []

for model_name in model_variants:
 print(f"Processing {model_name}")
 model_averages = []

 for i, example in enumerate(dataset):
 print(f" Processing example {i+1}/{len(dataset)}")
 gradients, model = calc_func(model_name, example)
 layer_averages = calculate_layer_averages(model, gradients)
 model_averages.append(layer_averages)

 del model
 torch.cuda.empty_cache()

 # Average across examples for this model
 avg_layer_gradients = np.mean(model_averages, axis=0)
 all_layer_averages.append(avg_layer_gradients)

Save results to CSV
save_gradients_to_csv(np.array(all_layer_averages), model_variants, task_type)

Visualize results
visualize_layer_gradients(all_layer_averages, model_variants, task_type)

def run_squad_analysis(force_recalculate=False):
 """Run SQuAD analysis with optional force recalculation."""
 print("Loading SQuAD dataset...")
 dataset = load_dataset("squad", split="validation")

 model_variants = [
 "shng2025/GPT-Valkyrie_LN-124m_noNorm_SQuAD",
 "shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD",
 "shng2025/GPT-Valkyrie_LN-124m_FFNonly_SQuAD",
 "shng2025/GPT-Valkyrie_LN-124m_baseModel_SQuAD",
 "shng2025/GPT-Valkyrie_RMSN-124m_noNorm_SQuAD",
 "shng2025/GPT-Valkyrie_RMSN-124m_AttnOnly_SQuAD",
 "shng2025/GPT-Valkyrie_RMSN-124m_FFNonly_SQuAD",
 "shng2025/GPT-Valkyrie_RMSN-124m_baseModel_SQuAD",
]

```

```

run_analysis(model_variants, dataset, calculate_gradient_magnitudes_squad,
 "SQuAD", force_recalculate=force_recalculate)

def run_billsum_analysis(force_recalculate=False):
 """Run Billsum analysis with optional force recalculation."""
 print("Loading Billsum dataset...")
 dataset = load_dataset("billsum", split="ca_test")

 model_variants = [
 "shng2025/GPT-Valkyrie_LN-124m_noNorm_Billsum",
 "shng2025/GPT-Valkyrie_LN-124m_AttnOnly_Billsum",
 "shng2025/GPT-Valkyrie_LN-124m_FFNonly_Billsum",
 "shng2025/GPT-Valkyrie_LN-124m_baseModel_Billsum",
 "shng2025/GPT-Valkyrie_RMSN-124m_noNorm_Billsum",
 "shng2025/GPT-Valkyrie_RMSN-124m_AttnOnly_Billsum",
 "shng2025/GPT-Valkyrie_RMSN-124m_FFNonly_Billsum",
 "shng2025/GPT-Valkyrie_RMSN-124m_baseModel_Billsum",
]
 run_analysis(model_variants, dataset, calculate_gradient_magnitudes_billsum,
 "Billsum", force_recalculate=force_recalculate)

if __name__ == "__main__":
 # Run analyses
 # Set force_recalculate=True to ignore cached results and recalculate
 run_squad_analysis(force_recalculate=True)
 run_billsum_analysis(force_recalculate=True)

```

**Figure-27: Comparison of models' average performance across GPT-4o's evaluation metrics &**

**Figure-28: Violin plots of models' score distributions across GPT-4o evaluation metrics**

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kruskal
from scikit_posthocs import posthoc_dunn
import warnings
import os

warnings.filterwarnings('ignore')

Create a directory for outputs
output_dir = 'text_generation_analysis_results'
os.makedirs(output_dir, exist_ok=True)

Load the data
df = pd.read_csv('combined_data_truncated.csv')

Ensure 'model_name' is treated as a categorical variable
df['model_name'] = df['model_name'].astype('category')

```

```

Simplify model names
def simplify_model_name(name):
 if '_' in name:
 return name.replace('_', '-')
 return name

df['model_name_simple'] = df['model_name'].apply(simplify_model_name)

Define score columns
score_columns = ['Overall Score', 'Creativity Score', 'Coherence Score', 'Fluency Score',
'Relevance Score', 'Engagement Score']

1. Effect Size Calculation (Eta-squared)
def calculate_effect_size(df, metric):
 try:
 h_statistic, _ = kruskal(*[group[metric].values for name, group in
df.groupby('model_name')])
 n = len(df)
 eta_squared = (h_statistic - len(df['model_name'].unique()) + 1) / (n - 1)
 return max(0, eta_squared) # Ensure non-negative value
 except Exception as e:
 print(f"Error calculating effect size for {metric}: {e}")
 return np.nan

effect_sizes = {metric: calculate_effect_size(df, metric) for metric in score_columns}
effect_sizes_df = pd.DataFrame.from_dict(effect_sizes, orient='index', columns=['Effect Size
(\eta^2)'])
effect_sizes_df.to_csv(os.path.join(output_dir, 'effect_sizes.csv'))

2. Non-parametric Tests (Kruskal-Wallis and Dunn's test)
def perform_kruskal_dunn(df, metric):
 try:
 # Kruskal-Wallis test
 kruskal_result = kruskal(*[group[metric].values for name, group in
df.groupby('model_name')])

 # Dunn's test
 dunn_result = posthoc_dunn(df, val_col=metric, group_col='model_name',
p_adjust='bonferroni')

 # Calculate effect size for Dunn's test
 dunn_effect_size = calculate_effect_size(df, metric)

 return {
 'kruskal': pd.DataFrame({'statistic': [kruskal_result.statistic],
 'p-value': [kruskal_result.pvalue]}),
 'dunn_pval': dunn_result,
 'dunn_effect': dunn_effect_size
 }
 except Exception as e:
 print(f"Error performing Kruskal-Wallis and Dunn's test for {metric}: {e}")
 return None

```

```

test_results = {metric: perform_kruskal_dunn(df, metric) for metric in score_columns}

Save results
for metric, results in test_results.items():
 if results is not None:
 results['kruskal'].to_csv(os.path.join(output_dir, f'kruskal_wallis_{metric.replace(" ", "_")}.csv'),
 index=False)
 results['dunn_pval'].to_csv(os.path.join(output_dir, f'dunn_test_pvalues_{metric.replace(" ", "_")}.csv'))
 pd.DataFrame({'Effect Size (η^2)': [results['dunn_effect']]}) .to_csv(
 os.path.join(output_dir, f'dunn_test_effect_size_{metric.replace(" ", "_")}.csv'))

3. Basic Descriptive Statistics
descriptive_stats = df.groupby('model_name_simple')[score_columns].agg(['mean', 'std', 'min', 'max'])
descriptive_stats.to_csv(os.path.join(output_dir, 'descriptive_statistics.csv'))

4. Visualizations
Violin Plots
plt.figure(figsize=(20, 15))
for i, metric in enumerate(score_columns, 1):
 plt.subplot(2, 3, i)
 sns.violinplot(x='model_name_simple', y=metric, data=df)
 plt.title(metric)
 plt.xticks(rotation=45, ha='right')
 plt.xlabel('')
 plt.ylim(1, 5)
 if i % 3 != 1:
 plt.ylabel('')
plt.tight_layout()
plt.savefig(os.path.join(output_dir, 'violin_plots.png'), bbox_inches='tight')
plt.close()

Radar Chart
def radar_chart(df, metrics):
 means = df.groupby('model_name_simple')[metrics].mean()
 angles = np.linspace(0, 2*np.pi, len(metrics), endpoint=False)
 means = pd.concat([means, means.iloc[:, :1]], axis=1)
 angles = np.concatenate((angles, [angles[0]]))

 fig, ax = plt.subplots(figsize=(14, 10), subplot_kw=dict(projection='polar'))
 for model in means.index:
 values = means.loc[model].values
 ax.plot(angles, values, 'o-', linewidth=2, label=model)
 ax.fill(angles, values, alpha=0.25)
 ax.set_thetagrids(angles[:-1] * 180/np.pi, metrics)
 ax.set_yticks([1, 5]) # Assuming scores are between 0 and 5
 plt.legend(loc='center left', bbox_to_anchor=(1.1, 0.5))
 plt.title("Model Performance Across Text Generation Metrics")
 plt.tight_layout()
 plt.savefig(os.path.join(output_dir, 'radar_chart.png'), bbox_inches='tight')

```

```

plt.close()

radar_chart(df, score_columns)

print(f"\nAnalysis complete. Results saved in the '{output_dir}' directory.")
print("Check the generated PNG files for visualizations.")

```

**Figure-31: Violin plots of models' score distributions across BillSum traditional metrics**

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from statsmodels.stats.multicomp import pairwise_tukeyhsd
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from scipy.stats import kruskal
from scikit_posthocs import posthoc_dunn
import warnings
import os

warnings.filterwarnings('ignore')

def clean_model_name(name):
 """Convert full model name to concise format."""
 try:
 parts = name.split('_')
 norm_type = parts[1].split('-')[0] # Get LN or RMSN
 variant = parts[3].split('__')[0] # Get modelBuilder, noNorm, AttnOnly, or FFNonly
 return f"{norm_type}-{variant}"
 except:
 return name

def perform_kruskal_dunn(df, metric):
 """Perform Kruskal-Wallis and Dunn's test for a metric."""
 try:
 # Check if all values are identical
 if df[metric].nunique() == 1:
 print(f"Note: All {metric} values are identical ({df[metric].iloc[0]}), skipping Kruskal-Wallis test.")
 return pd.DataFrame({
 'statistic': [0],
 'p-value': [1.0],
 'note': ['All values identical']
 }), None
 kruskal_result = kruskal(*[group[metric].values for name, group in df.groupby('model_name')])
 if kruskal_result.pvalue < 0.05:
 dunn_result = posthoc_dunn(df, val_col=metric, group_col='model_name',
p_adjust='bonferroni')
 return pd.DataFrame({'statistic': [kruskal_result.statistic], 'p-value':

```

```

[kruskal_result.pvalue]}), dunn_result
 return pd.DataFrame({'statistic': [kruskal_result.statistic], 'p-value':
[kruskal_result.pvalue]}), None
except Exception as e:
 print(f"Error performing Kruskal-Wallis and Dunn's test for {metric}: {e}")
 return None, None

def create_radar_chart(df, metrics, condition_name, output_dir):
 """Create radar chart with specific formatting."""
 means = df.groupby('model_name')[metrics].mean()
 angles = np.linspace(0, 2*np.pi, len(metrics), endpoint=False)
 means = pd.concat([means, means.iloc[:, :1]], axis=1)
 angles = np.concatenate((angles, [angles[0]]))

 fig, ax = plt.subplots(figsize=(14, 10), subplot_kw=dict(projection='polar'))
 for model in means.index:
 values = means.loc[model].values
 ax.plot(angles, values, 'o-', linewidth=2, label=model)
 ax.fill(angles, values, alpha=0.25)
 ax.set_thetagrids(angles[:-1] * 180/np.pi, metrics)
 ax.set_ylim(0, max(means.max().max(), 1.0))
 plt.legend(loc='center left', bbox_to_anchor=(1.1, 0.5))
 plt.title(f"Model Performance Across Metrics - {condition_name}")
 plt.tight_layout()
 plt.savefig(os.path.join(output_dir, 'radar_chart.png'), bbox_inches='tight')
 plt.close()

def calculate_anova_summary(df, metrics):
 """Calculate ANOVA summary for each metric."""
 anova_results = []

 for metric in metrics:
 groups = [group[metric].values for name, group in df.groupby('model_name')]
 f_value, p_value = stats.f_oneway(*groups)

 eta_squared = f_value / (f_value + df.groupby('model_name').size().iloc[0] - 1)

 anova_results.append({
 'Metric': metric,
 'F_value': f_value,
 'p_value': p_value,
 'eta_squared': eta_squared
 })

 return pd.DataFrame(anova_results)

def create_analysis_for_dataset(df, dataset_name, output_dir):
 """Perform statistical analysis for the dataset."""
 # Define metrics for detailed analysis
 score_metrics = ['rouge1_score', 'rouge2_score', 'rougeL_score', 'bleu_score']
 generation_metrics = ['sequence_score', 'mean_token_probability', 'generation_length']
 all_metrics = score_metrics + generation_metrics

```

```

os.makedirs(output_dir, exist_ok=True)

1. Basic Descriptive Statistics
desc_stats = df.groupby('model_name')[all_metrics].agg(['mean', 'std', 'min', 'max',
'count'])
desc_stats.to_csv(os.path.join(output_dir, 'descriptive_statistics.csv'))

2. Effect Size Calculation
effect_sizes = {}
for metric in all_metrics:
 try:
 f_value, _ = stats.f_oneway(*[group[metric] for name, group in
df.groupby('model_name')])
 effect_sizes[metric] = f_value / (f_value +
df.groupby('model_name').size().iloc[0] - 1)
 except Exception as e:
 print(f"Error calculating effect size for {metric}: {e}")
 effect_sizes[metric] = np.nan

effect_sizes_df = pd.DataFrame.from_dict(effect_sizes, orient='index', columns=['Effect
Size'])
effect_sizes_df.to_csv(os.path.join(output_dir, 'effect_sizes.csv'))

3. ANOVA summary
anova_summary = calculate_anova_summary(df, all_metrics)
anova_summary.to_csv(os.path.join(output_dir, 'anova_summary.csv'), index=False)

4. Score distribution summary
score_dist_summary = df[all_metrics].agg([
 'mean', 'std', 'min', 'max',
 lambda x: x.quantile(0.25),
 lambda x: x.quantile(0.75),
 'skew', 'kurt'
]).round(4)

score_dist_summary.index = ['Mean', 'Std Dev', 'Min', 'Max', 'Q1', 'Q3', 'Skewness',
'Kurtosis']
score_dist_summary.to_csv(os.path.join(output_dir, 'score_distribution_summary.csv'))

5. Post-hoc Tests
for metric in all_metrics:
 tukey = pairwise_tukeyhsd(df[metric], df['model_name'])
 pd.DataFrame(data=tukey._results_table.data[1:],

 columns=tukey._results_table.data[0]).to_csv(
 os.path.join(output_dir, f'tukey_hsd_{metric}.csv'), index=False)

6. Correlation Analysis
correlation_matrix = df[all_metrics].corr()
correlation_matrix.to_csv(os.path.join(output_dir, 'correlation_matrix.csv'))

7. PCA Analysis
scaler = StandardScaler()
pca = PCA()

```

```

pca_result = pca.fit_transform(scaler.fit_transform(df[score_metrics]))
pca_df = pd.DataFrame({
 'Principal Component': range(1, len(pca.explained_variance_ratio_) + 1),
 'Explained Variance Ratio': pca.explained_variance_ratio_,
 'Cumulative Explained Variance Ratio': np.cumsum(pca.explained_variance_ratio_)
})
pca_df.to_csv(os.path.join(output_dir, 'pca_results.csv'), index=False)

8. Kruskal-Wallis and Dunn's Tests
for metric in all_metrics:
 kruskal_result, dunn_result = perform_kruskal_dunn(df, metric)
 if kruskal_result is not None:
 kruskal_result.to_csv(os.path.join(output_dir, f'kruskal_{metric}.csv'),
index=False)
 if dunn_result is not None:
 dunn_result.to_csv(os.path.join(output_dir, f'dunn_test_{metric}.csv'))

9. Visualizations

Correlation Heatmap
plt.figure(figsize=(12, 10))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title(f'Correlation Heatmap - {dataset_name}')
plt.tight_layout()
plt.savefig(os.path.join(output_dir, 'correlation_heatmap.png'))
plt.close()

Violin Plots - Separate for score metrics and generation metrics
for metrics_group, group_name in [(score_metrics, 'score_metrics'),
 (generation_metrics, 'generation_metrics')]:
 plt.figure(figsize=(20, 10))
 for i, metric in enumerate(metrics_group, 1):
 plt.subplot(1, len(metrics_group), i)
 sns.violinplot(x='model_name', y=metric, data=df)
 plt.title(f'{metric}')
 plt.xticks(rotation=45, ha='right')
 plt.ylim(0, df[metric].max() * 1.1) # Set y-axis limits with 10% padding
 if i != 1:
 plt.ylabel('')
 plt.tight_layout()
 plt.savefig(os.path.join(output_dir, f'violin_plots_{group_name}.png'),
bbox_inches='tight')
 plt.close()

Distribution Plots - Separate for score metrics and generation metrics
for metrics_group, group_name in [(score_metrics, 'score_metrics'),
 (generation_metrics, 'generation_metrics')]:
 plt.figure(figsize=(20, 10))
 for i, metric in enumerate(metrics_group, 1):
 plt.subplot(1, len(metrics_group), i)
 for model in df['model_name'].unique():
 sns.kdeplot(data=df[df['model_name'] == model], x=metric, label=model)
 plt.title(f'Distribution of {metric}')

```

```

 plt.xlabel('Value')
 plt.ylabel('Density')
 if i == len(metrics_group): # Place legend for last plot
 plt.legend(title='Model Name', bbox_to_anchor=(1.05, 1), loc='upper left')
 else:
 plt.legend([])
 plt.tight_layout()
 plt.savefig(os.path.join(output_dir, f'distribution_plots_{group_name}.png'),
bbox_inches='tight')
 plt.close()

Box Plots - Separate for score metrics and generation metrics
for metrics_group, group_name in [(score_metrics, 'score_metrics'),
 (generation_metrics, 'generation_metrics')]:
 plt.figure(figsize=(20, 10))
 for i, metric in enumerate(metrics_group, 1):
 plt.subplot(1, len(metrics_group), i)
 sns.boxplot(x='model_name', y=metric, data=df)
 plt.title(metric)
 plt.xticks(rotation=45, ha='right')
 plt.ylim(0, df[metric].max() * 1.1) # Set y-axis limits with 10% padding
 plt.xlabel('')
 if i != 1:
 plt.ylabel('')
 plt.tight_layout()
 plt.savefig(os.path.join(output_dir, f'boxplots_{group_name}.png'),
bbox_inches='tight')
 plt.close()

Radar Charts - Separate for score metrics and generation metrics
for metrics_group, group_name in [(score_metrics, 'score_metrics'),
 (generation_metrics, 'generation_metrics')]:
 create_radar_chart(df, metrics_group, f'{dataset_name}_{group_name}', output_dir)

return desc_stats

def main():
 output_dir = 'billsum_analysis_results'
 os.makedirs(output_dir, exist_ok=True)

 print("Loading detailed data...")
 detailed_df = pd.read_csv('billsum_evaluation_detailed.csv')
 detailed_df['model_name'] = detailed_df['model_name'].apply(clean_model_name)
 detailed_df['model_name'] = detailed_df['model_name'].astype('category')

 print("\nAnalyzing detailed results...")
 detailed_stats = create_analysis_for_dataset(detailed_df, 'detailed_metrics',
 os.path.join(output_dir, 'detailed_analysis'))

 print("\nAnalysis complete. Results saved in 'billsum_analysis_results' directory.")

if __name__ == "__main__":
 main()

```



**Figure-32: Comparison of models' average performance across GPT-4o's evaluation metrics &**

**Figure-33: Violin plots of models' score distributions across GPT-4o evaluation metrics**

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kruskal
from scikit_posthocs import posthoc_dunn
import warnings
import os

warnings.filterwarnings('ignore')

Create a directory for outputs
output_dir = 'billsum_analysis_results'
os.makedirs(output_dir, exist_ok=True)

Load the data
df = pd.read_csv('combined_data_truncated.csv')

Ensure 'model_name' is treated as a categorical variable
df['model_name'] = df['model_name'].astype('category')

Modified name simplification function to handle both LN and RMSN models
def simplify_model_name(name):
 if '__' in name:
 parts = name.split('__')
 if 'LN-' in name:
 return f"LN-{parts[-2]}"
 elif 'RMSN' in name:
 return f"RMSN-{parts[-2]}"
 return name

df['model_name_simple'] = df['model_name'].apply(simplify_model_name)

df['model_name_simple'] = df['model_name'].apply(simplify_model_name)

Define score columns for summarization
score_columns = ['Overall Score', 'Relevance Score', 'Conciseness Score', 'Fluency Score',
'Accuracy Score', 'Coherence Score']

1. Effect Size Calculation (Eta-squared)
def calculate_effect_size(df, metric):
 try:
 h_statistic, _ = kruskal(*[group[metric].values for name, group in
df.groupby('model_name')])
 n = len(df)
 eta_squared = (h_statistic - len(df['model_name'].unique()) + 1) / (n - 1)
 return max(0, eta_squared) # Ensure non-negative value
 except Exception as e:
 print(f"Error calculating effect size for {metric}: {e}")
 return np.nan
```

```

effect_sizes = {metric: calculate_effect_size(df, metric) for metric in score_columns}
effect_sizes_df = pd.DataFrame.from_dict(effect_sizes, orient='index', columns=['Effect Size (η^2 ')])
effect_sizes_df.to_csv(os.path.join(output_dir, 'effect_sizes.csv'))

2. Non-parametric Tests (Kruskal-Wallis and Dunn's test)
def perform_kruskal_dunn(df, metric):
 try:
 # Kruskal-Wallis test
 kruskal_result = kruskal(*[group[metric].values for name, group in df.groupby('model_name')])

 # Dunn's test
 dunn_result = posthoc_dunn(df, val_col=metric, group_col='model_name',
 p_adjust='bonferroni')

 # Calculate effect size for Dunn's test
 dunn_effect_size = calculate_effect_size(df, metric)

 return {
 'kruskal': pd.DataFrame({'statistic': [kruskal_result.statistic],
 'p-value': [kruskal_result.pvalue]}),
 'dunn_pval': dunn_result,
 'dunn_effect': dunn_effect_size
 }
 except Exception as e:
 print(f"Error performing Kruskal-Wallis and Dunn's test for {metric}: {e}")
 return None

test_results = {metric: perform_kruskal_dunn(df, metric) for metric in score_columns}

Save results
for metric, results in test_results.items():
 if results is not None:
 results['kruskal'].to_csv(os.path.join(output_dir, f'kruskal_wallis_{metric.replace(" ", "_")}.csv'),
 index=False)
 results['dunn_pval'].to_csv(os.path.join(output_dir,
f'dunn_test_pvalues_{metric.replace(" ", "_")}.csv'))
 pd.DataFrame({'Effect Size (η^2)': [results['dunn_effect']]}) .to_csv(
 os.path.join(output_dir, f'dunn_test_effect_size_{metric.replace(" ", "_")}.csv'))

3. Basic Descriptive Statistics
descriptive_stats = df.groupby('model_name_simple')[score_columns].agg(['mean', 'std', 'min', 'max'])
descriptive_stats.to_csv(os.path.join(output_dir, 'descriptive_statistics.csv'))

4. Visualizations
Violin Plots
plt.figure(figsize=(20, 15))
for i, metric in enumerate(score_columns, 1):
 plt.subplot(2, 3, i)

```

```

sns.violinplot(x='model_name_simple', y=metric, data=df)
plt.title(metric)
plt.xticks(rotation=45, ha='right')
plt.xlabel('')
plt.ylim(1, 5)
if i % 3 != 1:
 plt.ylabel('')
plt.tight_layout()
plt.savefig(os.path.join(output_dir, 'violin_plots.png'), bbox_inches='tight')
plt.close()

Radar Chart
def radar_chart(df, metrics):
 means = df.groupby('model_name_simple')[metrics].mean()
 angles = np.linspace(0, 2*np.pi, len(metrics), endpoint=False)
 means = pd.concat([means, means.iloc[:, :1]], axis=1)
 angles = np.concatenate((angles, [angles[0]]))

 fig, ax = plt.subplots(figsize=(14, 10), subplot_kw=dict(projection='polar'))
 for model in means.index:
 values = means.loc[model].values
 ax.plot(angles, values, 'o-', linewidth=2, label=model)
 ax.fill(angles, values, alpha=0.25)
 ax.set_thetagrids(angles[:-1] * 180/np.pi, metrics)
 ax.set_yticks([1, 5]) # Assuming scores are between 0 and 5
 plt.legend(loc='center left', bbox_to_anchor=(1.1, 0.5))
 plt.title("Model Performance Across Metrics")
 plt.tight_layout()
 plt.savefig(os.path.join(output_dir, 'radar_chart.png'), bbox_inches='tight')
 plt.close()

radar_chart(df, score_columns)

print(f"\nAnalysis complete. Results saved in the '{output_dir}' directory.")
print("Check the generated PNG files for visualizations.")

```

**Figure-36: Violin Plot showcasing distribution of F1 Scores**

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from statsmodels.stats.multicomp import pairwise_tukeyhsd
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from scipy.stats import kruskal
from scikit_posthocs import posthoc_dunn
import warnings
import os

warnings.filterwarnings('ignore')

def perform_kruskal_dunn(df, metric):
 """Perform Kruskal-Wallis and Dunn's test for a metric."""
 try:
 kruskal_result = kruskal(*[group[metric].values for name, group in df.groupby('model_name')])
 if kruskal_result.pvalue < 0.05:
 dunn_result = posthoc_dunn(df, val_col=metric, group_col='model_name',
 p_adjust='bonferroni')
 return pd.DataFrame({'statistic': [kruskal_result.statistic], 'p-value': [kruskal_result.pvalue]}), dunn_result
 return pd.DataFrame({'statistic': [kruskal_result.statistic], 'p-value': [kruskal_result.pvalue]}), None
 except Exception as e:
 print(f"Error performing Kruskal-Wallis and Dunn's test for {metric}: {e}")
 return None, None

def create_radar_chart(df, metrics, condition_name, output_dir):
 """Create radar chart with specific formatting."""
 means = df.groupby('model_name')[metrics].mean()
 angles = np.linspace(0, 2*np.pi, len(metrics), endpoint=False)
 means = pd.concat([means, means.iloc[:, :1]], axis=1)
 angles = np.concatenate((angles, [angles[0]]))

 fig, ax = plt.subplots(figsize=(14, 10), subplot_kw=dict(projection='polar'))
 for model in means.index:
 values = means.loc[model].values
 ax.plot(angles, values, 'o-', linewidth=2, label=model)
 ax.fill(angles, values, alpha=0.25)
 ax.set_thetagrids(angles[:-1] * 180/np.pi, metrics)
 ax.set_yticks([0, max(means.max().max(), 1.0)])
 plt.legend(loc='center left', bbox_to_anchor=(1.1, 0.5))
 plt.title(f"Model Performance Across Metrics - {condition_name}")
 plt.tight_layout()
 plt.savefig(os.path.join(output_dir, 'radar_chart.png'), bbox_inches='tight')
 plt.close()

def calculate_anova_summary(df, metrics):
```

```

"""Calculate ANOVA summary for each metric."""
anova_results = []

for metric in metrics:
 groups = [group[metric].values for name, group in df.groupby('model_name')]
 f_value, p_value = stats.f_oneway(*groups)

 # Calculate effect size (eta-squared)
 eta_squared = f_value / (f_value + df.groupby('model_name').size().iloc[0] - 1)

 anova_results.append({
 'Metric': metric,
 'F_value': f_value,
 'p_value': p_value,
 'eta_squared': eta_squared
 })

return pd.DataFrame(anova_results)

def create_score_distribution_plots(df, metrics, condition_name, output_dir):
 """Create detailed score distribution plots."""
 plt.figure(figsize=(20, 15))

 for i, metric in enumerate(metrics, 1):
 plt.subplot(2, 3, i)
 for model in df['model_name'].unique():
 model_data = df[df['model_name'] == model]
 sns.kdeplot(data=model_data, x=metric, label=model)

 plt.title(f'Distribution of {metric}')
 plt.xlabel('Score')
 plt.ylabel('Density')
 if i == 3: # Place legend for third plot
 plt.legend(title='Model Name', bbox_to_anchor=(1.05, 1), loc='upper left')
 else:
 plt.legend([]) # Remove individual legends

 plt.tight_layout()
 plt.savefig(os.path.join(output_dir, 'detailed_score_distributions.png'),
 bbox_inches='tight')
 plt.close()

def create_analysis_for_condition(df, condition_name, output_dir):
 """Perform statistical analysis for a given condition."""
 metrics = ['f1_score', 'confidence_score', 'no_answer_probability']

 condition_dir = os.path.join(output_dir, condition_name)
 os.makedirs(condition_dir, exist_ok=True)

 # 1. Basic Descriptive Statistics
 desc_stats = df.groupby('model_name')[metrics].agg(['mean', 'std', 'min', 'max', 'count'])
 desc_stats.to_csv(os.path.join(condition_dir, 'descriptive_statistics.csv'))

```

```

2. Effect Size Calculation
effect_sizes = {}
for metric in metrics:
 try:
 f_value, _ = stats.f_oneway(*[group[metric] for name, group in
df.groupby('model_name')])
 effect_sizes[metric] = f_value / (f_value +
df.groupby('model_name').size().iloc[0] - 1)
 except Exception as e:
 print(f"Error calculating effect size for {metric}: {e}")
 effect_sizes[metric] = np.nan

effect_sizes_df = pd.DataFrame.from_dict(effect_sizes, orient='index', columns=['Effect
Size'])
effect_sizes_df.to_csv(os.path.join(condition_dir, 'effect_sizes.csv'))

Add ANOVA summary
anova_summary = calculate_anova_summary(df, metrics)
anova_summary.to_csv(os.path.join(condition_dir, 'anova_summary.csv'), index=False)

Create detailed score distribution plots
create_score_distribution_plots(df, metrics, condition_name, condition_dir)

score_dist_summary = df[metrics].agg([
 'mean', 'std', 'min', 'max',
 lambda x: x.quantile(0.25),
 lambda x: x.quantile(0.75),
 'skew', 'kurt'
]).round(4)

score_dist_summary.index = ['Mean', 'Std Dev', 'Min', 'Max', 'Q1', 'Q3', 'Skewness',
'Kurtosis']
score_dist_summary.to_csv(os.path.join(condition_dir, 'score_distribution_summary.csv'))

3. Post-hoc Tests
for metric in metrics:
 tukey = pairwise_tukeyhsd(df[metric], df['model_name'])
 pd.DataFrame(data=tukey._results_table.data[1:],
 columns=tukey._results_table.data[0]).to_csv(
 os.path.join(condition_dir, f'tukey_hsd_{metric}.csv'), index=False)

4. Correlation Analysis
correlation_matrix = df[metrics].corr()
correlation_matrix.to_csv(os.path.join(condition_dir, 'correlation_matrix.csv'))

5. PCA Analysis
scaler = StandardScaler()
pca = PCA()
pca_result = pca.fit_transform(scaler.fit_transform(df[metrics]))
pca_df = pd.DataFrame({
 'Principal Component': range(1, len(pca.explained_variance_ratio_) + 1),
 'Explained Variance Ratio': pca.explained_variance_ratio_,
 'Cumulative Explained Variance Ratio': np.cumsum(pca.explained_variance_ratio_)
})

```

```

 })
 pca_df.to_csv(os.path.join(condition_dir, 'pca_results.csv'), index=False)

6. Kruskal-Wallis and Dunn's Tests
for metric in metrics:
 kruskal_result, dunn_result = perform_kruskal_dunn(df, metric)
 if kruskal_result is not None:
 kruskal_result.to_csv(os.path.join(condition_dir, f'kruskal_{metric}.csv'),
index=False)
 if dunn_result is not None:
 dunn_result.to_csv(os.path.join(condition_dir, f'dunn_test_{metric}.csv'))

7. Visualizations

Correlation Heatmap
plt.figure(figsize=(12, 10))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title(f'Correlation Heatmap - {condition_name}')
plt.tight_layout()
plt.savefig(os.path.join(condition_dir, 'correlation_heatmap.png'))
plt.close()

Violin Plots
plt.figure(figsize=(20, 10))
for i, metric in enumerate(metrics, 1):
 plt.subplot(1, 3, i)
 sns.violinplot(x='model_name', y=metric, data=df)
 plt.title(f'{metric} - {condition_name}')
 plt.xticks(rotation=45, ha='right')
 plt.ylim(0, 1) # Set y-axis limits
 if i % 3 != 1:
 plt.ylabel('')
plt.tight_layout()
plt.savefig(os.path.join(condition_dir, 'violin_plots.png'), bbox_inches='tight')
plt.close()

Distribution Plots
plt.figure(figsize=(20, 15))
for i, metric in enumerate(metrics, 1):
 plt.subplot(1, 3, i)
 for model in df['model_name'].unique():
 sns.kdeplot(data=df[df['model_name'] == model], x=metric, label=model, clip=(0,
1)) # Add clip parameter
 plt.title(f'Distribution of {metric} - {condition_name}')
 plt.xlabel('Score')
 plt.ylabel('Density')
 plt.xlim(0, 1) # Set x-axis limits
 if i == 3: # Place legend outside the plots
 plt.legend(title='Model Name', bbox_to_anchor=(1.05, 1), loc='upper left')
 else:
 plt.legend([])
plt.tight_layout()
plt.savefig(os.path.join(condition_dir, 'score_distributions.png'), bbox_inches='tight')

```

```

plt.close()

Box Plots
plt.figure(figsize=(20, 15))
for i, metric in enumerate(metrics, 1):
 plt.subplot(1, 3, i)
 sns.boxplot(x='model_name', y=metric, data=df)
 plt.title(metric)
 plt.xticks(rotation=45, ha='right')
 plt.ylim(0, 1) # Set y-axis limits
 plt.xlabel('')
 if i % 3 != 1:
 plt.ylabel('')
plt.tight_layout()
plt.savefig(os.path.join(condition_dir, 'boxplots.png'), bbox_inches='tight')
plt.close()

Radar Chart
create_radar_chart(df, metrics, condition_name, condition_dir)

return desc_stats

def clean_model_name(name):
 """Convert full model name to concise format."""
 # Extract the norm type and variant from the full name
 # Example: "shng2025/GPT-Valkyrie_LN-124m__AttnOnly__SQuAD" -> "LN-AttnOnly"
 try:
 # Split by underscores and extract relevant parts
 parts = name.split('_')
 norm_type = parts[1].split('-')[0] # Get LN or RMSN
 variant = parts[3].split('__')[0] # Get baseModel, noNorm, AttnOnly, or FFNonly
 return f"{norm_type}-{variant}"
 except:
 return name # Return original name if parsing fails

Then modify the main() function:
def main():
 output_dir = 'qa_analysis_results'
 os.makedirs(output_dir, exist_ok=True)

 print("Loading data...")
 df = pd.read_csv('squad_evaluation_detailed.csv')

 # Clean model names
 df['model_name'] = df['model_name'].apply(clean_model_name)
 df['model_name'] = df['model_name'].astype('category')

 exact_match_df = df[df['exact_match'] == 1]
 no_match_df = df[df['exact_match'] == 0]
 all_df = df

 print(f"\nAnalyzing exact matches ({len(exact_match_df)} samples)...")
 exact_stats = create_analysis_for_condition(exact_match_df, 'exact_match', output_dir)

```

```

print(f"\nAnalyzing non-matches ({len(no_match_df)} samples)...")
no_match_stats = create_analysis_for_condition(no_match_df, 'no_match', output_dir)

print(f"\nAnalyzing all cases ({len(all_df)} samples)...")
all_stats = create_analysis_for_condition(all_df, 'all_cases', output_dir)

comparison_summary = pd.DataFrame({
 'Exact Match Count': len(exact_match_df),
 'No Match Count': len(no_match_df),
 'Total Count': len(all_df),
 'Exact Match Ratio': len(exact_match_df) / len(all_df),
 'No Match Ratio': len(no_match_df) / len(all_df)
}, index=['Summary'])

comparison_summary.to_csv(os.path.join(output_dir, 'match_distribution_summary.csv'))

print(f"\nAnalysis complete. Results saved in '{output_dir}' directory.")

if __name__ == "__main__":
 main()

```

**Figure-37: Comparison of models' average performance across GPT-4o's evaluation metrics &**

**Figure-38: Violin plots of models' score distributions across GPT-4o evaluation metrics**

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from statsmodels.stats.multicomp import pairwise_tukeyhsd
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from scipy.stats import kruskal
from scikit_posthocs import posthoc_dunn
import warnings
import os

warnings.filterwarnings('ignore')

Create a directory for CSV outputs
output_dir = 'squad_analysis_results'
os.makedirs(output_dir, exist_ok=True)

Load the data
df = pd.read_csv('combined_data_truncated_squad.csv')

print("Columns in the CSV file:")
print(df.columns)

Ensure 'model_name' is treated as a categorical variable

```

```

df['model_name'] = df['model_name'].astype('category')

Define new score columns for SQuAD task
score_columns = ['Overall Score', 'Correctness Score', 'Completeness Score', 'Relevance Score', 'Fluency Score', 'Conciseness Score']

print(f"\nUsing score columns: {score_columns}")

1. Effect Size Calculation
def calculate_effect_size(df, metric):
 try:
 f_value, _ = stats.f_oneway(*[group[metric] for name, group in df.groupby('model_name')])
 return f_value / (f_value + df.groupby('model_name').size().iloc[0] - 1)
 except Exception as e:
 print(f"Error calculating effect size for {metric}: {e}")
 return np.nan

effect_sizes = {metric: calculate_effect_size(df, metric) for metric in score_columns}
effect_sizes_df = pd.DataFrame.from_dict(effect_sizes, orient='index', columns=['Effect Size'])
effect_sizes_df.to_csv(os.path.join(output_dir, 'effect_sizes.csv'))

2. Post-hoc Tests
def perform_tukey_hsd(df, metric):
 try:
 tukey = pairwise_tukeyhsd(df[metric], df['model_name'])
 return pd.DataFrame(data=tukey._results_table.data[1:], columns=tukey._results_table.data[0])
 except Exception as e:
 print(f"Error performing Tukey HSD for {metric}: {e}")
 return pd.DataFrame()

tukey_results = {metric: perform_tukey_hsd(df, metric) for metric in score_columns}
for metric, result in tukey_results.items():
 result.to_csv(os.path.join(output_dir, f'tukey_hsd_{metric.replace(" ", "_")}.csv'), index=False)

3. Correlation Analysis
correlation_matrix = df[score_columns].corr()
correlation_matrix.to_csv(os.path.join(output_dir, 'correlation_matrix.csv'))

4. Principal Component Analysis
scaler = StandardScaler()
pca = PCA()
pca_result = pca.fit_transform(scaler.fit_transform(df[score_columns]))
pca_df = pd.DataFrame({'Explained Variance Ratio': pca.explained_variance_ratio_})
pca_df.to_csv(os.path.join(output_dir, 'pca_results.csv'), index=False)

5. Non-parametric Tests
def perform_kruskal_dunn(df, metric):
 try:
 groups = [group[metric].values for name, group in df.groupby('model_name')]

```

```

kruskal_result = kruskal(*groups)
dunn_result = posthoc_dunn(df, val_col=metric, group_col='model_name',
p_adjust='bonferroni')
kruskal_df = pd.DataFrame({'statistic': [kruskal_result.statistic], 'p-value':
[kruskal_result.pvalue]})

Diagnostic information
print(f"\nDiagnostic information for {metric}:")
print(f"Kruskal-Wallis p-value: {kruskal_result.pvalue}")
print(f"Number of groups: {len(groups)}")
print(f"Sample sizes: {[len(group) for group in groups]}")
print(f"Group means: {[group.mean() for group in groups]}")
print(f"Percentage of 1s in Dunn's test: {(dunn_result == 1).sum().sum() / dunn_result.size * 100:.2f}%")

return kruskal_df, dunn_result
except Exception as e:
 print(f"Error performing Kruskal-Wallis and Dunn's test for {metric}: {e}")
 return None, None

Use the function and save results
kruskal_dunn_results = {metric: perform_kruskal_dunn(df, metric) for metric in score_columns}
for metric, (kruskal_result, dunn_result) in kruskal_dunn_results.items():
 if kruskal_result is not None:
 kruskal_result.to_csv(os.path.join(output_dir, f'kruskal_wallis_{metric.replace(" ", "_")}.csv'), index=False)
 if dunn_result is not None:
 dunn_result.to_csv(os.path.join(output_dir, f'dunn_test_{metric.replace(" ", "_")}.csv'))

6. Visualizations (unchanged)
plt.figure(figsize=(12, 10))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap of SQuAD Metrics')
plt.tight_layout()
plt.savefig(os.path.join(output_dir, 'correlation_heatmap_squad_metrics.png'))
plt.close()

Violin Plots
plt.figure(figsize=(20, 15))
for i, metric in enumerate(score_columns, 1):
 plt.subplot(2, 3, i)
 sns.violinplot(x='model_name', y=metric, data=df)
 plt.title(metric)
 plt.xticks(rotation=45, ha='right')
 plt.xlabel('')
 if i % 3 != 1:
 plt.ylabel('')
plt.tight_layout()
plt.savefig(os.path.join(output_dir, 'violin_plots_squad_metrics.png'), bbox_inches='tight')
plt.close()

Radar Chart

```

```

def radar_chart(df, metrics):
 means = df.groupby('model_name')[metrics].mean()
 angles = np.linspace(0, 2*np.pi, len(metrics), endpoint=False)
 means = pd.concat([means, means.iloc[:, :1]], axis=1)
 angles = np.concatenate((angles, [angles[0]]))

 fig, ax = plt.subplots(figsize=(14, 10), subplot_kw=dict(projection='polar'))
 for model in means.index:
 values = means.loc[model].values
 ax.plot(angles, values, 'o-', linewidth=2, label=model)
 ax.fill(angles, values, alpha=0.25)
 ax.set_thetagrids(angles[:-1] * 180/np.pi, metrics)
 ax.set_ylim(0, 5) # Assuming scores are between 0 and 5
 plt.legend(loc='center left', bbox_to_anchor=(1.1, 0.5))
 plt.title("Model Performance Across SQuAD Metrics")
 plt.tight_layout()
 plt.savefig(os.path.join(output_dir, 'radar_chart_squad_metrics.png'),
bbox_inches='tight')
 plt.close()

radar_chart(df, score_columns)

Distribution of Scores
plt.figure(figsize=(20, 15))
for i, metric in enumerate(score_columns, 1):
 plt.subplot(2, 3, i)
 for model in df['model_name'].unique():
 sns.kdeplot(data=df[df['model_name'] == model], x=metric, label=model)
 plt.title(f'Distribution of {metric}')
 plt.xlabel('Score')
 plt.ylabel('Density')
 if i == 3: # Place legend outside the plots
 plt.legend(title='Model Name', bbox_to_anchor=(1.05, 1), loc='upper left')
 else:
 plt.legend([]) # Remove individual legends
plt.tight_layout()
plt.savefig(os.path.join(output_dir, 'score_distributions_squad_metrics.png'),
bbox_inches='tight')
plt.close()

Boxplots
plt.figure(figsize=(20, 15))
for i, metric in enumerate(score_columns, 1):
 plt.subplot(2, 3, i)
 sns.boxplot(x='model_name', y=metric, data=df)
 plt.title(metric)
 plt.xticks(rotation=45, ha='right')
 plt.xlabel('')
 if i % 3 != 1:
 plt.ylabel('')
plt.tight_layout()
plt.savefig(os.path.join(output_dir, 'boxplots_squad_metrics.png'), bbox_inches='tight')
plt.close()

```

```
7. Basic Descriptive Statistics
descriptive_stats = df.groupby('model_name')[score_columns].agg(['mean', 'std', 'min',
'max'])
descriptive_stats.to_csv(os.path.join(output_dir, 'descriptive_statistics.csv'))

print(f"\nAnalysis complete. Results saved in CSV files in the '{output_dir}' directory.")
print("Check the generated PNG files for visualizations.")
```

## Appendix D - pre training code

This section includes the code used for pre-training our models. As noted in Section 4, this code is heavily inspired by Andrej Karpathy's nanoGPT and is a reproduction of the GPT-2 124M model. The primary differences between each model lie in the implementation of different normalisation techniques for our investigation.

### D.1 Layer Normalisation Pre-training

```
import os
import math
import time
import inspect
from dataclasses import dataclass
import torch
import torch.nn as nn
from torch.nn import functional as F
from hellaswag import render_example, iterate_examples

from argparse import Namespace

import wandb

import tiktoken
import numpy as np

from torch.distributed import init_process_group, destroy_process_group
from torch.nn.parallel import DistributedDataParallel as DDP
import torch.distributed as dist

from huggingface_hub import Repository, create_branch
import logging
import glob

class CausalSelfAttention(nn.Module):

 def __init__(self, config):
 super().__init__()
 assert config.n_embd % config.n_head == 0
 # key, query, value projections for all heads, but in a batch
 self.c_attn = nn.Linear(config.n_embd, 3 * config.n_embd)
 # output projection
 self.c_proj = nn.Linear(config.n_embd, config.n_embd)
 self.c_proj.NANOGLP_SCALE_INIT = 1
 # regularization
 self.n_head = config.n_head
 self.n_embd = config.n_embd

 def forward(self, x):
 B, T, C = x.size() # batch size, sequence length, embedding dimensionality (n_embd)
 # calculate query, key, values for all heads in batch and move head forward to be the batch dim
 # nh is "number of heads", hs is "head size", and C (number of channels) = nh * hs
 # e.g. in GPT-2 (124M), n_head=12, hs=64, so nh*hs=C=768 channels in the Transformer
 qkv = self.c_attn(x)
 q, k, v = qkv.split(self.n_embd, dim=2)
 k = k.view(B, T, self.n_head, C // self.n_head).transpose(1, 2) # (B, nh, T, hs)
 q = q.view(B, T, self.n_head, C // self.n_head).transpose(1, 2) # (B, nh, T, hs)
 v = v.view(B, T, self.n_head, C // self.n_head).transpose(1, 2) # (B, nh, T, hs)
 y = F.scaled_dot_product_attention(q, k, v, is_causal=True) # flash attention
 y = y.transpose(1, 2).contiguous().view(B, T, C) # re-assemble all head outputs side by side
 # output projection
```

```

y = self.c_proj(y)
return y

class MLP(nn.Module):

 def __init__(self, config):
 super().__init__()
 self.c_fc = nn.Linear(config.n_embd, 4 * config.n_embd)
 self.gelu = nn.GELU(approximate='tanh')
 self.c_proj = nn.Linear(4 * config.n_embd, config.n_embd)
 self.c_proj.NANOGLP_SCALE_INIT = 1

 def forward(self, x):
 x = self.c_fc(x)
 x = self.gelu(x)
 x = self.c_proj(x)
 return x

from torch.nn.parameter import Parameter
from torch import Tensor, Size
from typing import Union, List, Tuple, Optional
from torch.nn import init
import numbers

_shape_t = Union[int, List[int], Size]

class LayerNorm(nn.Module):

 __constants__ = ['normalized_shape', 'eps', 'elementwise_affine']
 normalized_shape: Tuple[int, ...]
 eps: float
 elementwise_affine: bool

 def __init__(self, normalized_shape: _shape_t, eps: float = 1e-5, elementwise_affine: bool = True, bias: bool = True, device=None, dtype=None) -> None:
 factory_kwargs = {'device': device, 'dtype': dtype}
 super().__init__()
 if isinstance(normalized_shape, numbers.Integral):
 # mypy error: incompatible types in assignment
 normalized_shape = (normalized_shape,) # type: ignore[assignment]
 self.normalized_shape = tuple(normalized_shape) # type: ignore[arg-type]
 self.eps = eps
 self.elementwise_affine = elementwise_affine
 if self.elementwise_affine:
 self.gain = Parameter(torch.empty(self.normalized_shape, **factory_kwargs))
 if bias:
 self.bias = Parameter(torch.empty(self.normalized_shape, **factory_kwargs))
 else:
 self.register_parameter('bias', None)
 else:
 self.register_parameter('gain', None)
 self.register_parameter('bias', None)

 self.reset_parameters()

 def reset_parameters(self) -> None:
 if self.elementwise_affine:
 init.ones_(self.gain)
 if self.bias is not None:

```

```

 init.zeros_(self.bias)

 def forward(self, x: Tensor) -> Tensor:
 """Forward pass of LayerNorm."""
 mean = x.mean(dim=-1, keepdim=True)
 var = torch.var(x, dim=-1, keepdim=True, unbiased=False) # mean_x2 = torch.square(x).mean(dim=-1,
keepdim=True) // alternate code for variance calculation, faster torch.compile
 # var = mean_x2 - torch.square(mean)
 x_norm = (x - mean) / torch.sqrt(var + self.eps)

 if self.elementwise_affine:
 x_norm = self.gain * x_norm + self.bias # changed name of weight parameter, to gain parameter to
concur with naming convention

 return x_norm

 def extra_repr(self) -> str:
 return '{normalized_shape}, eps={eps},\n' \
 'elementwise_affine={elementwise_affine}'.format(**self.__dict__)

class Block(nn.Module):

 def __init__(self, config):
 super().__init__()
 self.ln_1 = LayerNorm(config.n_embd)
 self.attn = CausalSelfAttention(config)
 self.ln_2 = LayerNorm(config.n_embd)
 self.mlp = MLP(config)

 def forward(self, x):
 x = x + self.attn(self.ln_1(x))
 x = x + self.mlp(self.ln_2(x))
 return x

@dataclass
class GPTConfig:
 block_size: int = 1024 # max sequence length
 vocab_size: int = 50257 # number of tokens: 50,000 BPE merges + 256 bytes tokens + 1 <|endoftext|> token
 n_layer: int = 12 # number of layers
 n_head: int = 12 # number of heads
 n_embd: int = 768 # embedding dimension

class GPT(nn.Module):

 def __init__(self, config):
 super().__init__()
 self.config = config

 self.transformer = nn.ModuleDict(dict(
 wte = nn.Embedding(config.vocab_size, config.n_embd),
 wpe = nn.Embedding(config.block_size, config.n_embd),
 h = nn.ModuleList([Block(config) for _ in range(config.n_layer)]),
 ln_f = LayerNorm(config.n_embd),
))
 self.lm_head = nn.Linear(config.n_embd, config.vocab_size, bias=False)

 # weight sharing scheme

```

```

 self.transformer.wte.weight = self.lm_head.weight

 # init params
 self.apply(self._init_weights)

def _init_weights(self, module):
 if isinstance(module, nn.Linear):
 std = 0.02
 if hasattr(module, 'NANOGLP_SCALE_INIT'):
 std *= (2 * self.config.n_layer) ** -0.5
 torch.nn.init.normal_(module.weight, mean=0.0, std=std)
 if module.bias is not None:
 torch.nn.init.zeros_(module.bias)
 elif isinstance(module, nn.Embedding):
 torch.nn.init.normal_(module.weight, mean=0.0, std=0.02)

def forward(self, idx, targets=None):
 # idx is of shape (B, T)
 B, T = idx.size()
 assert T <= self.config.block_size, f"Cannot forward sequence of length {T}, block size is only {self.config.block_size}"
 # forward the token and position embeddings
 pos = torch.arange(0, T, dtype=torch.long, device=idx.device) # shape (T)
 pos_emb = self.transformer.wpe(pos) # position embeddings of shape (T, n_embd)
 tok_emb = self.transformer.wte(idx) # token embeddings of shape (B, T, n_embd)
 x = tok_emb + pos_emb
 # forward the blocks of the transformer
 for block in self.transformer.h:
 x = block(x)
 # forward the final layernorm and the classifier
 x = self.transformer.ln_f(x)
 logits = self.lm_head(x) # (B, T, vocab_size)
 loss = None
 if targets is not None:
 loss = F.cross_entropy(logits.view(-1, logits.size(-1)), targets.view(-1))
 return logits, loss

@classmethod
def from_pretrained(cls, model_type):
 """Loads pretrained GPT-2 model weights from huggingface"""
 assert model_type in {'gpt2', 'gpt2-medium', 'gpt2-large', 'gpt2-xl'}
 from transformers import GPT2LMHeadModel
 print("loading weights from pretrained gpt: %s" % model_type)

 # n_layer, n_head and n_embd are determined from model_type
 config_args = {
 'gpt2': dict(n_layer=12, n_head=12, n_embd=768), # 124M params
 'gpt2-medium': dict(n_layer=24, n_head=16, n_embd=1024), # 350M params
 'gpt2-large': dict(n_layer=36, n_head=20, n_embd=1280), # 774M params
 'gpt2-xl': dict(n_layer=48, n_head=25, n_embd=1600), # 1558M params
 }[model_type]
 config_args['vocab_size'] = 50257 # always 50257 for GPT model checkpoints
 config_args['block_size'] = 1024 # always 1024 for GPT model checkpoints
 # create a from-scratch initialized minGPT model
 config = GPTConfig(**config_args)
 model = GPT(config)
 sd = model.state_dict()
 sd_keys = sd.keys()
 sd_keys = [k for k in sd_keys if not k.endswith('.attn.bias')]

param
 # discard this mask / buffer, not a
 # init a huggingface/transformers model

```

```

model_hf = GPT2LMHeadModel.from_pretrained(model_type)
sd_hf = model_hf.state_dict()

copy while ensuring all of the parameters are aligned and match in names and shapes
sd_keys_hf = sd_hf.keys()
sd_keys_hf = [k for k in sd_keys_hf if not k.endswith('.attn.masked_bias')] # ignore these, just a
buffer
sd_keys_hf = [k for k in sd_keys_hf if not k.endswith('.attn.bias')] # same, just the mask (buffer)
transposed = ['attn.c_attn.weight', 'attn.c_proj.weight', 'mlp.c_fc.weight', 'mlp.c_proj.weight']
basically the openai checkpoints use a "Conv1D" module, but we only want to use a vanilla Linear
this means that we have to transpose these weights when we import them
assert len(sd_keys_hf) == len(sd_keys), f"mismatched keys: {len(sd_keys_hf)} != {len(sd_keys)}"
for k in sd_keys_hf:
 if any(k.endswith(w) for w in transposed):
 # special treatment for the Conv1D weights we need to transpose
 assert sd_hf[k].shape[::-1] == sd[k].shape
 with torch.no_grad():
 sd[k].copy_(sd_hf[k].t())
 else:
 # vanilla copy over the other parameters
 assert sd_hf[k].shape == sd[k].shape
 with torch.no_grad():
 sd[k].copy_(sd_hf[k])

return model

def configure_optimizers(self, weight_decay, learning_rate, device_type):
 # start with all of the candidate parameters (that require grad)
 param_dict = {pn: p for pn, p in self.named_parameters()}
 param_dict = {pn: p for pn, p in param_dict.items() if p.requires_grad}
 # create optim groups. Any parameters that is 2D will be weight decayed, otherwise no.
 # i.e. all weight tensors in matmuls + embeddings decay, all biases and layernorms don't.
 decay_params = [p for n, p in param_dict.items() if p.dim() >= 2]
 nodecay_params = [p for n, p in param_dict.items() if p.dim() < 2]
 optim_groups = [
 {'params': decay_params, 'weight_decay': weight_decay},
 {'params': nodecay_params, 'weight_decay': 0.0}
]
 num_decay_params = sum(p.numel() for p in decay_params)
 num_nodecay_params = sum(p.numel() for p in nodecay_params)
 if master_process:
 print(f"num decayed parameter tensors: {len(decay_params)}, with {num_decay_params:,} parameters")
 print(f"num non-decayed parameter tensors: {len(nodecay_params)}, with {num_nodecay_params:,} parameters")
 # Create AdamW optimizer and use the fused version if it is available
 fused_available = 'fused' in inspect.signature(torch.optim.AdamW).parameters
 use_fused = fused_available and device_type == "cuda"
 if master_process:
 print(f"using fused AdamW: {use_fused}")
 optimizer = torch.optim.AdamW(optim_groups, lr=learning_rate, betas=(0.9, 0.95), eps=1e-8,
fused=fused)
 return optimizer

def load_tokens(filename):
 npt = np.load(filename)
 npt = npt.astype(np.int32) # added after video
 ptt = torch.tensor(npt, dtype=torch.long)
 return ptt

class DataLoaderLite:
 def __init__(self, B, T, process_rank, num_processes, split):

```

```

self.B = B
self.T = T
self.process_rank = process_rank
self.num_processes = num_processes
assert split in {'train', 'val'}

get the shard filenames
data_root = "../../edu_fineweb10B"
shards = os.listdir(data_root)
shards = [s for s in shards if split in s] # listing out shards file in the data_root dir
shards = sorted(shards)
shards = [os.path.join(data_root, s) for s in shards]
self.shards = shards
assert len(shards) > 0, f"no shards found for split {split}"
if master_process:
 print(f"found {len(shards)} shards for split {split}")

NEW IMPL
self.current_shard = 0
self.current_position = self.B * self.T * self.process_rank

self.reset()

def reset(self):
 # state, init at shard zero
 self.current_shard = 0
 self.tokens = load_tokens(self.shards[self.current_shard])
 self.current_position = self.B * self.T * self.process_rank

def set_state(state):
 self.current_shard = state['current_shard']
 self.current_position = state['current_position']
 self.tokens = load_tokens(self.shards[self.current_shard])

def get_state(self):
 return {
 'current_shard': self.current_shard,
 'current_position': self.current_position
 }

def next_batch(self):
 B, T = self.B, self.T
 buf = self.tokens[self.current_position : self.current_position+B*T+1]
 x = (buf[:-1]).view(B, T) # inputs
 y = (buf[1:]).view(B, T) # targets
 # advance the position in the tensor
 self.current_position += B * T * self.num_processes
 # if loading the next batch would be out of bounds, advance to next shard
 if self.current_position + (B * T * self.num_processes + 1) > len(self.tokens):
 self.current_shard = (self.current_shard + 1) % len(self.shards)
 self.tokens = load_tokens(self.shards[self.current_shard])
 self.current_position = B * T * self.process_rank
 return x, y

def setup_logging(project_name, args):
 logger = logging.getLogger(__name__)
 dir_name = "./log"
 os.makedirs(dir_name, exist_ok=True)
 print(f"Directory '{dir_name}' {'already exists' if os.path.exists(dir_name) else 'was created'}.")

logging.basicConfig(

```

```

format="%(asctime)s - %(levelname)s - %(name)s - %(message)s",
datefmt="%m/%d/%Y %H:%M:%S",
level=logging.INFO,
handlers=[
 logging.FileHandler(f"log/debug_{ddp_rank}.log"),
 logging.StreamHandler(),
],
)

if master_process:
 wandb.init(project=project_name, config=args, dir="..")
 run_name = wandb.run.name
 wandb_id = wandb.run.id
 logger.setLevel(logging.INFO)
 print(f"Starting new run: {run_name}")
else:
 run_name = ""
 wandb_id = ""
 logger.setLevel(logging.ERROR)

return logger, run_name, wandb_id

def resume_logging(project_name, run_id, args):
 logger = logging.getLogger(__name__)
 dir_name = "./log"
 os.makedirs(dir_name, exist_ok=True)
 print(f"Directory '{dir_name}' {'already exists' if os.path.exists(dir_name) else 'was created'}.")

 logging.basicConfig(
 format="%(asctime)s - %(levelname)s - %(name)s - %(message)s",
 datefmt="%m/%d/%Y %H:%M:%S",
 level=logging.INFO,
 handlers=[
 logging.FileHandler(f"log/debug_{ddp_rank}.log"),
 logging.StreamHandler(),
],
)

 if master_process:
 wandb.init(project=project_name, id=run_id, resume="must", config=args, dir='..')
 run_name = wandb.run.name
 logger.setLevel(logging.INFO)
 print(f"Resuming run: {run_name}")
 else:
 run_name = ""
 logger.setLevel(logging.ERROR)

 return logger, run_name

def log_metrics(metrics):
 if master_process:
 wandb.log(metrics)

def save_checkpoint(model, optimizer, step, val_loss, run_name, train_loader_state, val_loader_state, wandb_id):
 checkpoint = {
 'model': model.state_dict(),
 'optimizer': optimizer.state_dict(),
 'step': step,
 'val_loss': val_loss,
 'run_name': run_name,
 }

```

```

 'train_loader_state': train_loader_state,
 'val_loader_state': val_loader_state,
 'wandb_id': wandb_id
 }
checkpoint_path = os.path.join(log_dir, f"checkpoint_{step}.pt")
torch.save(checkpoint, checkpoint_path)
return checkpoint_path

def load_checkpoint(checkpoint_path, model, optimizer, train_loader, val_loader):
 checkpoint = torch.load(checkpoint_path)
 model.load_state_dict(checkpoint['model'])
 optimizer.load_state_dict(checkpoint['optimizer'])
 train_loader.set_state(checkpoint['train_loader_state'])
 val_loader.set_state(checkpoint['val_loader_state'])
 return checkpoint['step'], checkpoint['val_loss'], checkpoint['run_name'], checkpoint['wandb_id']

helper function for HellaSwag eval
takes tokens, mask, and logits, returns the index of the completion with the lowest loss

def get_most_likely_row(tokens, mask, logits):
 # evaluate the autoregressive loss at all positions
 shift_logits = (logits[..., :-1, :]).contiguous()
 shift_tokens = (tokens[..., 1:]).contiguous()
 flat_shift_logits = shift_logits.view(-1, shift_logits.size(-1))
 flat_shift_tokens = shift_tokens.view(-1)
 shift_losses = F.cross_entropy(flat_shift_logits, flat_shift_tokens, reduction='none')
 shift_losses = shift_losses.view(tokens.size(0), -1)
 # now get the average loss just for the completion region (where mask == 1), in each row
 shift_mask = (mask[..., 1:]).contiguous() # we must shift mask, so we start at the last prompt token
 masked_shift_losses = shift_losses * shift_mask
 # sum and divide by the number of 1s in the mask
 sum_loss = masked_shift_losses.sum(dim=1)
 avg_loss = sum_loss / shift_mask.sum(dim=1)
 # now we have a loss for each of the 4 completions
 # the one with the lowest loss should be the most likely
 pred_norm = avg_loss.argmin().item()
 return pred_norm

simple launch:
python train_gpt2.py
DDP launch for e.g. 8 GPUs:
torchrun --standalone --nproc_per_node=8 train_gpt2.py

run the training loop

set up DDP (distributed data parallel).
torchrun command sets the env variables RANK, LOCAL_RANK, and WORLD_SIZE
ddp = int(os.environ.get('RANK', -1)) != -1 # is this a ddp run?
if ddp:
 # use of DDP atm demands CUDA, we set the device appropriately according to rank
 assert torch.cuda.is_available(), "for now i think we need CUDA for DDP"
 init_process_group(backend='nccl')
 ddp_rank = int(os.environ['RANK'])
 ddp_local_rank = int(os.environ['LOCAL_RANK'])
 ddp_world_size = int(os.environ['WORLD_SIZE'])
 device = f'cuda:{ddp_local_rank}'
 torch.cuda.set_device(device)
 master_process = ddp_rank == 0 # this process will do logging, checkpointing etc.

```

```

else:
 # vanilla, non-DDP run
 ddp_rank = 0
 ddp_local_rank = 0
 ddp_world_size = 1
 master_process = True
 # attempt to autodetect device
 device = "cpu"
 if torch.cuda.is_available():
 device = "cuda"
 elif hasattr(torch.backends, "mps") and torch.backends.mps.is_available():
 device = "mps"
 print(f"using device: {device}")

GPTesla - 111M param setup in comment. Modification to make lighter training requirement needed
config = {
 "weight_decay": 0.1,
 # "lr_scheduler_type": "cosine",
 "gradient_accumulation_steps": 2**19 // (64 * 1024 * ddp_world_size), # total_batch_size // (B * T *
ddp_world_size
 "max_eval_steps": 20,
 "seq_length": 1024,

 # New centralised parameters
 "project_name": "shng2025/GPT-Valkyrie_LN-124m",
 "total_batch_size": 2**19, # temporarily because 6 GPUs # 2**19, ~0.5M, in number of tokens
 "micro_batch_size": 64,
 "max_lr": 6e-4,
 "min_lr": 6e-5, # 10% of max_lr // not used, as we are using weight_decay instead
 "warmup_steps": 715,
 "max_steps": 19073,
 "val_every": 500, # EVALUATION
 "generate_every": 500, # EVALUATION
 "hellaswag_every": 500, # EVALUATION
 "save_every": 2000, # SAVE CHECKPOINTING
 "log_dir": "./log",
 "device": "auto", # "auto", "cpu", "cuda", or "mps"
 "use_compile": True,
 "grad_clip": 1.0,
 "num_return_sequences": 4,
 "max_generate_length": 32,
 "top_k": 50,

 "resume_from_checkpoint": True,
}

args = Namespace(**config)
samples_per_step = torch.cuda.device_count() * args.micro_batch_size

project_name = args.project_name

Logging - DEPRECATED
if master_process:
 pass
 # run_name, wandb_id = setup_logging(project_name.split("/")[1])
 # print(f"Weights and Biases run name: {run_name}")
 # print(f"Weights and Biases run id : {wandb_id}")

```

```

added after video, pytorch can be serious about it's device vs. device_type distinction
device_type = "cuda" if device.startswith("cuda") else "cpu"

torch.manual_seed(1337)
if torch.cuda.is_available():
 torch.cuda.manual_seed(1337)

enc = tiktoken.get_encoding("gpt2")

total_batch_size = args.total_batch_size # 2**19, ~0.5M, in number of tokens
B = args.micro_batch_size # micro batch size
T = args.seq_length # sequence length
assert total_batch_size % (B * T * ddp_world_size) == 0, "make sure total_batch_size is divisible by B * T * ddp_world_size"
grad_accum_steps = total_batch_size // (B * T * ddp_world_size)
if master_process:
 print(f"total desired batch size: {total_batch_size}")
 print(f"=> calculated gradient accumulation steps: {grad_accum_steps}")

train_loader = DataLoaderLite(B=B, T=T, process_rank=ddp_rank, num_processes=ddp_world_size, split="train")
val_loader = DataLoaderLite(B=B, T=T, process_rank=ddp_rank, num_processes=ddp_world_size, split="val")

torch.set_float32_matmul_precision('medium')

create model
model = GPT(GPTConfig(vocab_size=50304))
model = GPT.from_pretrained("gpt2") # or init from OpenAI GPT-2
model.to(device)
use_compile = True # torch.compile interferes with HellaSwag eval and Generation. TODO fix
if use_compile:
 model = torch.compile(model)
if ddp:
 model = DDP(model, device_ids=[ddp_local_rank])
raw_model = model.module if ddp else model # always contains the "raw" unwrapped model

max_lr = args.max_lr
min_lr = max_lr * args.weight_decay
warmup_steps = args.warmup_steps
max_steps = args.max_steps # 19,073 steps is ~1 epoch, if data is 10B tokens and batch size 0.5M tokens
def get_lr(it):
 # 1) linear warmup for warmup_iters steps
 if it < warmup_steps:
 return max_lr * (it+1) / warmup_steps
 # 2) if it > lr_decay_iters, return min learning rate
 if it > max_steps:
 return min_lr
 # 3) in between, use cosine decay down to min learning rate
 decay_ratio = (it - warmup_steps) / (max_steps - warmup_steps)
 assert 0 <= decay_ratio <= 1
 coeff = 0.5 * (1.0 + math.cos(math.pi * decay_ratio)) # coeff starts at 1 and goes to 0
 return min_lr + coeff * (max_lr - min_lr)

optimize!
optimizer = raw_model.configure_optimizers(weight_decay=args.weight_decay, learning_rate=args.max_lr,
device_type=device_type)

create the log directory we will write checkpoints to and log to
log_dir = args.log_dir
os.makedirs(log_dir, exist_ok=True)
log_file = os.path.join(log_dir, f"log.txt")

```

```

with open(log_file, "w") as f: # open for writing to clear the file
 pass

"""

Initialize HuggingFace repository
if master_process:
 new_branch_name = run_name
 create_branch(project_name, repo_type="model", branch=new_branch_name)
 hf_repo = Repository("./", clone_from=project_name, revision=run_name)
"""

Training loop
starting_step = 0

if args.resume_from_checkpoint:
 checkpoint_dir = args.log_dir
 checkpoint_pattern = os.path.join(checkpoint_dir, "checkpoint_*.pt")
 checkpoint_files = glob.glob(checkpoint_pattern)
 latest_checkpoint = max(checkpoint_files, key=os.path.getctime)
 checkpoint_path = latest_checkpoint

 # Use the load_checkpoint function here
 starting_step, val_loss, run_name, wandb_id = load_checkpoint(checkpoint_path, raw_model, optimizer,
train_loader, val_loader)
 starting_step += 1 # to make sure step 80 isn't repeated

 logger, run_name = resume_logging(project_name.split("/")[1], wandb_id, args)
 print(f"Resuming from checkpoint: {checkpoint_path}")
 print(f"Weights and Biases run name: {run_name}")
 print(f"Resuming from step: {starting_step}")

Initialize HuggingFace repository <-- UNSURE IF NEEDED
if master_process:
 new_branch_name = run_name
 # create_branch(project_name, repo_type="model", branch=new_branch_name)
 hf_repo = Repository("./", clone_from=project_name, revision=run_name)

 # Local subprocess for git pulling and checking out the newest branch
 if master_process:
 import subprocess
 subprocess.run(["git", "fetch", "origin"])
 subprocess.run(["git", "checkout", new_branch_name])
 subprocess.run(["git", "pull", "origin", new_branch_name])

 if master_process:
 print(f"Resuming from checkpoint at step {starting_step}")
 else:
 starting_step = 0
 logger, run_name, wandb_id = setup_logging(project_name.split("/")[1], args)
 print(f"Weights and Biases run name: {run_name}")
 print(f"Weights and Biases run id : {wandb_id}")

Initialize HuggingFace repository
if master_process:
 new_branch_name = run_name
 create_branch(project_name, repo_type="model", branch=new_branch_name)
 hf_repo = Repository("./", clone_from=project_name, revision=run_name)

 if master_process:
 print(f"Starting new run: {run_name}")

Training Loop

```

```

for step in range(starting_step, max_steps):
 t0 = time.time()
 last_step = (step == max_steps - 1)

 # once in a while evaluate our validation loss
 if step % args.val_every == 0 or last_step:
 model.eval()
 val_loader.reset()
 with torch.no_grad():
 val_loss_accum = 0.0
 val_loss_steps = args.max_eval_steps
 for _ in range(val_loss_steps):
 x, y = val_loader.next_batch()
 x, y = x.to(device), y.to(device)
 with torch.autocast(device_type=device_type, dtype=torch.bfloat16):
 logits, loss = model(x, y)
 loss = loss / val_loss_steps
 val_loss_accum += loss.detach()

 if ddp:
 dist.all_reduce(val_loss_accum, op=dist.ReduceOp.AVG)

 # once in a while evaluate hellaswag
 if (step % args.hellaswag_every == 0 or last_step) and (not use_compile):
 num_correct_norm = 0
 num_total = 0
 for i, example in enumerate(iterate_examples("val")):
 # only process examples where i % ddp_world_size == ddp_rank
 if i % ddp_world_size != ddp_rank:
 continue
 # render the example into tokens and labels
 _, tokens, mask, label = render_example(example)
 tokens = tokens.to(device)
 mask = mask.to(device)
 # get the logits
 with torch.no_grad():
 with torch.autocast(device_type=device_type, dtype=torch.bfloat16):
 logits, loss = model(tokens)
 pred_norm = get_most_likely_row(tokens, mask, logits)
 num_total += 1
 num_correct_norm += int(pred_norm == label)
 # reduce the stats across all processes
 if ddp:
 num_total = torch.tensor(num_total, dtype=torch.long, device=device)
 num_correct_norm = torch.tensor(num_correct_norm, dtype=torch.long, device=device)
 dist.all_reduce(num_total, op=dist.ReduceOp.SUM)
 dist.all_reduce(num_correct_norm, op=dist.ReduceOp.SUM)
 num_total = num_total.item()
 num_correct_norm = num_correct_norm.item()
 acc_norm = num_correct_norm / num_total
 if master_process:
 log_metrics({"hella/swag": acc_norm, "hella/correct_norm": num_correct_norm, "hella/num total": num_total})
 print(f"HellaSwag accuracy: {num_correct_norm}/{num_total}={acc_norm:.4f}")
 with open(log_file, "a") as f:
 f.write(f"{step} hella {acc_norm:.4f}\n")

 # once in a while generate from the model (except step 0, which is noise)
 if ((step > 0 and step % args.generate_every == 0) or last_step) and (not use_compile):
 model.eval()
 num_return_sequences = args.num_return_sequences
 max_length = args.max_generate_length
 tokens = enc.encode("Hello, I'm a language model,")

```

```

tokens = torch.tensor(tokens, dtype=torch.long)
tokens = tokens.unsqueeze(0).repeat(num_return_sequences, 1)
xgen = tokens.to(device)
sample_rng = torch.Generator(device=device)
sample_rng.manual_seed(42 + ddp_rank)
while xgen.size(1) < max_length:
 # forward the model to get the logits
 with torch.no_grad():
 with torch.autocast(device_type=device_type, dtype=torch.bfloat16):
 logits, loss = model(xgen) # (B, T, vocab_size)
 # take the logits at the last position
 logits = logits[:, -1, :] # (B, vocab_size)
 # get the probabilities
 probs = F.softmax(logits, dim=-1)
 # do top-k sampling of 50 (huggingface pipeline default)
 # topk_probs here becomes (5, 50), topk_indices is (5, 50)
 topk_probs, topk_indices = torch.topk(probs, args.top_k, dim=-1) # top_k == 50 in this case
 # select a token from the top-k probabilities
 # note: multinomial does not demand the input to sum to 1
 ix = torch.multinomial(topk_probs, 1, generator=sample_rng) # (B, 1)
 # gather the corresponding indices
 xcol = torch.gather(topk_indices, -1, ix) # (B, 1)
 # append to the sequence
 xgen = torch.cat((xgen, xcol), dim=1)
 # print the generated text
 for i in range(num_return_sequences):
 tokens = xgen[i, :max_length].tolist()
 decoded = enc.decode(tokens)
 print(f"rank {ddp_rank} sample {i}: {decoded}")

if step % args.save_every == 0 or last_step:
 if master_process:
 # Save checkpoint and push to HuggingFace
 checkpoint_path = save_checkpoint(
 raw_model, optimizer, step, val_loss_accum.item(), run_name,
 train_loader.get_state(), val_loader.get_state(), wandb_id
)
 hf_repo.push_to_hub(commit_message=f"Checkpoint at step {step}")
 print(f"Saved checkpoint and pushed to HuggingFace at step {step}")
 log_metrics({"loss/validation": val_loss_accum.item()})

do one step of the optimization
model.train()
optimizer.zero_grad()
loss_accum = 0.0
for micro_step in range(grad_accum_steps):
 x, y = train_loader.next_batch()
 x, y = x.to(device), y.to(device)
 # added after video, this field is also used by the forward pass.
 if ddp:
 model.require_backward_grad_sync = (micro_step == grad_accum_steps - 1)
 with torch.autocast(device_type=device_type, dtype=torch.bfloat16):
 logits, loss = model(x, y)
 # we have to scale the loss to account for gradient accumulation,
 # because the gradients just add on each successive backward().
 # addition of gradients corresponds to a SUM in the objective, but
 # instead of a SUM we want MEAN. Scale the loss here so it comes out right
 loss = loss / grad_accum_steps
 loss_accum += loss.detach()
 loss.backward()
if ddp:
 dist.all_reduce(loss_accum, op=dist.ReduceOp.AVG)

```

```

norm = torch.nn.utils.clip_grad_norm_(model.parameters(), args.grad_clip) # grad_clip == 1.0 by default
determine and set the learning rate for this iteration
lr = get_lr(step)
for param_group in optimizer.param_groups:
 param_group['lr'] = lr
optimizer.step()
if device_type == "cuda":
 torch.cuda.synchronize() # wait for the GPU to finish work
t1 = time.time()
dt = t1 - t0 # time difference in seconds
tokens_processed = train_loader.B * train_loader.T * grad_accum_steps * ddp_world_size
tokens_per_sec = tokens_processed / dt
if master_process:
 log_metrics({
 "lr": lr, # get_lr()
 "samples": step * samples_per_step,
 "steps": step,
 "loss/train": loss_accum.item(),
 # file specific addition
 "global gradient norm": norm,
 "dt": dt,
 "tok per sec": tokens_per_sec
 })
 print(f"step {step:5d} | loss: {loss_accum.item():.6f} | lr {lr:.4e} | norm: {norm:.4f} | dt: {dt*1000:.2f}ms | tok/sec: {tokens_per_sec:.2f}")
 with open(log_file, "a") as f:
 f.write(f"{step} train {loss_accum.item():.6f}\n")

if ddp:
 destroy_process_group()

Final save and push
if master_process:
 final_checkpoint_path = checkpoint_path = save_checkpoint(raw_model, optimizer, step,
val_loss_accum.item(), run_name, train_loader.get_state(), val_loader.get_state(), wandb_id)
 hf_repo.push_to_hub(commit_message="Final model")
 print("Training completed. Final model pushed to HuggingFace.")

```

## D.2 Root Mean Square normalisation pretraining

```

import os
import math
import time
import inspect
from dataclasses import dataclass
import torch
import torch.nn as nn
from torch.nn import functional as F
from hellaswag import render_example, iterate_examples

from argparse import Namespace

import wandb

import tiktoken
import numpy as np

from torch.distributed import init_process_group, destroy_process_group
from torch.nn.parallel import DistributedDataParallel as DDP
import torch.distributed as dist

```

```

from huggingface_hub import Repository, create_branch
import logging
import glob

class CausalSelfAttention(nn.Module):

 def __init__(self, config):
 super().__init__()
 assert config.n_embd % config.n_head == 0
 # key, query, value projections for all heads, but in a batch
 self.c_attn = nn.Linear(config.n_embd, 3 * config.n_embd)
 # output projection
 self.c_proj = nn.Linear(config.n_embd, config.n_embd)
 self.c_proj.NANOGLPT_SCALE_INIT = 1
 # regularization
 self.n_head = config.n_head
 self.n_embd = config.n_embd

 def forward(self, x):
 B, T, C = x.size() # batch size, sequence length, embedding dimensionality (n_embd)
 # calculate query, key, values for all heads in batch and move head forward to be the batch dim
 # nh is "number of heads", hs is "head size", and C (number of channels) = nh * hs
 # e.g. in GPT-2 (124M), n_head=12, hs=64, so nh*hs=C=768 channels in the Transformer
 qkv = self.c_attn(x)
 q, k, v = qkv.split(self.n_embd, dim=2)
 k = k.view(B, T, self.n_head, C // self.n_head).transpose(1, 2) # (B, nh, T, hs)
 q = q.view(B, T, self.n_head, C // self.n_head).transpose(1, 2) # (B, nh, T, hs)
 v = v.view(B, T, self.n_head, C // self.n_head).transpose(1, 2) # (B, nh, T, hs)
 y = F.scaled_dot_product_attention(q, k, v, is_causal=True) # flash attention
 y = y.transpose(1, 2).contiguous().view(B, T, C) # re-assemble all head outputs side by side
 # output projection
 y = self.c_proj(y)
 return y

class MLP(nn.Module):

 def __init__(self, config):
 super().__init__()
 self.c_fc = nn.Linear(config.n_embd, 4 * config.n_embd)
 self.gelu = nn.GELU(approximate='tanh')
 self.c_proj = nn.Linear(4 * config.n_embd, config.n_embd)
 self.c_proj.NANOGLPT_SCALE_INIT = 1

 def forward(self, x):
 x = self.c_fc(x)
 x = self.gelu(x)
 x = self.c_proj(x)
 return x

from torch.nn.parameter import Parameter
from torch import Tensor, Size
from typing import Union, List, Tuple, Optional
from torch.nn import init
import numbers

_shape_t = Union[int, List[int], Size]

```

```

class RMSNorm(nn.Module):
 __constants__ = ['normalized_shape', 'eps', 'elementwise_affine']
 normalized_shape: Tuple[int, ...]
 eps: float
 elementwise_affine: bool

 def __init__(self, normalized_shape: Union[int, Tuple[int, ...]], eps: float = 1e-5, elementwise_affine:
bool = True, device=None, dtype=None) -> None:
 factory_kwargs = {'device': device, 'dtype': dtype}
 super().__init__()
 if isinstance(normalized_shape, numbers.Integral):
 normalized_shape = (normalized_shape,)
 self.normalized_shape = tuple(normalized_shape)
 self.eps = eps
 self.elementwise_affine = elementwise_affine
 if self.elementwise_affine:
 self.gain = nn.Parameter(torch.empty(self.normalized_shape, **factory_kwargs))
 else:
 self.register_parameter('gain', None)

 self.reset_parameters()

 def reset_parameters(self) -> None:
 if self.elementwise_affine:
 nn.init.ones_(self.gain)

 def forward(self, x: Tensor) -> Tensor:
 """Forward pass of RMSNorm."""
 mean_square = torch.mean(torch.square(x), dim=-1, keepdim=True)
 x_norm = x * torch.rsqrt(mean_square + self.eps)
 if self.elementwise_affine:
 x_norm = self.gain * x_norm
 return x_norm

 def extra_repr(self) -> str:
 return '{normalized_shape}, eps={eps},\n' \
 'elementwise_affine={elementwise_affine}'.format(**self.__dict__)

class Block(nn.Module):

 def __init__(self, config):
 super().__init__()
 self.ln_1 = RMSNorm(config.n_embd)
 self.attn = CausalSelfAttention(config)
 self.ln_2 = RMSNorm(config.n_embd)
 self.mlp = MLP(config)

 def forward(self, x):
 x = x + self.attn(self.ln_1(x))
 x = x + self.mlp(self.ln_2(x))
 return x

@dataclass
class GPTConfig:
 block_size: int = 1024 # max sequence length
 vocab_size: int = 50257 # number of tokens: 50,000 BPE merges + 256 bytes tokens + 1 <|endoftext|> token
 n_layer: int = 12 # number of layers

```

```

n_head: int = 12 # number of heads
n_embd: int = 768 # embedding dimension

class GPT(nn.Module):

 def __init__(self, config):
 super().__init__()
 self.config = config

 self.transformer = nn.ModuleDict(dict(
 wte = nn.Embedding(config.vocab_size, config.n_embd),
 wpe = nn.Embedding(config.block_size, config.n_embd),
 h = nn.ModuleList([Block(config) for _ in range(config.n_layer)]),
 ln_f = RMSNorm(config.n_embd),
))
 self.lm_head = nn.Linear(config.n_embd, config.vocab_size, bias=False)

 # weight sharing scheme
 self.transformer.wte.weight = self.lm_head.weight

 # init params
 self.apply(self._init_weights)

 def _init_weights(self, module):
 if isinstance(module, nn.Linear):
 std = 0.02
 if hasattr(module, 'NANO_GPT_SCALE_INIT'):
 std *= (2 * self.config.n_layer) ** -0.5
 torch.nn.init.normal_(module.weight, mean=0.0, std=std)
 if module.bias is not None:
 torch.nn.init.zeros_(module.bias)
 elif isinstance(module, nn.Embedding):
 torch.nn.init.normal_(module.weight, mean=0.0, std=0.02)

 def forward(self, idx, targets=None):
 # idx is of shape (B, T)
 B, T = idx.size()
 assert T <= self.config.block_size, f"Cannot forward sequence of length {T}, block size is only {self.config.block_size}"
 # forward the token and position embeddings
 pos = torch.arange(0, T, dtype=torch.long, device=idx.device) # shape (T)
 pos_emb = self.transformer.wpe(pos) # position embeddings of shape (T, n_embd)
 tok_emb = self.transformer.wte(idx) # token embeddings of shape (B, T, n_embd)
 x = tok_emb + pos_emb
 # forward the blocks of the transformer
 for block in self.transformer.h:
 x = block(x)
 # forward the final RMSNorm and the classifier
 x = self.transformer.ln_f(x)
 logits = self.lm_head(x) # (B, T, vocab_size)
 loss = None
 if targets is not None:
 loss = F.cross_entropy(logits.view(-1, logits.size(-1)), targets.view(-1))
 return logits, loss

 @classmethod
 def from_pretrained(cls, model_type):
 """Loads pretrained GPT-2 model weights from huggingface"""
 assert model_type in {'gpt2', 'gpt2-medium', 'gpt2-large', 'gpt2-xl'}
 from transformers import GPT2LMHeadModel
 print("loading weights from pretrained gpt: %s" % model_type)

```

```

n_layer, n_head and n_embd are determined from model_type
config_args = {
 'gpt2': dict(n_layer=12, n_head=12, n_embd=768), # 124M params
 'gpt2-medium': dict(n_layer=24, n_head=16, n_embd=1024), # 350M params
 'gpt2-large': dict(n_layer=36, n_head=20, n_embd=1280), # 774M params
 'gpt2-xl': dict(n_layer=48, n_head=25, n_embd=1600), # 1558M params
}[model_type]
config_args['vocab_size'] = 50257 # always 50257 for GPT model checkpoints
config_args['block_size'] = 1024 # always 1024 for GPT model checkpoints
create a from-scratch initialized minGPT model
config = GPTConfig(**config_args)
model = GPT(config)
sd = model.state_dict()
sd_keys = sd.keys()
sd_keys = [k for k in sd_keys if not k.endswith('.attn.bias')] # discard this mask / buffer, not a
param

init a huggingface/transformers model
model_hf = GPT2LMHeadModel.from_pretrained(model_type)
sd_hf = model_hf.state_dict()

copy while ensuring all of the parameters are aligned and match in names and shapes
sd_keys_hf = sd_hf.keys()
sd_keys_hf = [k for k in sd_keys_hf if not k.endswith('.attn.masked_bias')] # ignore these, just a
buffer
sd_keys_hf = [k for k in sd_keys_hf if not k.endswith('.attn.bias')] # same, just the mask (buffer)
transposed = ['attn.c_attn.weight', 'attn.c_proj.weight', 'mlp.c_fc.weight', 'mlp.c_proj.weight']
basically the openai checkpoints use a "Conv1D" module, but we only want to use a vanilla Linear
this means that we have to transpose these weights when we import them
assert len(sd_keys_hf) == len(sd_keys), f"mismatched keys: {len(sd_keys_hf)} != {len(sd_keys)}"
for k in sd_keys_hf:
 if any(k.endswith(w) for w in transposed):
 # special treatment for the Conv1D weights we need to transpose
 assert sd_hf[k].shape[::-1] == sd[k].shape
 with torch.no_grad():
 sd[k].copy_(sd_hf[k].t())
 else:
 # vanilla copy over the other parameters
 assert sd_hf[k].shape == sd[k].shape
 with torch.no_grad():
 sd[k].copy_(sd_hf[k])

return model

def configure_optimizers(self, weight_decay, learning_rate, device_type):
 # start with all of the candidate parameters (that require grad)
 param_dict = {pn: p for pn, p in self.named_parameters()}
 param_dict = {pn: p for pn, p in param_dict.items() if p.requires_grad}
 # create optim groups. Any parameters that is 2D will be weight decayed, otherwise no.
 # i.e. all weight tensors in matmul + embeddings decay, all biases and RMSNorms don't.
 decay_params = [p for n, p in param_dict.items() if p.dim() >= 2]
 nodecay_params = [p for n, p in param_dict.items() if p.dim() < 2]
 optim_groups = [
 {'params': decay_params, 'weight_decay': weight_decay},
 {'params': nodecay_params, 'weight_decay': 0.0}
]
 num_decay_params = sum(p.numel() for p in decay_params)
 num_nodecay_params = sum(p.numel() for p in nodecay_params)
 if master_process:
 print(f"num decayed parameter tensors: {len(decay_params)}, with {num_decay_params:,} parameters")
 print(f"num non-decayed parameter tensors: {len(nodecay_params)}, with {num_nodecay_params:,} parameters")

```

```

parameters")
 # Create AdamW optimizer and use the fused version if it is available
 fused_available = 'fused' in inspect.signature(torch.optim.AdamW).parameters
 use_fused = fused_available and device_type == "cuda"
 if master_process:
 print(f"using fused AdamW: {use_fused}")
 optimizer = torch.optim.AdamW(optim_groups, lr=learning_rate, betas=(0.9, 0.95), eps=1e-8,
fused=use_fused)
 return optimizer

def load_tokens(filename):
 npt = np.load(filename)
 npt = npt.astype(np.int32) # added after video
 ptt = torch.tensor(npt, dtype=torch.long)
 return ptt

class DataLoaderLite:
 def __init__(self, B, T, process_rank, num_processes, split):
 self.B = B
 self.T = T
 self.process_rank = process_rank
 self.num_processes = num_processes
 assert split in {'train', 'val'}

 # get the shard filenames
 data_root = "../../edu_fineweb10B"
 shards = os.listdir(data_root)
 shards = [s for s in shards if split in s] # listing out shards file in the data_root dir
 shards = sorted(shards)
 shards = [os.path.join(data_root, s) for s in shards]
 self.shards = shards
 assert len(shards) > 0, f"no shards found for split {split}"
 if master_process:
 print(f"found {len(shards)} shards for split {split}")

 # NEW IMPL
 self.current_shard = 0
 self.current_position = self.B * self.T * self.process_rank

 self.reset()

 def reset(self):
 # state, init at shard zero
 self.current_shard = 0
 self.tokens = load_tokens(self.shards[self.current_shard])
 self.current_position = self.B * self.T * self.process_rank

 def set_state(self, state):
 self.current_shard = state['current_shard']
 self.current_position = state['current_position']
 self.tokens = load_tokens(self.shards[self.current_shard])

 def get_state(self):
 return {
 'current_shard': self.current_shard,
 'current_position': self.current_position
 }

 def next_batch(self):
 B, T = self.B, self.T
 buf = self.tokens[self.current_position : self.current_position+B*T+1]

```

```

x = (buf[:-1]).view(B, T) # inputs
y = (buf[1:]).view(B, T) # targets
advance the position in the tensor
self.current_position += B * T * self.num_processes
if loading the next batch would be out of bounds, advance to next shard
if self.current_position + (B * T * self.num_processes + 1) > len(self.tokens):
 self.current_shard = (self.current_shard + 1) % len(self.shards)
 self.tokens = load_tokens(self.shards[self.current_shard])
 self.current_position = B * T * self.process_rank
return x, y

def setup_logging(project_name, args):
 logger = logging.getLogger(__name__)
 dir_name = "./log"
 os.makedirs(dir_name, exist_ok=True)
 print(f"Directory '{dir_name}' {'already exists' if os.path.exists(dir_name) else 'was created'}.")

 logging.basicConfig(
 format"%(asctime)s - %(levelname)s - %(name)s - %(message)s",
 datefmt="%m/%d/%Y %H:%M:%S",
 level=logging.INFO,
 handlers=[
 logging.FileHandler(f"log/debug_{ddp_rank}.log"),
 logging.StreamHandler(),
],
)

 if master_process:
 wandb.init(project=project_name, config=args, dir="..")
 run_name = wandb.run.name
 wandb_id = wandb.run.id
 logger.setLevel(logging.INFO)
 print(f"Starting new run: {run_name}")
 else:
 run_name = ""
 wandb_id = ""
 logger.setLevel(logging.ERROR)

 return logger, run_name, wandb_id

def resume_logging(project_name, run_id, args):
 logger = logging.getLogger(__name__)
 dir_name = "./log"
 os.makedirs(dir_name, exist_ok=True)
 print(f"Directory '{dir_name}' {'already exists' if os.path.exists(dir_name) else 'was created'}.")

 logging.basicConfig(
 format"%(asctime)s - %(levelname)s - %(name)s - %(message)s",
 datefmt="%m/%d/%Y %H:%M:%S",
 level=logging.INFO,
 handlers=[
 logging.FileHandler(f"log/debug_{ddp_rank}.log"),
 logging.StreamHandler(),
],
)

 if master_process:
 wandb.init(project=project_name, id=run_id, resume="must", config=args, dir='..')
 run_name = wandb.run.name
 logger.setLevel(logging.INFO)
 print(f"Resuming run: {run_name}")

```

```

else:
 run_name = ""
 logger.setLevel(logging.ERROR)

 return logger, run_name

def log_metrics(metrics):
 if master_process:
 wandb.log(metrics)

def save_checkpoint(model, optimizer, step, val_loss, run_name, train_loader_state, val_loader_state,
wandb_id):
 checkpoint = {
 'model': model.state_dict(),
 'optimizer': optimizer.state_dict(),
 'step': step,
 'val_loss': val_loss,
 'run_name': run_name,
 'train_loader_state': train_loader_state,
 'val_loader_state': val_loader_state,
 'wandb_id': wandb_id
 }
 checkpoint_path = os.path.join(log_dir, f"checkpoint_{step}.pt")
 torch.save(checkpoint, checkpoint_path)
 return checkpoint_path

def load_checkpoint(checkpoint_path, model, optimizer, train_loader, val_loader):
 checkpoint = torch.load(checkpoint_path)
 model.load_state_dict(checkpoint['model'])
 optimizer.load_state_dict(checkpoint['optimizer'])
 train_loader.set_state(checkpoint['train_loader_state'])
 val_loader.set_state(checkpoint['val_loader_state'])
 return checkpoint['step'], checkpoint['val_loss'], checkpoint['run_name'], checkpoint['wandb_id']

```

```

helper function for HellaSwag eval
takes tokens, mask, and logits, returns the index of the completion with the lowest loss

def get_most_likely_row(tokens, mask, logits):
 # evaluate the autoregressive loss at all positions
 shift_logits = (logits[..., :-1, :]).contiguous()
 shift_tokens = (tokens[..., 1:]).contiguous()
 flat_shift_logits = shift_logits.view(-1, shift_logits.size(-1))
 flat_shift_tokens = shift_tokens.view(-1)
 shift_losses = F.cross_entropy(flat_shift_logits, flat_shift_tokens, reduction='none')
 shift_losses = shift_losses.view(tokens.size(0), -1)
 # now get the average loss just for the completion region (where mask == 1), in each row
 shift_mask = (mask[..., 1:]).contiguous() # we must shift mask, so we start at the last prompt token
 masked_shift_losses = shift_losses * shift_mask
 # sum and divide by the number of 1s in the mask
 sum_loss = masked_shift_losses.sum(dim=1)
 avg_loss = sum_loss / shift_mask.sum(dim=1)
 # now we have a loss for each of the 4 completions
 # the one with the lowest loss should be the most likely
 pred_norm = avg_loss.argmin().item()
 return pred_norm

simple launch:

```

```

python train_gpt2.py
DDP launch for e.g. 8 GPUs:
torchrun --standalone --nproc_per_node=8 train_gpt2.py

run the training loop

set up DDP (distributed data parallel).
torchrun command sets the env variables RANK, LOCAL_RANK, and WORLD_SIZE
ddp = int(os.environ.get('RANK', -1)) != -1 # is this a ddp run?
if ddp:
 # use of DDP atm demands CUDA, we set the device appropriately according to rank
 assert torch.cuda.is_available(), "for now i think we need CUDA for DDP"
 init_process_group(backend='nccl')
 ddp_rank = int(os.environ['RANK'])
 ddp_local_rank = int(os.environ['LOCAL_RANK'])
 ddp_world_size = int(os.environ['WORLD_SIZE'])
 device = f'cuda:{ddp_local_rank}'
 torch.cuda.set_device(device)
 master_process = ddp_rank == 0 # this process will do logging, checkpointing etc.
else:
 # vanilla, non-DDP run
 ddp_rank = 0
 ddp_local_rank = 0
 ddp_world_size = 1
 master_process = True
 # attempt to autodetect device
 device = "cpu"
 if torch.cuda.is_available():
 device = "cuda"
 elif hasattr(torch.backends, "mps") and torch.backends.mps.is_available():
 device = "mps"
 print(f"using device: {device}")

GPTesla - 111M param setup in comment. Modification to make lighter training requirement needed
config = {
 "weight_decay": 0.1,
 # "lr_scheduler_type": "cosine",
 "gradient_accumulation_steps": 2**19 // (64 * 1024 * ddp_world_size), # total_batch_size // (B * T *
 ddp_world_size
 "max_eval_steps": 20,
 "seq_length": 1024,

 # New centralised parameters
 "project_name": "shng2025/GPT-Valkyrie_RMSN-124m",
 "total_batch_size": 2**19, # temporarily because 6 GPUs # 2**19, ~0.5M, in number of tokens
 "micro_batch_size": 64,
 "max_lr": 6e-4,
 "min_lr": 6e-5, # 10% of max_lr // not used, as we are using weight_decay instead
 "warmup_steps": 715,
 "max_steps": 19073,
 "val_every": 500, # EVALUATION
 "generate_every": 500, # EVALUATION
 "hellawag_every": 500, # EVALUATION
 "save_every": 2000, # SAVE CHECKPOINTING
 "log_dir": "./log",
 "device": "auto", # "auto", "cpu", "cuda", or "mps"
 "use_compile": True,
 "grad_clip": 1.0,
 "num_return_sequences": 4,
}

```

```

 "max_generate_length": 32,
 "top_k": 50,

 "resume_from_checkpoint": False,
}

args = Namespace(**config)
samples_per_step = torch.cuda.device_count() * args.micro_batch_size

project_name = args.project_name

Logging - DEPRECATED
if master_process:
 pass
 # run_name, wandb_id = setup_logging(project_name.split("/")[1])
 # print(f"Weights and Biases run name: {run_name}")
 # print(f"Weights and Biases run id : {wandb_id}")

added after video, pytorch can be serious about it's device vs. device_type distinction
device_type = "cuda" if device.startswith("cuda") else "cpu"

torch.manual_seed(1337)
if torch.cuda.is_available():
 torch.cuda.manual_seed(1337)

enc = tiktoken.get_encoding("gpt2")

total_batch_size = args.total_batch_size # 2**19, ~0.5M, in number of tokens
B = args.micro_batch_size # micro batch size
T = args.seq_length # sequence length
assert total_batch_size % (B * T * ddp_world_size) == 0, "make sure total_batch_size is divisible by B * T * ddp_world_size"
grad_accum_steps = total_batch_size // (B * T * ddp_world_size)
if master_process:
 print(f"total desired batch size: {total_batch_size}")
 print(f"=> calculated gradient accumulation steps: {grad_accum_steps}")

train_loader = DataLoaderLite(B=B, T=T, process_rank=ddp_rank, num_processes=ddp_world_size, split="train")
val_loader = DataLoaderLite(B=B, T=T, process_rank=ddp_rank, num_processes=ddp_world_size, split="val")

torch.set_float32_matmul_precision('medium')

create model
model = GPT(GPTConfig(vocab_size=50304))
model = GPT.from_pretrained("gpt2") # or init from OpenAI GPT-2
model.to(device)
use_compile = True # torch.compile interferes with HellaSwag eval and Generation. TODO fix
if use_compile:
 model = torch.compile(model)
if ddp:
 model = DDP(model, device_ids=[ddp_local_rank])
raw_model = model.module if ddp else model # always contains the "raw" unwrapped model

max_lr = args.max_lr
min_lr = max_lr * args.weight_decay
warmup_steps = args.warmup_steps
max_steps = args.max_steps # 19,073 steps is ~1 epoch, if data is 10B tokens and batch size 0.5M tokens
def get_lr(it):
 # 1) linear warmup for warmup_iters steps

```

```

if it < warmup_steps:
 return max_lr * (it+1) / warmup_steps
2) if it > lr_decay_iters, return min learning rate
if it > max_steps:
 return min_lr
3) in between, use cosine decay down to min learning rate
decay_ratio = (it - warmup_steps) / (max_steps - warmup_steps)
assert 0 <= decay_ratio <= 1
coeff = 0.5 * (1.0 + math.cos(math.pi * decay_ratio)) # coeff starts at 1 and goes to 0
return min_lr + coeff * (max_lr - min_lr)

optimize!
optimizer = raw_model.configure_optimizers(weight_decay=args.weight_decay, learning_rate=args.max_lr,
device_type=device_type)

create the log directory we will write checkpoints to and log to
log_dir = args.log_dir
os.makedirs(log_dir, exist_ok=True)
log_file = os.path.join(log_dir, "log.txt")
with open(log_file, "w") as f: # open for writing to clear the file
 pass

"""
Initialize HuggingFace repository
if master_process:
 new_branch_name = run_name
 create_branch(project_name, repo_type="model", branch=new_branch_name)
 hf_repo = Repository("./", clone_from=project_name, revision=run_name)
"""

Training loop
starting_step = 0

if args.resume_from_checkpoint:
 checkpoint_dir = args.log_dir
 checkpoint_pattern = os.path.join(checkpoint_dir, "checkpoint_*.pt")
 checkpoint_files = glob.glob(checkpoint_pattern)
 latest_checkpoint = max(checkpoint_files, key=os.path.getctime)
 checkpoint_path = latest_checkpoint

 # Use the load_checkpoint function here
 starting_step, val_loss, run_name, wandb_id = load_checkpoint(checkpoint_path, raw_model, optimizer,
train_loader, val_loader)
 starting_step += 1 # to make sure step 80 isn't repeated

 logger, run_name = resume_logging(project_name.split("/")[1], wandb_id, args)
 print(f"Resuming from checkpoint: {checkpoint_path}")
 print(f"Weights and Biases run name: {run_name}")
 print(f"Resuming from step: {starting_step}")

 # Initialize HuggingFace repository <- UNSURE IF NEEDED
 if master_process:
 new_branch_name = run_name
 # create_branch(project_name, repo_type="model", branch=new_branch_name)
 hf_repo = Repository("./", clone_from=project_name, revision=run_name)

 # Local subprocess for git pulling and checking out the newest branch
 if master_process:
 import subprocess
 subprocess.run(["git", "fetch", "origin"])
 subprocess.run(["git", "checkout", new_branch_name])
 subprocess.run(["git", "pull", "origin", new_branch_name])

```

```

if master_process:
 print(f"Resuming from checkpoint at step {starting_step}")
else:
 starting_step = 0
logger, run_name, wandb_id = setup_logging(project_name.split("/")[1], args)
print(f"Weights and Biases run name: {run_name}")
print(f"Weights and Biases run id : {wandb_id}")

Initialize HuggingFace repository
if master_process:
 new_branch_name = run_name
 create_branch(project_name, repo_type="model", branch=new_branch_name)
 hf_repo = Repository("./", clone_from=project_name, revision=run_name)

if master_process:
 print(f"Starting new run: {run_name}")

Training Loop
for step in range(starting_step, max_steps):
 t0 = time.time()
 last_step = (step == max_steps - 1)

 # once in a while evaluate our validation loss
 if step % args.val_every == 0 or last_step:
 model.eval()
 val_loader.reset()
 with torch.no_grad():
 val_loss_accum = 0.0
 val_loss_steps = args.max_eval_steps
 for _ in range(val_loss_steps):
 x, y = val_loader.next_batch()
 x, y = x.to(device), y.to(device)
 with torch.autocast(device_type=device_type, dtype=torch.bfloat16):
 logits, loss = model(x, y)
 loss = loss / val_loss_steps
 val_loss_accum += loss.detach()
 if ddp:
 dist.all_reduce(val_loss_accum, op=dist.ReduceOp.AVG)

 # once in a while evaluate hellaswag
 if (step % args.hellaswag_every == 0 or last_step) and (not use_compile):
 num_correct_norm = 0
 num_total = 0
 for i, example in enumerate(iterate_examples("val")):
 # only process examples where i % ddp_world_size == ddp_rank
 if i % ddp_world_size != ddp_rank:
 continue
 # render the example into tokens and labels
 _, tokens, mask, label = render_example(example)
 tokens = tokens.to(device)
 mask = mask.to(device)
 # get the logits
 with torch.no_grad():
 with torch.autocast(device_type=device_type, dtype=torch.bfloat16):
 logits, loss = model(tokens)
 pred_norm = get_most_likely_row(tokens, mask, logits)
 num_total += 1
 num_correct_norm += int(pred_norm == label)
 # reduce the stats across all processes
 if ddp:
 num_total = torch.tensor(num_total, dtype=torch.long, device=device)

```

```

num_correct_norm = torch.tensor(num_correct_norm, dtype=torch.long, device=device)
dist.all_reduce(num_total, op=dist.ReduceOp.SUM)
dist.all_reduce(num_correct_norm, op=dist.ReduceOp.SUM)
num_total = num_total.item()
num_correct_norm = num_correct_norm.item()
acc_norm = num_correct_norm / num_total
if master_process:
 log_metrics({"hella/swag": acc_norm, "hella/correct_norm": num_correct_norm, "hella/num total": num_total})
 print(f"HellaSwag accuracy: {num_correct_norm}/{num_total}={acc_norm:.4f}")
 with open(log_file, "a") as f:
 f.write(f"{step} hella {acc_norm:.4f}\n")

once in a while generate from the model (except step 0, which is noise)
if ((step > 0 and step % args.generate_every == 0) or last_step) and (not use_compile):
 model.eval()
 num_return_sequences = args.num_return_sequences
 max_length = args.max_generate_length
 tokens = enc.encode("Hello, I'm a language model,")
 tokens = torch.tensor(tokens, dtype=torch.long)
 tokens = tokens.unsqueeze(0).repeat(num_return_sequences, 1)
 xgen = tokens.to(device)
 sample_rng = torch.Generator(device=device)
 sample_rng.manual_seed(42 + ddp_rank)
 while xgen.size(1) < max_length:
 # forward the model to get the logits
 with torch.no_grad():
 with torch.autocast(device_type=device_type, dtype=torch.bfloat16):
 logits, loss = model(xgen) # (B, T, vocab_size)
 # take the logits at the last position
 logits = logits[:, -1, :] # (B, vocab_size)
 # get the probabilities
 probs = F.softmax(logits, dim=-1)
 # do top-k sampling of 50 (huggingface pipeline default)
 # topk_probs here becomes (5, 50), topk_indices is (5, 50)
 topk_probs, topk_indices = torch.topk(probs, args.top_k, dim=-1) # top_k == 50 in this case
 # select a token from the top-k probabilities
 # note: multinomial does not demand the input to sum to 1
 ix = torch.multinomial(topk_probs, 1, generator=sample_rng) # (B, 1)
 # gather the corresponding indices
 xcol = torch.gather(topk_indices, -1, ix) # (B, 1)
 # append to the sequence
 xgen = torch.cat((xgen, xcol), dim=1)
 # print the generated text
 for i in range(num_return_sequences):
 tokens = xgen[i, :max_length].tolist()
 decoded = enc.decode(tokens)
 print(f"rank {ddp_rank} sample {i}: {decoded}")

if step % args.save_every == 0 or last_step:
 if master_process:
 # Save checkpoint and push to HuggingFace
 checkpoint_path = save_checkpoint(
 raw_model, optimizer, step, val_loss_accum.item(), run_name,
 train_loader.get_state(), val_loader.get_state(), wandb_id
)
 hf_repo.push_to_hub(commit_message=f"Checkpoint at step {step}")
 print(f"Saved checkpoint and pushed to HuggingFace at step {step}")
 log_metrics({"loss/validation": val_loss_accum.item()})

do one step of the optimization
model.train()

```

```

optimizer.zero_grad()
loss_accum = 0.0
for micro_step in range(grad_accum_steps):
 x, y = train_loader.next_batch()
 x, y = x.to(device), y.to(device)
 # added after video, this field is also used by the forward pass.
 if ddp:
 model.require_backward_grad_sync = (micro_step == grad_accum_steps - 1)
 with torch.autocast(device_type=device_type, dtype=torch.bfloat16):
 logits, loss = model(x, y)
 # we have to scale the loss to account for gradient accumulation,
 # because the gradients just add on each successive backward().
 # addition of gradients corresponds to a SUM in the objective, but
 # instead of a SUM we want MEAN. Scale the loss here so it comes out right
 loss = loss / grad_accum_steps
 loss_accum += loss.detach()
 loss.backward()
if ddp:
 dist.all_reduce(loss_accum, op=dist.ReduceOp.AVG)
norm = torch.nn.utils.clip_grad_norm_(model.parameters(), args.grad_clip) # grad_clip == 1.0 by default
determine and set the learning rate for this iteration
lr = get_lr(step)
for param_group in optimizer.param_groups:
 param_group['lr'] = lr
optimizer.step()
if device_type == "cuda":
 torch.cuda.synchronize() # wait for the GPU to finish work
t1 = time.time()
dt = t1 - t0 # time difference in seconds
tokens_processed = train_loader.B * train_loader.T * grad_accum_steps * ddp_world_size
tokens_per_sec = tokens_processed / dt
if master_process:
 log_metrics({
 "lr": lr, # get_lr()
 "samples": step * samples_per_step,
 "steps": step,
 "loss/train": loss_accum.item(),
 # file specific addition
 "global gradient norm": norm,
 "dt": dt,
 "tok per sec": tokens_per_sec
 })
 print(f"step {step:5d} | loss: {loss_accum.item():.6f} | lr {lr:.4e} | norm: {norm:.4f} | dt: {dt*1000:.2f}ms | tok/sec: {tokens_per_sec:.2f}")
 with open(log_file, "a") as f:
 f.write(f"{step} train {loss_accum.item():.6f}\n")

if ddp:
 destroy_process_group()

Final save and push
if master_process:
 final_checkpoint_path = checkpoint_path = save_checkpoint(raw_model, optimizer, step,
val_loss_accum.item(), run_name, train_loader.get_state(), val_loader.get_state(), wandb_id)
 hf_repo.push_to_hub(commit_message="Final model")
 print("Training completed. Final model pushed to HuggingFace.")

```

### D.3 Power normalisation pretraining

```
import os
import math
import time
import inspect
from dataclasses import dataclass
import torch
import torch.nn as nn
from torch.nn import functional as F
from hellaswag import render_example, iterate_examples

from argparse import Namespace

import wandb

import tiktoken
import numpy as np

from torch.distributed import init_process_group, destroy_process_group
from torch.nn.parallel import DistributedDataParallel as DDP
import torch.distributed as dist

from huggingface_hub import Repository, create_branch
import logging
import glob

os.environ['NUMEXPR_MAX_THREADS'] = '96'

class CausalSelfAttention(nn.Module):

 def __init__(self, config):
 super().__init__()
 assert config.n_embd % config.n_head == 0
 # key, query, value projections for all heads, but in a batch
 self.c_attn = nn.Linear(config.n_embd, 3 * config.n_embd)
 # output projection
 self.c_proj = nn.Linear(config.n_embd, config.n_embd)
 self.c_proj.NANOGLPT_SCALE_INIT = 1
 # regularization
 self.n_head = config.n_head
 self.n_embd = config.n_embd

 def forward(self, x):
 B, T, C = x.size() # batch size, sequence length, embedding dimensionality (n_embd)
 # calculate query, key, values for all heads in batch and move head forward to be the batch dim
 # nh is "number of heads", hs is "head size", and C (number of channels) = nh * hs
 # e.g. in GPT-2 (124M), n_head=12, hs=64, so nh*hs=C=768 channels in the Transformer
 qkv = self.c_attn(x)
 q, k, v = qkv.split(self.n_embd, dim=2)
 k = k.view(B, T, self.n_head, C // self.n_head).transpose(1, 2) # (B, nh, T, hs)
 q = q.view(B, T, self.n_head, C // self.n_head).transpose(1, 2) # (B, nh, T, hs)
 v = v.view(B, T, self.n_head, C // self.n_head).transpose(1, 2) # (B, nh, T, hs)
 y = F.scaled_dot_product_attention(q, k, v, is_causal=True) # flash attention
 y = y.transpose(1, 2).contiguous().view(B, T, C) # re-assemble all head outputs side by side
 # output projection
 y = self.c_proj(y)
 return y

class MLP(nn.Module):
```

```

def __init__(self, config):
 super().__init__()
 self.c_fc = nn.Linear(config.n_embd, 4 * config.n_embd)
 self.gelu = nn.GELU(approximate='tanh')
 self.c_proj = nn.Linear(4 * config.n_embd, config.n_embd)
 self.c_proj.NANOGPT_SCALE_INIT = 1

def forward(self, x):
 x = self.c_fc(x)
 x = self.gelu(x)
 x = self.c_proj(x)
 return x

given up
class GroupScaling1D(nn.Module):
 def __init__(self, eps=1e-5, group_num=1):
 super(GroupScaling1D, self).__init__()
 self.eps = eps
 self.group_num = group_num

 def extra_repr(self):
 return f'eps={self.eps}, group={self.group_num}'

 def forward(self, input):
 B, T, C = input.shape
 if self.group_num == 1:
 moment2 = torch.mean(torch.square(input), dim=-1, keepdim=True)
 else:
 Cg = C // self.group_num
 input = input.view(B, T, self.group_num, Cg)
 moment2 = torch.mean(input * input, dim=-1, keepdim=True)
 moment2 = moment2.repeat(1, 1, 1, Cg).view(B, T, C)
 return input / torch.sqrt(moment2 + self.eps)

class SyncPowerFunction(torch.autograd.Function):
 @staticmethod
 def forward(ctx, x, weight, bias, running_phi, eps, afwd, abkw, ema_gz, warmup_iters, current_iter,
 process_group, affine):
 ctx.affine = affine
 ctx.eps = eps
 current_iter = current_iter.item()
 ctx.process_group = process_group
 ctx.abkw = abkw

 B, T, C = x.size()
 x2 = torch.mean(torch.square(x), dim=(0, 1))
 var = x2.view(1, 1, C)

 if current_iter <= warmup_iters:
 y = x * torch.rsqrt(var + eps) # Shape: (B, T, C)
 else:
 y = x * torch.rsqrt(running_phi + eps) # Shape: (B, T, C)

 ctx.save_for_backward(y, var, weight, ema_gz)

 if current_iter < warmup_iters:
 running_phi.copy_(running_phi * (current_iter-1)/current_iter + var/current_iter) # MISTAKE?
UNSURE <-- I think its correct, because we want to value for every feature inside dim=-1 calculated across
the batch

```

```

running_phi.copy_(afwd*running_phi + (1-afwd)*var) # MISTAKE? UNSURE

if process_group is not None: # and (current_iter % 100 or current_iter == args.max_steps): #
experimental, to try and reduce time spent accessing memory
 torch.distributed.all_reduce(running_phi, op=torch.distributed.ReduceOp.AVG, group=process_group)

if affine:
 y = weight.view(1, 1, C) * y + bias.view(1, 1, C)

return y

@staticmethod
def backward(ctx, grad_output):
 eps = ctx.eps
 abkw = ctx.abkw

 y, var, weight, ema_gz = ctx.saved_tensors

 if ctx.affine:
 g = grad_output * weight.view(1, 1, -1)
 else:
 g = grad_output

 approx_grad_g = g - (1 - abkw) * ema_gz * y
 ema_gz.add_(torch.mean(approx_grad_g * y, dim=(0, 1), keepdim=True))

 if ctx.process_group is not None: # and (current_iter % 100 or current_iter == args.max_steps): #
experimental, to try and reduce time spent accessing memory
 dist.all_reduce(ema_gz, op=dist.ReduceOp.AVG, group=ctx.process_group)

 gx = torch.rsqrt(var + eps) * approx_grad_g

 if ctx.affine:
 grad_weight = torch.sum(grad_output * y, dim=(0, 1))
 grad_bias = torch.sum(grad_output, dim=(0, 1))
 if ctx.process_group is not None:
 dist.all_reduce(grad_weight, op=dist.ReduceOp.AVG, group=ctx.process_group)
 dist.all_reduce(grad_bias, op=dist.ReduceOp.AVG, group=ctx.process_group)
 else:
 grad_weight = None
 grad_bias = None

 return gx, grad_weight, grad_bias, None, None, None, None, None, None, None, None, None, None

class SyncPowerNorm(nn.Module):
 def __init__(self, num_features, eps=1e-3, alpha_fwd=0.9, alpha_bkw=0.9,
 affine=True, warmup_iters=715, process_group=None, group_num=1):
 super().__init__()

 self.num_features = num_features
 self.eps = eps
 self.affine = affine
 self.process_group = process_group

 if self.affine:
 self.weight = nn.Parameter(torch.ones(num_features))
 self.bias = nn.Parameter(torch.zeros(num_features))
 else:
 self.weight = None
 self.bias = None

 self.register_buffer('running_phi', torch.ones(1, 1, num_features)) # I THINK THIS IS LIKELY THE

```

```

ISSUE, ITS NOT MEANT TO BE A PARAMETER BUT WE ARE TREATING IT LIKE ONE???
 self.register_buffer('ema_gz', torch.zeros(1, 1, num_features))
 self.register_buffer('iters', torch.ones(1).type(torch.LongTensor))

 self.afwd = alpha_fwd
 self.abkw = alpha_bkw

 self.warmup_iters = warmup_iters
 self.grad_accum_steps = args.gradient_accumulation_steps
 self.accum_count = 0
 self.group_scaling = GroupScaling1D(eps=eps, group_num=group_num)

def extra_repr(self):
 return f'{self.num_features}, eps={self.eps}, alpha_fwd={self.afwd}, alpha_bkw={self.abkw}, '\
 f'affiliate={self.affiliate}, warmup={self.warmup_iters}'

def forward(self, input):
 B, T, C = input.size()
 assert C == self.num_features, f"Input features {C} doesn't match num_features {self.num_features}""

 # Apply GroupScaling1D
 input = self.group_scaling(input)

 if self.training:
 self.accum_count += 1
 if self.accum_count >= self.grad_accum_steps:
 self.iters.add_(1)
 self.accum_count = 0

 output = SyncPowerFunction.apply(input, self.weight if self.affiliate else None, self.bias if
self.affiliate else None,
 self.running_phi, self.eps, self.afwd, self.abkw, self.ema_gz,
self.warmup_iters, self.iters,
 self.process_group, self.affiliate)

 else:
 # var = self.running_phi
 output = input * torch.rsqrt(self.running_phi + self.eps)
 if self.affiliate: # if not, do nothing.
 output = self.weight.reshape(1, 1, C) * output + self.bias.reshape(1, 1, C)

 return output # Shape: (B, T, C)

class Block(nn.Module):

 def __init__(self, config):
 super().__init__()
 self.ln_1 = SyncPowerNorm(config.n_embd)
 self.attn = CausalSelfAttention(config)
 self.ln_2 = SyncPowerNorm(config.n_embd)
 self.mlp = MLP(config)

 def forward(self, x):
 x = x + self.attn(self.ln_1(x))
 x = x + self.mlp(self.ln_2(x))
 return x

@dataclass
class GPTConfig:

```

```

block_size: int = 1024 # max sequence length
vocab_size: int = 50257 # number of tokens: 50,000 BPE merges + 256 bytes tokens + 1 <|endoftext|> token
n_layer: int = 12 # number of layers
n_head: int = 12 # number of heads
n_embd: int = 768 # embedding dimension

class GPT(nn.Module):

 def __init__(self, config):
 super().__init__()
 self.config = config

 self.transformer = nn.ModuleDict(dict(
 wte = nn.Embedding(config.vocab_size, config.n_embd),
 wpe = nn.Embedding(config.block_size, config.n_embd),
 h = nn.ModuleList([Block(config) for _ in range(config.n_layer)]),
 ln_f = SyncPowerNorm(config.n_embd),
))
 self.lm_head = nn.Linear(config.n_embd, config.vocab_size, bias=False)

 # weight sharing scheme
 self.transformer.wte.weight = self.lm_head.weight

 # init params
 self.apply(self._init_weights)

 def _init_weights(self, module):
 if isinstance(module, nn.Linear):
 std = 0.02
 if hasattr(module, 'NANOGLT_SCALE_INIT'):
 std *= (2 * self.config.n_layer) ** -0.5
 torch.nn.init.normal_(module.weight, mean=0.0, std=std)
 if module.bias is not None:
 torch.nn.init.zeros_(module.bias)
 elif isinstance(module, nn.Embedding):
 torch.nn.init.normal_(module.weight, mean=0.0, std=0.02)

 def forward(self, idx, targets=None):
 # idx is of shape (B, T)
 B, T = idx.size()
 assert T <= self.config.block_size, f"Cannot forward sequence of length {T}, block size is only {self.config.block_size}"
 # forward the token and position embeddings
 pos = torch.arange(0, T, dtype=torch.long, device=idx.device) # shape (T)
 pos_emb = self.transformer.wpe(pos) # position embeddings of shape (T, n_embd)
 tok_emb = self.transformer.wte(idx) # token embeddings of shape (B, T, n_embd)
 x = tok_emb + pos_emb
 # forward the blocks of the transformer
 for block in self.transformer.h:
 x = block(x)
 # forward the final layernorm and the classifier
 x = self.transformer.ln_f(x)
 logits = self.lm_head(x) # (B, T, vocab_size)
 loss = None
 if targets is not None:
 loss = F.cross_entropy(logits.view(-1, logits.size(-1)), targets.view(-1))
 return logits, loss

 @classmethod
 def from_pretrained(cls, model_type):
 """Loads pretrained GPT-2 model weights from huggingface"""

```

```

assert model_type in {'gpt2', 'gpt2-medium', 'gpt2-large', 'gpt2-xl'}
from transformers import GPT2LMHeadModel
print("loading weights from pretrained gpt: %s" % model_type)

n_layer, n_head and n_embd are determined from model_type
config_args = {
 'gpt2': dict(n_layer=12, n_head=12, n_embd=768), # 124M params
 'gpt2-medium': dict(n_layer=24, n_head=16, n_embd=1024), # 350M params
 'gpt2-large': dict(n_layer=36, n_head=20, n_embd=1280), # 774M params
 'gpt2-xl': dict(n_layer=48, n_head=25, n_embd=1600), # 1558M params
}[model_type]
config_args['vocab_size'] = 50257 # always 50257 for GPT model checkpoints
config_args['block_size'] = 1024 # always 1024 for GPT model checkpoints
create a from-scratch initialized minGPT model
config = GPTConfig(**config_args)
model = GPT(config)
sd = model.state_dict()
sd_keys = sd.keys()
sd_keys = [k for k in sd_keys if not k.endswith('.attn.bias')] # discard this mask / buffer, not a
param

init a huggingface/transformers model
model_hf = GPT2LMHeadModel.from_pretrained(model_type)
sd_hf = model_hf.state_dict()

copy while ensuring all of the parameters are aligned and match in names and shapes
sd_keys_hf = sd_hf.keys()
sd_keys_hf = [k for k in sd_keys_hf if not k.endswith('.attn.masked_bias')] # ignore these, just a
buffer
sd_keys_hf = [k for k in sd_keys_hf if not k.endswith('.attn.bias')] # same, just the mask (buffer)
transposed = ['attn.c_attn.weight', 'attn.c_proj.weight', 'mlp.c_fc.weight', 'mlp.c_proj.weight']
basically the openai checkpoints use a "Conv1D" module, but we only want to use a vanilla Linear
this means that we have to transpose these weights when we import them
assert len(sd_keys_hf) == len(sd_keys), f"mismatched keys: {len(sd_keys_hf)} != {len(sd_keys)}"
for k in sd_keys_hf:
 if any(k.endswith(w) for w in transposed):
 # special treatment for the Conv1D weights we need to transpose
 assert sd_hf[k].shape[::-1] == sd[k].shape
 with torch.no_grad():
 sd[k].copy_(sd_hf[k].t())
 else:
 # vanilla copy over the other parameters
 assert sd_hf[k].shape == sd[k].shape
 with torch.no_grad():
 sd[k].copy_(sd_hf[k])

return model

def configure_optimizers(self, weight_decay, learning_rate, device_type):
 # start with all of the candidate parameters (that require grad)
 param_dict = {pn: p for pn, p in self.named_parameters()}
 param_dict = {pn: p for pn, p in param_dict.items() if p.requires_grad}
 # create optim groups. Any parameters that is 2D will be weight decayed, otherwise no.
 # i.e. all weight tensors in matmul + embeddings decay, all biases and layernorms don't.
 decay_params = [p for n, p in param_dict.items() if p.dim() >= 2]
 nodecay_params = [p for n, p in param_dict.items() if p.dim() < 2]
 optim_groups = [
 {'params': decay_params, 'weight_decay': weight_decay},
 {'params': nodecay_params, 'weight_decay': 0.0}
]
 num_decay_params = sum(p.numel() for p in decay_params)
 num_nodecay_params = sum(p.numel() for p in nodecay_params)

```

```

if master_process:
 print(f"num decayed parameter tensors: {len(decay_params)}, with {num_decay_params:,} parameters")
 print(f"num non-decayed parameter tensors: {len(nodecay_params)}, with {num_noddecay_params:,} parameters")
Create AdamW optimizer and use the fused version if it is available
fused_available = 'fused' in inspect.signature(torch.optim.AdamW).parameters
use_fused = fused_available and device_type == "cuda"
if master_process:
 print(f"using fused AdamW: {use_fused}")
optimizer = torch.optim.AdamW(optim_groups, lr=learning_rate, betas=(0.9, 0.95), eps=1e-8,
fused=use_fused)
return optimizer

def load_tokens(filename):
 npt = np.load(filename)
 npt = npt.astype(np.int32) # added after video
 ptt = torch.tensor(npt, dtype=torch.long)
 return ptt

class DataLoaderLite:
 def __init__(self, B, T, process_rank, num_processes, split):
 self.B = B
 self.T = T
 self.process_rank = process_rank
 self.num_processes = num_processes
 assert split in {'train', 'val'}

 # get the shard filenames
 data_root = "../../edu_fineweb10B"
 shards = os.listdir(data_root)
 shards = [s for s in shards if split in s] # listing out shards file in the data_root dir
 shards = sorted(shards)
 shards = [os.path.join(data_root, s) for s in shards]
 self.shards = shards
 assert len(shards) > 0, f"no shards found for split {split}"
 if master_process:
 print(f"found {len(shards)} shards for split {split}")

 # NEW IMPL
 self.current_shard = 0
 self.current_position = self.B * self.T * self.process_rank

 self.reset()

 def reset(self):
 # state, init at shard zero
 self.current_shard = 0
 self.tokens = load_tokens(self.shards[self.current_shard])
 self.current_position = self.B * self.T * self.process_rank

 def set_state(self, state):
 self.current_shard = state['current_shard']
 self.current_position = state['current_position']
 self.tokens = load_tokens(self.shards[self.current_shard])

 def get_state(self):
 return {
 'current_shard': self.current_shard,
 'current_position': self.current_position
 }

```

```

def next_batch(self):
 B, T = self.B, self.T
 buf = self.tokens[self.current_position : self.current_position+B*T+1]
 x = (buf[:-1]).view(B, T) # inputs
 y = (buf[1:]).view(B, T) # targets
 # advance the position in the tensor
 self.current_position += B * T * self.num_processes
 # if loading the next batch would be out of bounds, advance to next shard
 if self.current_position + (B * T * self.num_processes + 1) > len(self.tokens):
 self.current_shard = (self.current_shard + 1) % len(self.shards)
 self.tokens = load_tokens(self.shards[self.current_shard])
 self.current_position = B * T * self.process_rank
 return x, y

def setup_logging(project_name, args):
 logger = logging.getLogger(__name__)
 dir_name = "./log"
 os.makedirs(dir_name, exist_ok=True)
 print(f"Directory '{dir_name}' {'already exists' if os.path.exists(dir_name) else 'was created'}.")

 logging.basicConfig(
 format"%(asctime)s - %(levelname)s - %(name)s - %(message)s",
 datefmt="%m/%d/%Y %H:%M:%S",
 level=logging.INFO,
 handlers=[
 logging.FileHandler(f"log/debug_{ddp_rank}.log"),
 logging.StreamHandler(),
],
)

 if master_process:
 wandb.init(project=project_name, config=args, dir="..")
 run_name = wandb.run.name
 wandb_id = wandb.run.id
 logger.setLevel(logging.INFO)
 print(f"Starting new run: {run_name}")
 else:
 run_name = ""
 wandb_id = ""
 logger.setLevel(logging.ERROR)

 return logger, run_name, wandb_id

def resume_logging(project_name, run_id, args):
 logger = logging.getLogger(__name__)
 dir_name = "./log"
 os.makedirs(dir_name, exist_ok=True)
 print(f"Directory '{dir_name}' {'already exists' if os.path.exists(dir_name) else 'was created'}.")

 logging.basicConfig(
 format"%(asctime)s - %(levelname)s - %(name)s - %(message)s",
 datefmt="%m/%d/%Y %H:%M:%S",
 level=logging.INFO,
 handlers=[
 logging.FileHandler(f"log/debug_{ddp_rank}.log"),
 logging.StreamHandler(),
],
)

 if master_process:
 wandb.init(project=project_name, id=run_id, resume="must", config=args, dir='..')

```

```

run_name = wandb.run.name
logger.setLevel(logging.INFO)
print(f'Resuming run: {run_name}')
else:
 run_name = ""
 logger.setLevel(logging.ERROR)

return logger, run_name

def log_metrics(metrics):
 if master_process:
 wandb.log(metrics)

def save_checkpoint(model, optimizer, step, val_loss, run_name, train_loader_state, val_loader_state,
wandb_id):
 checkpoint = {
 'model': model.state_dict(),
 'optimizer': optimizer.state_dict(),
 'step': step,
 'val_loss': val_loss,
 'run_name': run_name,
 'train_loader_state': train_loader_state,
 'val_loader_state': val_loader_state,
 'wandb_id': wandb_id,
 }

 checkpoint_path = os.path.join(log_dir, f"checkpoint_{step}.pt")
 torch.save(checkpoint, checkpoint_path)
 return checkpoint_path

def load_checkpoint(checkpoint_path, model, optimizer, train_loader, val_loader):
 checkpoint = torch.load(checkpoint_path)
 model.load_state_dict(checkpoint['model'])
 optimizer.load_state_dict(checkpoint['optimizer'])
 train_loader.set_state(checkpoint['train_loader_state'])
 val_loader.set_state(checkpoint['val_loader_state'])

 return checkpoint['step'], checkpoint['val_loss'], checkpoint['run_name'], checkpoint['wandb_id']

logging powernorm stats
def log_pownorm_stats(model):
 stats = {}
 for name, module in model.named_modules():
 if isinstance(module, SyncPowerNorm):
 running_phi = module.get_buffer('running_phi')
 if running_phi is not None:
 stats[f"{name}/running_phi"] = running_phi.mean().item()
 ema_gz = module.get_buffer('ema_gz')
 if ema_gz is not None:
 stats[f"{name}/ema_gz"] = ema_gz.mean().item()
 iters = module.get_buffer('iters')
 if iters is not None:
 stats[f"{name}/iters"] = iters.item()
 return stats

helper function for HellaSwag eval
takes tokens, mask, and logits, returns the index of the completion with the lowest loss

def get_most_likely_row(tokens, mask, logits):
 # evaluate the autoregressive loss at all positions

```

```

shift_logits = (logits[..., :-1, :]).contiguous()
shift_tokens = (tokens[..., 1:]).contiguous()
flat_shift_logits = shift_logits.view(-1, shift_logits.size(-1))
flat_shift_tokens = shift_tokens.view(-1)
shift_losses = F.cross_entropy(flat_shift_logits, flat_shift_tokens, reduction='none')
shift_losses = shift_losses.view(tokens.size(0), -1)
now get the average loss just for the completion region (where mask == 1), in each row
shift_mask = (mask[..., 1:]).contiguous() # we must shift mask, so we start at the last prompt token
masked_shift_losses = shift_losses * shift_mask
sum and divide by the number of 1s in the mask
sum_loss = masked_shift_losses.sum(dim=1)
avg_loss = sum_loss / shift_mask.sum(dim=1)
now we have a loss for each of the 4 completions
the one with the lowest loss should be the most likely
pred_norm = avg_loss.argmin().item()
return pred_norm

simple launch:
python train_gpt2.py
DDP launch for e.g. 8 GPUs:
torchrun --standalone --nproc_per_node=8 train_gpt2.py

run the training loop

set up DDP (distributed data parallel).
torchrun command sets the env variables RANK, LOCAL_RANK, and WORLD_SIZE
ddp = int(os.environ.get('RANK', -1)) != -1 # is this a ddp run?
if ddp:
 # use of DDP atm demands CUDA, we set the device appropriately according to rank
 assert torch.cuda.is_available(), "for now i think we need CUDA for DDP"
 init_process_group(backend='nccl')
 ddp_rank = int(os.environ['RANK'])
 ddp_local_rank = int(os.environ['LOCAL_RANK'])
 ddp_world_size = int(os.environ['WORLD_SIZE'])
 device = f'cuda:{ddp_local_rank}'
 torch.cuda.set_device(device)
 master_process = ddp_rank == 0 # this process will do logging, checkpointing etc.
else:
 # vanilla, non-DDP run
 ddp_rank = 0
 ddp_local_rank = 0
 ddp_world_size = 1
 master_process = True
 # attempt to autodetect device
 device = "cpu"
 if torch.cuda.is_available():
 device = "cuda"
 elif hasattr(torch.backends, "mps") and torch.backends.mps.is_available():
 device = "mps"
 print(f"using device: {device}")

GPTesla - 111M param setup in comment. Modification to make lighter training requirement needed
config = {
 "weight_decay": 0.1,
 # "lr_scheduler_type": "cosine",
 "gradient_accumulation_steps": (2**16 * 7) // (56 * 1024 * ddp_world_size), # total_batch_size // (B * T
* ddp_world_size
 "max_eval_steps": 20,
}

```

```

"seq_length": 1024,

New centralised parameters
"project_name": "shng2025/GPT-Valkyrie_PN-124m",
"total_batch_size": 2**16 * 7, # temporarily because 6 GPUs # 2**19, ~0.5M, in number of tokens
"micro_batch_size": 56,
"max_lr": 6e-4,
"min_lr": 6e-5, # 10% of max_lr // not used, as we are using weight_decay instead
"warmup_steps": 715,
"max_steps": 21797, # had to be scaled up after 2001st step, as memory ran out when DDP
"val_every": 500, # EVALUATION
"generate_every": 500, # EVALUATION
"hellaswag_every": 500, # EVALUATION
"save_every": 2000, # SAVE CHECKPOINTING
"log_dir": "./log",
"device": "auto", # "auto", "cpu", "cuda", or "mps"
"use_compile": True,
"grad_clip": 1.0,
"num_return_sequences": 4,
"max_generate_length": 32,
"top_k": 50,

"resume_from_checkpoint": False,
}

args = Namespace(**config)
samples_per_step = torch.cuda.device_count() * args.micro_batch_size

project_name = args.project_name

Logging - DEPRECATED
if master_process:
 pass
 # run_name, wandb_id = setup_logging(project_name.split("/")[1])
 # print(f"Weights and Biases run name: {run_name}")
 # print(f"Weights and Biases run id : {wandb_id}")

added after video, pytorch can be serious about it's device vs. device_type distinction
device_type = "cuda" if device.startswith("cuda") else "cpu"

torch.manual_seed(1337)
if torch.cuda.is_available():
 torch.cuda.manual_seed(1337)

enc = tiktoken.get_encoding("gpt2")

total_batch_size = args.total_batch_size # 2**19, ~0.5M, in number of tokens
B = args.micro_batch_size # micro batch size
T = args.seq_length # sequence length
assert total_batch_size % (B * T * ddp_world_size) == 0, "make sure total_batch_size is divisible by B * T * ddp_world_size"
grad_accum_steps = total_batch_size // (B * T * ddp_world_size)
if master_process:
 print(f"total desired batch size: {total_batch_size}")
 print(f"=> calculated gradient accumulation steps: {grad_accum_steps}")

train_loader = DataLoaderLite(B=B, T=T, process_rank=ddp_rank, num_processes=ddp_world_size, split="train")
val_loader = DataLoaderLite(B=B, T=T, process_rank=ddp_rank, num_processes=ddp_world_size, split="val")

```

```

torch.set_float32_matmul_precision('high')

create model
model = GPT(GPTConfig(vocab_size=50304))
model = GPT.from_pretrained("gpt2") # or init from OpenAI GPT-2
model.to(device)
use_compile = True # torch.compile interferes with HellaSwag eval and Generation. TODO fix
if use_compile:
 model = torch.compile(model)
if ddp:
 model = DDP(model, device_ids=[ddp_local_rank])
raw_model = model.module if ddp else model # always contains the "raw" unwrapped model

max_lr = args.max_lr
min_lr = max_lr * args.weight_decay
warmup_steps = args.warmup_steps
max_steps = args.max_steps # 19,073 steps is ~1 epoch, if data is 10B tokens and batch size 0.5M tokens
def get_lr(it):
 # 1) linear warmup for warmup_iters steps
 if it < warmup_steps:
 return max_lr * (it+1) / warmup_steps
 # 2) if it > lr_decay_iters, return min learning rate
 if it > max_steps:
 return min_lr
 # 3) in between, use cosine decay down to min learning rate
 decay_ratio = (it - warmup_steps) / (max_steps - warmup_steps)
 assert 0 <= decay_ratio <= 1
 coeff = 0.5 * (1.0 + math.cos(math.pi * decay_ratio)) # coeff starts at 1 and goes to 0
 return min_lr + coeff * (max_lr - min_lr)

optimize!
optimizer = raw_model.configure_optimizers(weight_decay=args.weight_decay, learning_rate=args.max_lr,
device_type=device_type)

create the log directory we will write checkpoints to and log to
log_dir = args.log_dir
os.makedirs(log_dir, exist_ok=True)
log_file = os.path.join(log_dir, "log.txt")
with open(log_file, "w") as f: # open for writing to clear the file
 pass

"""
Initialize HuggingFace repository
if master_process:
 new_branch_name = run_name
 create_branch(project_name, repo_type="model", branch=new_branch_name)
 hf_repo = Repository("./", clone_from=project_name, revision=run_name)
"""

Training loop
starting_step = 0

if args.resume_from_checkpoint:
 checkpoint_dir = args.log_dir
 checkpoint_pattern = os.path.join(checkpoint_dir, "checkpoint_*.pt")
 checkpoint_files = glob.glob(checkpoint_pattern)
 latest_checkpoint = max(checkpoint_files, key=os.path.getctime)
 checkpoint_path = latest_checkpoint

 # Use the load_checkpoint function here
 starting_step, val_loss, run_name, wandb_id = load_checkpoint(checkpoint_path, raw_model, optimizer,
train_loader, val_loader)

```

```

starting_step += 1 # to make sure step 80 isn't repeated

logger, run_name = resume_logging(project_name.split("/")[1], wandb_id, args)
print(f"Resuming from checkpoint: {checkpoint_path}")
print(f"Weights and Biases run name: {run_name}")
print(f"Resuming from step: {starting_step}")

Initialize HuggingFace repository <-- UNSURE IF NEEDED
if master_process:
 new_branch_name = run_name
 # create_branch(project_name, repo_type="model", branch=new_branch_name)
 hf_repo = Repository("./", clone_from=project_name, revision=run_name)

Local subprocess for git pulling and checking out the newest branch
if master_process:
 import subprocess
 subprocess.run(["git", "fetch", "origin"])
 subprocess.run(["git", "checkout", new_branch_name])
 subprocess.run(["git", "pull", "origin", new_branch_name])

if master_process:
 print(f"Resuming from checkpoint at step {starting_step}")
else:
 starting_step = 0
 logger, run_name, wandb_id = setup_logging(project_name.split("/")[1], args)
 print(f"Weights and Biases run name: {run_name}")
 print(f"Weights and Biases run id : {wandb_id}")

Initialize HuggingFace repository
if master_process:
 new_branch_name = run_name
 create_branch(project_name, repo_type="model", branch=new_branch_name)
 hf_repo = Repository("./", clone_from=project_name, revision=run_name)

if master_process:
 print(f"Starting new run: {run_name}")

Training Loop
for step in range(starting_step, max_steps):
 t0 = time.time()
 last_step = (step == max_steps - 1)

 # once in a while evaluate our validation loss
 if step % args.val_every == 0 or last_step:
 model.eval()
 val_loader.reset()
 with torch.no_grad():
 val_loss_accum = 0.0
 val_loss_steps = args.max_eval_steps
 for _ in range(val_loss_steps):
 x, y = val_loader.next_batch()
 x, y = x.to(device), y.to(device)
 with torch.autocast(device_type=device_type, dtype=torch.bfloat16):
 logits, loss = model(x, y)
 loss = loss / val_loss_steps
 val_loss_accum += loss.detach()
 if ddp:
 dist.all_reduce(val_loss_accum, op=dist.ReduceOp.AVG)
 log_metrics({"loss/validation": val_loss_accum.item()})

 # once in a while evaluate hellaswag
 if (step % args.hellaswag_every == 0 or last_step) and (not use_compile):

```

```

num_correct_norm = 0
num_total = 0
for i, example in enumerate(iterate_examples("val")):
 # only process examples where i % ddp_world_size == ddp_rank
 if i % ddp_world_size != ddp_rank:
 continue
 # render the example into tokens and labels
 _, tokens, mask, label = render_example(example)
 tokens = tokens.to(device)
 mask = mask.to(device)
 # get the logits
 with torch.no_grad():
 with torch.autocast(device_type=device_type, dtype=torch.bfloat16):
 logits, loss = model(tokens)
 pred_norm = get_most_likely_row(tokens, mask, logits)
 num_total += 1
 num_correct_norm += int(pred_norm == label)
reduce the stats across all processes
if ddp:
 num_total = torch.tensor(num_total, dtype=torch.long, device=device)
 num_correct_norm = torch.tensor(num_correct_norm, dtype=torch.long, device=device)
 dist.all_reduce(num_total, op=dist.ReduceOp.SUM)
 dist.all_reduce(num_correct_norm, op=dist.ReduceOp.SUM)
 num_total = num_total.item()
 num_correct_norm = num_correct_norm.item()
acc_norm = num_correct_norm / num_total
if master_process:
 log_metrics({"hella/swag": acc_norm, "hella/correct_norm": num_correct_norm, "hella/num_total": num_total})
 print(f"HellaSwag accuracy: {num_correct_norm}/{num_total}={acc_norm:.4f}")
 with open(log_file, "a") as f:
 f.write(f"step hella {acc_norm:.4f}\n")

once in a while generate from the model (except step 0, which is noise)
if ((step > 0 and step % args.generate_every == 0) or last_step) and (not use_compile):
 model.eval()
 num_return_sequences = args.num_return_sequences
 max_length = args.max_generate_length
 tokens = enc.encode("Hello, I'm a language model,")
 tokens = torch.tensor(tokens, dtype=torch.long)
 tokens = tokens.unsqueeze(0).repeat(num_return_sequences, 1)
 xgen = tokens.to(device)
 sample_rng = torch.Generator(device=device)
 sample_rng.manual_seed(42 + ddp_rank)
 while xgen.size(1) < max_length:
 # forward the model to get the logits
 with torch.no_grad():
 with torch.autocast(device_type=device_type, dtype=torch.bfloat16):
 logits, loss = model(xgen) # (B, T, vocab_size)
 # take the logits at the last position
 logits = logits[:, -1, :] # (B, vocab_size)
 # get the probabilities
 probs = F.softmax(logits, dim=-1)
 # do top-k sampling of 50 (huggingface pipeline default)
 # topk_probs here becomes (5, 50), topk_indices is (5, 50)
 topk_probs, topk_indices = torch.topk(probs, args.top_k, dim=-1) # top_k == 50 in this case
 # select a token from the top-k probabilities
 # note: multinomial does not demand the input to sum to 1
 ix = torch.multinomial(topk_probs, 1, generator=sample_rng) # (B, 1)
 # gather the corresponding indices
 xcol = torch.gather(topk_indices, -1, ix) # (B, 1)
 # append to the sequence

```

```

 xgen = torch.cat((xgen, xcol), dim=1)
print the generated text
for i in range(num_return_sequences):
 tokens = xgen[i, :max_length].tolist()
 decoded = enc.decode(tokens)
 print(f"rank {ddp_rank} sample {i}: {decoded}")

if (step % args.save_every == 0 or last_step) and step > 0:
 if master_process:
 # Save checkpoint and push to HuggingFace
 checkpoint_path = save_checkpoint(
 raw_model, optimizer, step, val_loss_accum.item(), run_name,
 train_loader.get_state(), val_loader.get_state(), wandb_id
)
 hf_repo.push_to_hub(commit_message=f"Checkpoint at step {step}")
 print(f"Saved checkpoint and pushed to HuggingFace at step {step}")

do one step of the optimization
model.train()
optimizer.zero_grad()
loss_accum = 0.0

for micro_step in range(grad_accum_steps):
 x, y = train_loader.next_batch()
 x, y = x.to(device), y.to(device)
 # added after video, this field is also used by the forward pass.
 if ddp:
 model.require_backward_grad_sync = (micro_step == grad_accum_steps - 1)
 with torch.autocast(device_type=device_type, dtype=torch.bfloat16):
 logits, loss = model(x, y)
 # we have to scale the loss to account for gradient accumulation,
 # because the gradients just add on each successive backward().
 # addition of gradients corresponds to a SUM in the objective, but
 # instead of a SUM we want MEAN. Scale the loss here so it comes out right
 loss = loss / grad_accum_steps
 loss_accum += loss.detach()

 # Add NaN check here
 if not torch.isnan(loss):
 loss.backward()
 else:
 print(f"NaN loss detected at step {step}, micro_step {micro_step}. Skipping backward.")
 if master_process:
 wandb.alert(
 title="NaN Loss Detected",
 text=f"NaN loss detected at step {step}, micro_step {micro_step}. Skipping backward.",
 level=wandb.AlertLevel.WARN
)
 break

 if ddp:
 dist.all_reduce(loss_accum, op=dist.ReduceOp.AVG)

Add another NaN check here before optimizer step
if not torch.isnan(loss_accum):
 norm = torch.nn.utils.clip_grad_norm_(model.parameters(), args.grad_clip) # grad_clip == 1.0 by
default
 # determine and set the learning rate for this iteration
 lr = get_lr(step)
 for param_group in optimizer.param_groups:
 param_group['lr'] = lr
 optimizer.step()

```

```

else:
 print("successfully broke")
 print(f"NaN accumulated loss detected at step {step}. Skipping optimizer step.")
 wandb.alert(
 title="NaN Accumulated Loss Detected",
 text=f"NaN accumulated loss detected at step {step}. Skipping optimizer step.",
 level=wandb.AlertLevel.ERROR
)
 break

if device_type == "cuda":
 torch.cuda.synchronize() # wait for the GPU to finish work
t1 = time.time()
dt = t1 - t0 # time difference in seconds
tokens_processed = train_loader.B * train_loader.T * grad_accum_steps * ddp_world_size
tokens_per_sec = tokens_processed / dt
if master_process:
 pn_stats = log_pownorm_stats(raw_model)
 log_metrics({
 "lr": lr, # get_lr()
 "samples": step * samples_per_step,
 "steps": step,
 "loss/train": loss_accum.item(),
 # file specific addition
 "global gradient norm": norm,
 "dt": dt,
 "tok per sec": tokens_per_sec,
 **pn_stats # Unpack the PowerNorm stats into the metrics
 })
 print(f"step {step:5d} | loss: {loss_accum.item():.6f} | lr {lr:.4e} | norm: {norm:.4f} | dt: {dt*1000:.2f}ms | tok/sec: {tokens_per_sec:.2f}")
 with open(log_file, "a") as f:
 f.write(f"{step} train {loss_accum.item():.6f}\n")

if ddp:
 destroy_process_group()

Final save and push
if master_process:
 final_checkpoint_path = checkpoint_path = save_checkpoint(raw_model, optimizer, step,
 val_loss_accum.item(), run_name, train_loader.get_state(), val_loader.get_state(), wandb_id)
 hf_repo.push_to_hub(commit_message="Final model")
 print("Training completed. Final model pushed to HuggingFace.")

```

## D.4 Tokenization of EduFineweb-10B dataset using GPT-2 tokenizer

```

import os
import multiprocessing as mp
import numpy as np
import tiktoken
from datasets import load_dataset # pip install datasets
from tqdm import tqdm # pip install tqdm

local_dir = "edu_fineweb10B"
remote_name = "sample-10BT"
shard_size = int(1e8) # 100M tokens per shard, total of 100 shards

create the cache the local directory if it doesn't exist yet

```

```

DATA_CACHE_DIR = os.path.join(os.path.dirname(__file__), local_dir)
os.makedirs(DATA_CACHE_DIR, exist_ok=True)

download the dataset
fw = load_dataset("HuggingFaceFW/fineweb-edu", name=remote_name, split="train")

init the tokenizer
enc = tiktoken.get_encoding("gpt2")
eot = enc._special_tokens['<|endoftext|>'] # end of text token
def tokenize(doc):
 # tokenizes a single document and returns a numpy array of uint16 tokens
 tokens = [eot] # the special <|endoftext|> token delimits all documents
 tokens.extend(enc.encode_ordinary(doc["text"]))
 tokens_np = np.array(tokens)
 assert (0 <= tokens_np).all() and (tokens_np < 2**16).all(), "token dictionary too large for uint16"
 tokens_np_uint16 = tokens_np.astype(np.uint16)
 return tokens_np_uint16

def write_datafile(filename, tokens_np):
 np.save(filename, tokens_np)

tokenize all documents and write output shards, each of shard_size tokens (last shard has remainder)
nprocs = max(1, os.cpu_count()//2)
with mp.Pool(nprocs) as pool:
 shard_index = 0
 # preallocate buffer to hold current shard
 all_tokens_np = np.empty((shard_size,), dtype=np.uint16)
 token_count = 0
 progress_bar = None
 for tokens in pool imap(tokenize, fw, chunksize=16):

 # is there enough space in the current shard for the new tokens?
 if token_count + len(tokens) < shard_size:
 # simply append tokens to current shard
 all_tokens_np[token_count:token_count+len(tokens)] = tokens
 token_count += len(tokens)
 # update progress bar
 if progress_bar is None:
 progress_bar = tqdm(total=shard_size, unit="tokens", desc=f"Shard {shard_index}")
 progress_bar.update(len(tokens))
 else:
 # write the current shard and start a new one
 split = "val" if shard_index == 0 else "train"
 filename = os.path.join(DATA_CACHE_DIR, f"edufineweb_{split}_{shard_index:06d}")
 # split the document into whatever fits in this shard; the remainder goes to next one
 remainder = shard_size - token_count
 progress_bar.update(remainder)
 all_tokens_np[token_count:token_count+remainder] = tokens[:remainder]
 write_datafile(filename, all_tokens_np)
 shard_index += 1
 progress_bar = None
 # populate the next shard with the leftovers of the current doc
 all_tokens_np[0:len(tokens)-remainder] = tokens[remainder:]
 token_count = len(tokens)-remainder

 # write any remaining tokens as the last shard
if token_count != 0:
 split = "val" if shard_index == 0 else "train"
 filename = os.path.join(DATA_CACHE_DIR, f"edufineweb_{split}_{shard_index:06d}")
 write_datafile(filename, all_tokens_np[:token_count])

```

## Appendix E - Ablation Code

Below contains the code that we used to substitute the original `LayerNorm` class with an identity function.

### E.1 - LayerNorm Ablation Code

```
import torch
from transformers import GPT2LMHeadModel, GPT2Config
import os
import torch.nn as nn

class IdentityLayerNorm(nn.LayerNorm):
 def __init__(self, normalized_shape, eps=1e-5, elementwise_affine=True):
 super().__init__(normalized_shape, eps, elementwise_affine)

 # Set weights to 1 and bias to 0
 if self.elementwise_affine:
 nn.init.ones_(self.weight)
 nn.init.zeros_(self.bias)

 def forward(self, x):
 return x # Identity function

def replace_layer_norm(module, name):
 if isinstance(getattr(module, name), nn.LayerNorm):
 setattr(module, name, IdentityLayerNorm(768))

def create_noNorm_model(base_model):
 model = GPT2LMHeadModel(base_model.config)
 model.load_state_dict(base_model.state_dict())

 for block in model.transformer.h:
 replace_layer_norm(block, 'ln_1')
 replace_layer_norm(block, 'ln_2')

 # Ensure the changes are reflected in the state dict
 model.transformer.h = nn.ModuleList(model.transformer.h)

 # Store the variant type in the config
 model.config.variant_type = "noNorm"
 model.config.name_or_path = "shng2025/GPT-Valkyrie_LN-124m_noNorm_"
 return model

def create_AttnOnly_model(base_model):
 model = GPT2LMHeadModel(base_model.config)
 model.load_state_dict(base_model.state_dict())

 for block in model.transformer.h:
 # replace_layer_norm(block, 'ln_1')
 replace_layer_norm(block, 'ln_2')

 # Ensure the changes are reflected in the state dict
 model.transformer.h = nn.ModuleList(model.transformer.h)

 # Store the variant type in the config
 model.config.variant_type = "AttnOnly"
 model.config.name_or_path = "shng2025/GPT-Valkyrie_LN-124m_AttnOnly_"
 return model

def create_FFNonly_model(base_model):
 model = GPT2LMHeadModel(base_model.config)
```

```

model.load_state_dict(base_model.state_dict())

for block in model.transformer.h:
 replace_layer_norm(block, 'ln_1')
 # replace_layer_norm(block, 'ln_2')

Ensure the changes are reflected in the state dict
model.transformer.h = nn.ModuleList(model.transformer.h)

Store the variant type in the config
model.config.variant_type = "FFNonly"
model.config.name_or_path = "shng2025/GPT-Valkyrie_LN-124m__FFNonly__"
return model

```

```

def examine_layers(model, model_name):
 # Create a sample input
 sample_input = torch.randn(1, 1, model.config.n_embd)

 print(f"Examining model: {model_name}")
 print(f"Variant type: {getattr(model.config, 'variant_type', 'Not specified')}")
 print("\nNormalization layer details:")
 for i, block in enumerate(model.transformer.h):
 print(f" Block {i}:")

 # ln_1 (before attention)
 ln1_output = block.ln_1(sample_input)
 print(f" ln_1 (before attention):")
 print(f" Type: {type(block.ln_1).__name__}")
 print(f" Input mean: {sample_input.mean().item():.4f}, std: {sample_input.std().item():.4f}")
 print(f" Output mean: {ln1_output.mean().item():.4f}, std: {ln1_output.std().item():.4f}")
 print(f" Is identity: {torch.allclose(sample_input, ln1_output, atol=1e-6)}")

 # ln_2 (before FFN)
 ln2_output = block.ln_2(sample_input)
 print(f" ln_2 (before FFN):")
 print(f" Type: {type(block.ln_2).__name__}")
 print(f" Input mean: {sample_input.mean().item():.4f}, std: {sample_input.std().item():.4f}")
 print(f" Output mean: {ln2_output.mean().item():.4f}, std: {ln2_output.std().item():.4f}")
 print(f" Is identity: {torch.allclose(sample_input, ln2_output, atol=1e-6)}")

 # Final layer norm
 ln_f_output = model.transformer.ln_f(sample_input)
 print(f" Final layer norm:")
 print(f" Type: {type(model.transformer.ln_f).__name__}")
 print(f" Input mean: {sample_input.mean().item():.4f}, std: {sample_input.std().item():.4f}")
 print(f" Output mean: {ln_f_output.mean().item():.4f}, std: {ln_f_output.std().item():.4f}")
 print(f" Is identity: {torch.allclose(sample_input, ln_f_output, atol=1e-6)}")

Load the base model
base_model_path = "shng2025/GPT-Valkyrie_LN-124m__baseModel__"
base_model = GPT2LMHeadModel.from_pretrained(base_model_path)

Create the noNorm variant
noNorm_model = create_noNorm_model(base_model)
AttnOnly_model = create_AttnOnly_model(base_model)
FFNonly_model = create_FFNonly_model(base_model)

Save the noNorm model
save_path = "GPT-Valkyrie_LN-124m__noNorm__"
os.makedirs(save_path, exist_ok=True)
noNorm_model.save_pretrained(save_path)

```

```
print(f"Saved noNorm model to {save_path}")

Save the noNorm model
save_path = "GPT-Valkyrie_LN-124m_AttnOnly_"
os.makedirs(save_path, exist_ok=True)
AttnOnly_model.save_pretrained(save_path)
print(f"Saved noNorm model to {save_path}")

Save the noNorm model
save_path = "GPT-Valkyrie_LN-124m_FFNonly_"
os.makedirs(save_path, exist_ok=True)
FFNonly_model.save_pretrained(save_path)
print(f"Saved noNorm model to {save_path}")

Examine both models
print("\nBase Model:")
examine_layers(base_model, "Base Model")

print("\n" + "="*50 + "\n")

print("noNorm Model:")
examine_layers(noNorm_model, "noNorm Model")

print("\n" + "="*50 + "\n")

print("AttnOnly Model:")
examine_layers(AttnOnly_model, "AttnOnly Model")

print("\n" + "="*50 + "\n")

print("FFNonly Model:")
examine_layers(FFNonly_model, "FFNonly Model")
```

## E.2 - RMSN Ablation Code

```
import torch
from transformers import GPT2LMHeadModel, GPT2Config
import os
import torch.nn as nn

class IdentityLayerNorm(nn.LayerNorm):
 def __init__(self, normalized_shape, eps=1e-5, elementwise_affine=True):
 super().__init__(normalized_shape, eps, elementwise_affine)

 # Set weights to 1 and bias to 0
 if self.elementwise_affine:
 nn.init.ones_(self.weight)
 nn.init.zeros_(self.bias)

 def forward(self, x):
 return x # Identity function

def replace_layer_norm(module, name):
 if isinstance(getattr(module, name), nn.LayerNorm):
 setattr(module, name, IdentityLayerNorm(768))

def create_noNorm_model(base_model):
 model = GPT2LMHeadModel(base_model.config)
 model.load_state_dict(base_model.state_dict())

 for block in model.transformer.h:
 replace_layer_norm(block, 'ln_1')
 replace_layer_norm(block, 'ln_2')

 # Ensure the changes are reflected in the state dict
 model.transformer.h = nn.ModuleList(model.transformer.h)

 # Store the variant type in the config
 model.config.variant_type = "noNorm"
 model.config.name_or_path = "shng2025/GPT-Valkyrie_RMSN-124m_noNorm_"
 return model

def create_AttnOnly_model(base_model):
 model = GPT2LMHeadModel(base_model.config)
 model.load_state_dict(base_model.state_dict())

 for block in model.transformer.h:
 # replace_layer_norm(block, 'ln_1')
 replace_layer_norm(block, 'ln_2')

 # Ensure the changes are reflected in the state dict
 model.transformer.h = nn.ModuleList(model.transformer.h)

 # Store the variant type in the config
 model.config.variant_type = "AttnOnly"
 model.config.name_or_path = "shng2025/GPT-Valkyrie_RMSN-124m_AttnOnly_"
 return model

def create_FFNonly_model(base_model):
 model = GPT2LMHeadModel(base_model.config)
 model.load_state_dict(base_model.state_dict())

 for block in model.transformer.h:
 replace_layer_norm(block, 'ln_1')
```

```

replace_layer_norm(block, 'ln_2')

Ensure the changes are reflected in the state dict
model.transformer.h = nn.ModuleList(model.transformer.h)

Store the variant type in the config
model.config.variant_type = "FFNonly"
model.config.name_or_path = "shng2025/GPT-Valkyrie_RMSN-124m__FFNonly__"
return model

def examine_layers(model, model_name):
 # Create a sample input
 sample_input = torch.randn(1, 1, model.config.n_embd)

 print(f"Examining model: {model_name}")
 print(f"Variant type: {getattr(model.config, 'variant_type', 'Not specified')}")
 print("\nNormalization layer details:")
 for i, block in enumerate(model.transformer.h):
 print(f" Block {i}:")

 # ln_1 (before attention)
 ln1_output = block.ln_1(sample_input)
 print(f" ln_1 (before attention):")
 print(f" Type: {type(block.ln_1).__name__}")
 print(f" Input mean: {sample_input.mean().item():.4f}, std: {sample_input.std().item():.4f}")
 print(f" Output mean: {ln1_output.mean().item():.4f}, std: {ln1_output.std().item():.4f}")
 print(f" Is identity: {torch.allclose(sample_input, ln1_output, atol=1e-6)}")

 # ln_2 (before FFN)
 ln2_output = block.ln_2(sample_input)
 print(f" ln_2 (before FFN):")
 print(f" Type: {type(block.ln_2).__name__}")
 print(f" Input mean: {sample_input.mean().item():.4f}, std: {sample_input.std().item():.4f}")
 print(f" Output mean: {ln2_output.mean().item():.4f}, std: {ln2_output.std().item():.4f}")
 print(f" Is identity: {torch.allclose(sample_input, ln2_output, atol=1e-6)}")

 # Final layer norm
 ln_f_output = model.transformer.ln_f(sample_input)
 print(f" Final layer norm:")
 print(f" Type: {type(model.transformer.ln_f).__name__}")
 print(f" Input mean: {sample_input.mean().item():.4f}, std: {sample_input.std().item():.4f}")
 print(f" Output mean: {ln_f_output.mean().item():.4f}, std: {ln_f_output.std().item():.4f}")
 print(f" Is identity: {torch.allclose(sample_input, ln_f_output, atol=1e-6)}")

Load the base model
base_model_path = "shng2025/GPT-Valkyrie_RMSN-124m__baseModel__"
base_model = GPT2LMHeadModel.from_pretrained(base_model_path)

Create the noNorm variant
noNorm_model = create_noNorm_model(base_model)
AttnOnly_model = create_AttnOnly_model(base_model)
FFNonly_model = create_FFNonly_model(base_model)

Save the noNorm model
save_path = "GPT-Valkyrie_RMSN-124m__noNorm__"
os.makedirs(save_path, exist_ok=True)
noNorm_model.save_pretrained(save_path)
print(f"Saved noNorm model to {save_path}")

Save the noNorm model
save_path = "GPT-Valkyrie_RMSN-124m__AttnOnly__"

```

```
os.makedirs(save_path, exist_ok=True)
AttnOnly_model.save_pretrained(save_path)
print(f"Saved noNorm model to {save_path}")

Save the noNorm model
save_path = "GPT-Valkyrie_RMSN-124m__FFNonly__"
os.makedirs(save_path, exist_ok=True)
FFNonly_model.save_pretrained(save_path)
print(f"Saved noNorm model to {save_path}")

Examine both models
print("\nBase Model:")
examine_layers(base_model, "Base Model")

print("\n" + "="*50 + "\n")

print("noNorm Model:")
examine_layers(noNorm_model, "noNorm Model")

print("\n" + "="*50 + "\n")

print("AttnOnly Model:")
examine_layers(AttnOnly_model, "AttnOnly Model")

print("\n" + "="*50 + "\n")

print("FFNonly Model:")
examine_layers(FFNonly_model, "FFNonly Model")
```

## Appendix F

### F.1 Fine Tuning Summarisation (BillSum)

```
import torch
from transformers import GPT2TokenizerFast, GPT2LMHeadModel, DataCollatorWithPadding
from transformers import Trainer, TrainingArguments
from datasets import load_dataset, concatenate_datasets, DatasetDict
import numpy as np
from rouge_score import rouge_scorer
import wandb
from transformers.integrations import WandbCallback

Load dataset
def load_billsum():
 """
 Load and combine the 'train' and 'test' splits of the Billsum dataset into a single 'train' set,
 and use 'ca_test' as the 'test' set.

 Returns:
 DatasetDict: A dictionary containing 'train' and 'test' datasets.
 """
 # Define the split names
 training_splits = ['train', 'test']
 test_split = 'ca_test'

 datasets_to_combine = []

 # Load and combine training splits
 for split in training_splits:
 try:
 ds = load_dataset("billsum", split=split)
 print(f"Loaded split: {split} with {len(ds)} examples.")
 datasets_to_combine.append(ds)
 except Exception as e:
 print(f"Could not load split '{split}': {e}")

 if not datasets_to_combine:
 raise ValueError("No training datasets were loaded. Please check the split names.")

 # Concatenate 'train' and 'test' splits into a single 'train' set
 combined_train = concatenate_datasets(datasets_to_combine)
 print(f"Combined train dataset size: {len(combined_train)} examples.")

 # Load the 'ca_test' split as the test set
 try:
 test_ds = load_dataset("billsum", split=test_split)
 print(f"Loaded test split: {test_split} with {len(test_ds)} examples.")
 except Exception as e:
 raise ValueError(f"Could not load test split '{test_split}': {e}")

 # Create a DatasetDict with 'train' and 'test' splits
 dataset_dict = DatasetDict({
 "train": combined_train,
 "test": test_ds
 })

 print(f"DatasetDict keys: {dataset_dict.keys()}")
 return dataset_dict
```

```

Initialize tokenizer
tokenizer = GPT2TokenizerFast.from_pretrained("gpt2")
tokenizer.pad_token = tokenizer.eos_token

Preprocess function
def preprocess_function(examples, tokenizer):
 inputs = [f"Summarize the following text in a concise manner:\n\n{doc}\n\nSummary:" for doc in examples["text"]]
 summaries = [summary for summary in examples["summary"]]

 concatenated = [inp + " " + summ for inp, summ in zip(inputs, summaries)]

 tokenized = tokenizer(
 concatenated,
 max_length=1024 + 150, # Adjust based on desired input and summary lengths
 truncation=True,
 padding="max_length"
)

 labels = []
 for inp, summ in zip(inputs, summaries):
 inp_ids = tokenizer.encode(inp, add_special_tokens=False)
 summ_ids = tokenizer.encode(summ, add_special_tokens=False)

 full_ids = inp_ids + summ_ids

 label = [-100] * len(inp_ids) + summ_ids
 label = label + [-100] * (max_length := 1024 + 150 - len(label))
 labels.append(label[:max_length])

 tokenized["labels"] = labels
 return tokenized

Load and preprocess the dataset
dataset = Load_billsum()
tokenized_datasets = dataset.map(
 lambda examples: preprocess_function(examples, tokenizer),
 batched=True,
 remove_columns=['text', 'summary', 'title'] # Remove original text columns to save memory
)
Limit the test set to 100 examples
tokenized_datasets["test"] = tokenized_datasets["test"].shuffle(seed=42).select(range(100))
print(f"Limited test dataset size: {len(tokenized_datasets['test'])} examples.")

Function to freeze layers based on variant type
def freeze_layers(model, variant_type):
 if variant_type == "noNorm":
 for name, param in model.named_parameters():
 if "ln" in name:
 param.requires_grad = False
 elif variant_type == "AttnOnly":
 for name, param in model.named_parameters():
 if "ln_2" in name: # Freeze FFN Layer norm
 param.requires_grad = False
 elif variant_type == "FFNonly":
 for name, param in model.named_parameters():
 if "ln_1" in name: # Freeze attention Layer norm
 param.requires_grad = False

```

```

For baseModel, we don't freeze any layers

import evaluate # Import the evaluate library

Initialize ROUGE and BLEU metrics
rouge = evaluate.load("rouge")
bleu = evaluate.load("bleu")

import sacrebleu

def compute_metrics(eval_pred):
 """
 Compute ROUGE and BLEU metrics for summarization using SacreBLEU with smoothing.

 Args:
 eval_pred (EvalPrediction): Contains predictions and label_ids.

 Returns:
 dict: Average ROUGE and BLEU scores.
 """
 predictions, labels = eval_pred

 # Convert Logits to token IDs by taking the argmax over the vocabulary dimension
 pred_ids = np.argmax(predictions, axis=-1)

 # Decode the predicted token IDs to text
 decoded_preds = tokenizer.batch_decode(pred_ids, skip_special_tokens=True)

 # Replace -100 in the labels with the pad token ID and decode
 labels = np.where(labels != -100, labels, tokenizer.pad_token_id)
 decoded_labels = tokenizer.batch_decode(labels, skip_special_tokens=True)

 # Ensure that the predictions and references are lists of strings
 # and remove any leading/trailing whitespace
 decoded_preds = [pred.strip() for pred in decoded_preds]
 decoded_labels = [label.strip() for label in decoded_labels]

 # Compute ROUGE scores using the evaluate library
 rouge_result = rouge.compute(
 predictions=decoded_preds,
 references=decoded_labels,
 use_stemmer=True
)

 # Compute BLEU scores using SacreBLEU with smoothing
 bleu_scores = sacrebleu.corpus_bleu(
 decoded_preds,
 [decoded_labels],
 smooth_method='exp', # Exponential smoothing
 smooth_value=0.1,
 force=True, # Force compute even if length mismatch
 lowercase=True, # Normalize case
 tokenize='13a' # Tokenizer type (SacreBLEU default)
)

 bleu_score = bleu_scores.score # SacreBLEU returns a score attribute

 # Aggregate the results
 result = {
 "rouge1": rouge_result["rouge1"],

```

```

 "rouge2": rouge_result["rouge2"],
 "rougeL": rouge_result["rougeL"],
 "bleu": bleu_score
}

Optional: Calculate average prediction length
prediction_lens = [len(pred.split()) for pred in decoded_preds]
result["gen_len"] = np.mean(prediction_lens)

Round the results to four decimal places for readability
result = {k: round(v, 4) for k, v in result.items()}

return result

Fine-tuning function
def fine_tune_model(model, tokenizer, dataset, output_dir, variant, norm_type):
 # Initialize wandb run
 wandb.init(project=f"GPT-Valkyrie_{norm_type}-124m_{variant}_Billsum", reinit=True)
 run_name = wandb.run.name

 training_args = TrainingArguments(
 output_dir=output_dir,
 num_train_epochs=3,
 per_device_train_batch_size=4,
 per_device_eval_batch_size=1,
 warmup_steps=500,
 weight_decay=0.01,
 Logging_dir=".//Logs",
 Logging_steps=50,
 evaluation_strategy="steps",
 eval_steps=200,
 save_steps=200,
 Load_best_model_at_end=True,
 metric_for_best_model="rougeL",
 report_to="wandb",
 run_name=run_name,
 save_total_limit=3,
 fp16=True,
)

 from transformers import DataCollatorForLanguageModeling
 data_collator = DataCollatorForLanguageModeling(
 tokenizer=tokenizer,
 mLM=False # Causal Language modeling
)

 trainer = Trainer(
 model=model,
 args=training_args,
 train_dataset=dataset["train"],
 eval_dataset=dataset["test"], # even tho its meant to be called "validation" instead
 tokenizer=tokenizer,
 data_collator=data_collator, # Use the updated data collator
 compute_metrics=compute_metrics,
 callbacks=[WandbCallback()],
)

 trainer.train()
 wandb.finish()
 return trainer.model, run_name

```

```

Main training loop
variants = ["baseModel", "noNorm", "AttnOnly", "FFNonly"]
norm_types = ["LN", "RMSN"]

for norm_type in norm_types:
 for variant in variants:
 print(f"Processing {norm_type} {variant} model...")

 # Use the correct base model for each variant
 model_path = f"shng2025/GPT-Valkyrie_{norm_type}-124m_{variant}_"
 model = GPT2LMHeadModel.from_pretrained(model_path)

 model.config.pad_token_id = tokenizer.pad_token_id
 # Print to verify
 print(f"Tokenizer pad token: {tokenizer.pad_token}")
 print(f"Tokenizer pad token ID: {tokenizer.pad_token_id}")
 print(f"Model pad token ID: {model.config.pad_token_id}")

 freeze_Layers(model, variant)

 output_dir = f"./results/{norm_type}/{variant}"
 fine_tuned_model, run_name = fine_tune_model(model, tokenizer, tokenized_datasets, output_dir,
 variant, norm_type)

 # Save the model locally
 local_save_dir = f"./Local_models/GPT-Valkyrie_{norm_type}-124m_{variant}_Billsum"
 fine_tuned_model.save_pretrained(local_save_dir)
 tokenizer.save_pretrained(local_save_dir)
 print(f"Model saved locally to {local_save_dir}")

 # Push the model to your HuggingFace Hub repository
 new_repo_name = f"shng2025/GPT-Valkyrie_{norm_type}-124m_{variant}_Billsum"
 fine_tuned_model.push_to_hub(new_repo_name, branch=run_name)
 tokenizer.push_to_hub(new_repo_name, branch=run_name)
 print(f"Model pushed to HuggingFace Hub: {new_repo_name}, branch: {run_name}")

print("Training completed for all variants and normalization types.")

```

## F.2 Fine Tuning Question Answering (SQuAD)

```
import torch
from transformers import GPT2ForQuestionAnswering, GPT2TokenizerFast, Trainer, TrainingArguments
from datasets import load_dataset, load_metric
from torch.nn.utils.rnn import pad_sequence
import os
import wandb
from transformers.integrations import WandbCallback
import numpy as np

from datasets import load_metric # used in compute_metrics
from transformers.trainer_utils import EvalPrediction
from typing import Dict, List

def freeze_layers(model, variant_type):
 if variant_type == "noNorm":
 for name, param in model.named_parameters():
 if "ln" in name:
 param.requires_grad = False
 elif variant_type == "AttnOnly":
 for name, param in model.named_parameters():
 if "ln_2" in name: # Freeze FFN layer norm
 param.requires_grad = False
 elif variant_type == "FFOnly":
 for name, param in model.named_parameters():
 if "ln_1" in name: # Freeze attention layer norm
 param.requires_grad = False
 # For baseModel, we don't freeze any layers

def prepare_squad_dataset(tokenizer):
 dataset = load_dataset("squad_v2")

 def preprocess_function(examples):
 questions = [q.strip() for q in examples["question"]]
 contexts = [c.strip() for c in examples["context"]]

 # Tokenize questions and contexts together
 tokenized_examples = tokenizer(
 [tokenizer.cls_token + q for q in questions], # new addition
 contexts,
 max_length=384,
 truncation="only_second",
 stride=128,
 return_overflowing_tokens=True,
 padding="max_length",
 return_offsets_mapping=True,
)

 sample_mapping = tokenized_examples.pop("overflow_to_sample_mapping")
 offset_mapping = tokenized_examples["offset_mapping"]

 tokenized_examples["start_positions"] = []
 tokenized_examples["end_positions"] = []

 for i, offsets in enumerate(offset_mapping):
 input_ids = tokenized_examples["input_ids"][i]
 cls_index = 0

 sequence_ids = tokenized_examples.sequence_ids(i)

```

```

sample_index = sample_mapping[i]
answers = examples["answers"][sample_index]

if len(answers["answer_start"]) == 0:
 tokenized_examples["start_positions"].append(cls_index)
 tokenized_examples["end_positions"].append(cls_index)
else:
 start_char = answers["answer_start"][0]
 end_char = start_char + len(answers["text"][0])

 token_start_index = 0
 while sequence_ids[token_start_index] != 1:
 token_start_index += 1

 token_end_index = len(input_ids) - 1
 while sequence_ids[token_end_index] != 1:
 token_end_index -= 1

 if not (offsets[token_start_index][0] <= start_char and offsets[token_end_index][1] >= end_char):
 tokenized_examples["start_positions"].append(cls_index)
 tokenized_examples["end_positions"].append(cls_index)
 else:
 while token_start_index < len(offsets) and offsets[token_start_index][0] <= start_char:
 token_start_index += 1
 tokenized_examples["start_positions"].append(token_start_index - 1)
 while offsets[token_end_index][1] >= end_char:
 token_end_index -= 1
 tokenized_examples["end_positions"].append(token_end_index + 1)
return tokenized_examples

tokenized_datasets = dataset.map(
 preprocess_function,
 batched=True,
 remove_columns=dataset["train"].column_names,
)
return tokenized_datasets

def fine_tune_model(model, tokenizer, dataset, output_dir, variant, ln_type, num_train_epochs=2):
 # Initialize wandb run
 wandb.init(project=f"GPT-Valkyrie_{ln_type}-124m_{variant}_SQuAD", reinit=True)
 run_name = wandb.run.name

 training_args = TrainingArguments(
 output_dir=output_dir,
 num_train_epochs=num_train_epochs,
 per_device_train_batch_size=112,
 per_device_eval_batch_size=112,
 warmup_ratio=0.1,
 weight_decay=0.01,
 logging_dir=".logs",
 logging_strategy="steps",
 logging_steps=1,
 evaluation_strategy="epoch",
 save_strategy="epoch",
 load_best_model_at_end=True,
 report_to="wandb",
 run_name=run_name,
)

```

```

Initialize the qa_outputs layer
model.qa_outputs = torch.nn.Linear(model.config.hidden_size, 2)
model.qa_outputs.weight.data.normal_(mean=0.0, std=model.config.initializer_range)
model.qa_outputs.bias.data.zero_()

trainer = Trainer(
 model=model,
 args=training_args,
 train_dataset=dataset["train"],
 eval_dataset=dataset["validation"],
 tokenizer=tokenizer,
 callbacks=[WandbCallback()],
)
trainer.train()
wandb.finish()
return trainer.model, run_name

MAIN LOOP
wandb.login()

variants = ["noNorm", "AttnOnly", "FFNonly", "baseModel"]

from transformers import GPT2TokenizerFast

tokenizer = GPT2TokenizerFast.from_pretrained("gpt2")
tokenizer.pad_token = tokenizer.eos_token
tokenizer.cls_token = "[CLS]"
tokenizer.add_special_tokens({'cls_token': '[CLS]'})

dataset = prepare_squad_dataset(tokenizer)

for variant in variants:
 print(f"Processing {variant} model...")

 # Use the correct base model for each variant
 model_path = f"shng2025/GPT-Valkyrie_LN-124m_{variant}_"
 model = GPT2ForQuestionAnswering.from_pretrained(model_path)

 freeze_layers(model, variant)

 output_dir = f"./results/{variant}"
 fine_tuned_model, run_name = fine_tune_model(model, tokenizer, dataset, output_dir, variant, "LN")

 # Save the model locally
 local_save_dir = f"./local_models/GPT-Valkyrie_LN-124m_{variant}_SQuAD"
 fine_tuned_model.save_pretrained(local_save_dir)
 tokenizer.save_pretrained(local_save_dir)
 print(f"Model saved locally to {local_save_dir}")

 # Push the model to your HuggingFace Hub repository
 new_repo_name = f"shng2025/GPT-Valkyrie_LN-124m_{variant}_SQuAD"
 fine_tuned_model.push_to_hub(new_repo_name, branch=run_name)
 tokenizer.push_to_hub(new_repo_name, branch=run_name)
 print(f"Model pushed to HuggingFace Hub: {new_repo_name}, branch: {run_name}")

for variant in variants:

```

```
print(f"Processing {variant} model...")

Use the correct base model for each variant
model_path = f"shng2025/GPT-Valkyrie_RMSN-124m_{variant}_"
model = GPT2ForQuestionAnswering.from_pretrained(model_path)

freeze_layers(model, variant)

output_dir = f"./results/{variant}"
fine_tuned_model, run_name = fine_tune_model(model, tokenizer, dataset, output_dir, variant, "RMSN")

Save the model locally
local_save_dir = f"./local_models/GPT-Valkyrie_RMSN-124m_{variant}_SQuAD"
fine_tuned_model.save_pretrained(local_save_dir)
tokenizer.save_pretrained(local_save_dir)
print(f"Model saved locally to {local_save_dir}")

Push the model to your HuggingFace Hub repository
new_repo_name = f"shng2025/GPT-Valkyrie_RMSN-124m_{variant}_SQuAD"
fine_tuned_model.push_to_hub(new_repo_name, branch=run_name)
tokenizer.push_to_hub(new_repo_name, branch=run_name)
print(f"Model pushed to HuggingFace Hub: {new_repo_name}, branch: {run_name}")
```

## Appendix G - Collecting CSV Files

"This section presents the code used to collect the CSV files. These CSV files contain the input provided to the model, the reference answers, and the model's responses for each downstream task (Text Generation, Summarization, and Question Answering). Additionally, they include the relevant evaluation metrics for each task, whether traditional metrics or GPT-4o evaluation metrics."

### G.1 TextGen (GPT-4o Metrics)

```
import os
import pandas as pd
from transformers import pipeline
import torch
from tqdm import tqdm

def main():
 # List of models to evaluate
 models = [
 "shng2025/GPT-Valkyrie_RMSN-124m_noNorm__",
 "shng2025/GPT-Valkyrie_RMSN-124m_FFNonly__",
 "shng2025/GPT-Valkyrie_RMSN-124m_AttnOnly__",
 "shng2025/GPT-Valkyrie_RMSN-124m_baseModel__",
 "shng2025/GPT-Valkyrie_LN-124m_noNorm__",
 "shng2025/GPT-Valkyrie_LN-124m_FFNonly__",
 "shng2025/GPT-Valkyrie_LN-124m_AttnOnly__",
 "shng2025/GPT-Valkyrie_LN-124m_baseModel__"
]

 # Ensure output directory exists
 output_dir = "generated_texts"
 os.makedirs(output_dir, exist_ok=True)

 # List of prompts
 prompts = [
 "Once upon a time in a distant galaxy,", "The secret to happiness is", "In the midst of the bustling city,", "As the sun set over the horizon,", "The mysterious package arrived on my doorstep,", "In a world where robots and humans coexist,", "The ancient prophecy foretold that", "Amidst the chaos of the storm,", "The scientist peered into the microscope and saw", "Every morning, she would start her day with", "The last thing I expected to find in the attic was", "Under the cover of darkness,", "With a heavy heart, he decided to", "On the eve of the grand festival,", "The sound of laughter filled the air as", "If time travel were possible,", "Deep beneath the ocean waves,", "In the quiet village nestled among the hills,", "The door creaked open to reveal", "Legends speak of a sword that", "In the year 2525, humanity has", "The aroma of freshly baked bread", "She looked into his eyes and knew that", "On the first day of school,", "The journey to the top of the mountain was",
]
]
```

```

for model_name in models:
 print(f"Processing model: {model_name}")
 # Load the text generation pipeline
 generator = pipeline(
 'text-generation',
 model=model_name,
 tokenizer="gpt2",
 device=0 if torch.cuda.is_available() else -1
)
 # Set pad_token to eos_token
 generator.tokenizer.pad_token = generator.tokenizer.eos_token

generated_texts = []

for prompt in tqdm(prompts, desc=f"Generating texts for {model_name}"):
 outputs = generator(
 prompt,
 max_length=1024,
 num_return_sequences=1,
 do_sample=True,
 top_k=50,
 top_p=0.95,
 temperature=1.0,
 pad_token_id=generator.tokenizer.eos_token_id
)
 generated_text = outputs[0]['generated_text']
 generated_texts.append({
 "model_name": model_name,
 "prompt": prompt,
 "generated_text": generated_text
 })

Save the outputs to a CSV file
df = pd.DataFrame(generated_texts)
Extract norm_type and variant from model_name
norm_type, variant = extract_norm_type_variant(model_name)
output_filename = f"{norm_type}_{variant}_generated_texts.csv"
output_filepath = os.path.join(output_dir, output_filename)
df.to_csv(output_filepath, index=False)
print(f"Saved generated texts to {output_filepath}\n")

def extract_norm_type_variant(model_name):
 """
 Extracts norm_type and variant from the model_name string.
 """
 try:
 # Assuming the model_name format is: shng2025/GPT-Valkyrie_<norm_type>-124m__<variant>_
 base_name = model_name.split('/')[-1]
 parts = base_name.split('__')
 config_part = parts[0] # e.g., 'GPT-Valkyrie_RMSN-124m'
 variant = parts[1] if len(parts) > 1 else 'Unknown'
 norm_type = config_part.split('_')[1].split('-')[0] # e.g., 'RMSN'
 return norm_type, variant
 except Exception as e:
 print(f"Error extracting norm_type and variant from model_name '{model_name}': {e}")
 return "Unknown", "Unknown"

if __name__ == "__main__":
 main()

```

## G.2 Summarisation (Traditional Metrics)

```
import torch
from transformers import GPT2TokenizerFast, GPT2LMHeadModel
from datasets import load_dataset
from tqdm import tqdm
import evaluate
import sacrebleu
import numpy as np
import pandas as pd

Function to load model and tokenizer
def load_model_and_tokenizer(model_name):
 tokenizer = GPT2TokenizerFast.from_pretrained(model_name)
 model = GPT2LMHeadModel.from_pretrained(model_name)
 tokenizer.pad_token = tokenizer.eos_token # Ensure pad_token is set
 model.config.pad_token_id = tokenizer.eos_token_id # Set pad_token_id in model config
 return model, tokenizer

Updated function to generate summary
def generate_summary(model, tokenizer, text, max_input_length=874, max_summary_length=150):
 """
 Generates a summary for the given text using the fine-tuned model.
 """
 # Prepare the prompt
 prompt = f"summarize: {text} summary:"

 # Tokenize the prompt
 inputs = tokenizer(
 prompt,
 return_tensors="pt",
 max_length=max_input_length,
 truncation=True,
 padding=False # Do not pad here
)

 input_ids = inputs.input_ids.to(model.device)
 attention_mask = inputs.attention_mask.to(model.device)

 # Record the length of the prompt
 prompt_length = input_ids.size(1)

 # Generate summary
 with torch.no_grad():
 outputs = model.generate(
 input_ids=input_ids,
 attention_mask=attention_mask,
 max_new_tokens=max_summary_length,
 num_beams=5,
 no_repeat_ngram_size=3,
 early_stopping=True,
 do_sample=False,
 pad_token_id=tokenizer.eos_token_id,
 eos_token_id=tokenizer.eos_token_id,
```

```

)

Extract generated token IDs
generated_ids = outputs[0]

Extract summary token IDs (tokens beyond the prompt)
summary_ids = generated_ids[prompt_length:]

Decode the summary tokens
generated_summary = tokenizer.decode(summary_ids, skip_special_tokens=True).strip()

Truncate the summary if it exceeds max_summary_length
generated_summary_tokens = tokenizer.tokenize(generated_summary)
if len(generated_summary_tokens) > max_summary_length:
 generated_summary_tokens = generated_summary_tokens[:max_summary_length]
generated_summary = tokenizer.convert_tokens_to_string(generated_summary_tokens)

Decode the full generated output for manual inspection (optional)
full_output = tokenizer.decode(generated_ids, skip_special_tokens=True)

return generated_summary, full_output

Function to compute metrics
def compute_metrics(predictions, references):
 """
 Compute ROUGE and BLEU metrics for summarization.
 """
 rouge = evaluate.load("rouge")

 # Compute ROUGE scores
 rouge_result = rouge.compute(
 predictions=predictions,
 references=references,
 use_stemmer=True,
 use_aggregator=True # Ensure aggregated scores
)

 # Compute BLEU scores using SacreBLEU with smoothing
 bleu_scores = sacrebleu.corpus_bleu(
 predictions,
 [references],
 smooth_method='exp', # Exponential smoothing
 smooth_value=0.1,
 force=True, # Force compute even if length mismatch
 lowercase=True, # Normalize case
 tokenize='13a' # Tokenizer type (SacreBLEU default)
)
 bleu_score = bleu_scores.score

 # Aggregate the results
 result = {
 "rouge1": round(rouge_result["rouge1"], 4),
 "rouge2": round(rouge_result["rouge2"], 4),

```

```

 "rougeL": round(rouge_result["rougeL"], 4),
 "bleu": round(bleu_score, 4)
}

Optional: Calculate average prediction length
prediction_lens = [len(pred.split()) for pred in predictions]
result["gen_len"] = round(np.mean(prediction_lens), 4)

return result

Function to evaluate the model
def evaluate_model(model_name, dataset, max_input_length=874, max_summary_length=150):
 """
 Evaluates the fine-tuned model on the provided dataset.
 """
 print(f"Evaluating model: {model_name}")
 try:
 model, tokenizer = load_model_and_tokenizer(model_name)
 except Exception as e:
 print(f"Error loading model {model_name}: {e}")
 return None

 device = 'cuda' if torch.cuda.is_available() else 'cpu'
 model.to(device)
 model.eval()

 generated_summaries = []
 reference_summaries_full = []
 reference_summaries_truncated = []

 for i, example in enumerate(tqdm(dataset, desc=f"Generating summaries for {model_name}")):
 text = example['text']
 reference_full = example['summary'] # Original full-length reference summary

 # Truncate the input text
 max_input_tokens = max_input_length
 text_tokens = tokenizer.tokenize(text)
 if len(text_tokens) > max_input_tokens:
 text_tokens = text_tokens[:max_input_tokens]
 text = tokenizer.convert_tokens_to_string(text_tokens)

 # Truncate the reference summary
 reference_tokens = tokenizer.tokenize(reference_full)
 if len(reference_tokens) > max_summary_length:
 reference_tokens_truncated = reference_tokens[:max_summary_length]
 reference_truncated =
 tokenizer.convert_tokens_to_string(reference_tokens_truncated)
 else:
 reference_truncated = reference_full

 generated_summary, _ = generate_summary(model, tokenizer, text, max_input_length,
max_summary_length)

```

```

if generated_summary:
 generated_summaries.append(generated_summary)
 reference_summaries_full.append(reference_full)
 reference_summaries_truncated.append(reference_truncated)
else:
 print(f"Warning: Empty summary generated for example {i+1}")

if not generated_summaries:
 print(f"Warning: No valid summaries generated for {model_name}")
 return None

Compute metrics using full reference summaries
scores_full = compute_metrics(generated_summaries, reference_summaries_full)
Compute metrics using truncated reference summaries
scores_truncated = compute_metrics(generated_summaries, reference_summaries_truncated)

Combine the scores into a single dictionary
scores = {
 'rouge1_full': scores_full['rouge1'],
 'rouge2_full': scores_full['rouge2'],
 'rougeL_full': scores_full['rougeL'],
 'bleu_full': scores_full['bleu'],
 'rouge1_truncated': scores_truncated['rouge1'],
 'rouge2_truncated': scores_truncated['rouge2'],
 'rougeL_truncated': scores_truncated['rougeL'],
 'bleu_truncated': scores_truncated['bleu'],
 'gen_len': scores_full['gen_len'] # Generated summary length is the same
}

print(f"\nScores for {model_name} (using full reference summaries):")
for metric in ['rouge1_full', 'rouge2_full', 'rougeL_full', 'bleu_full']:
 print(f"{metric}: {scores[metric]:.4f}")

print(f"\nScores for {model_name} (using truncated reference summaries):")
for metric in ['rouge1_truncated', 'rouge2_truncated', 'rougeL_truncated',
'bleu_truncated']:
 print(f"{metric}: {scores[metric]:.4f}")

return scores

Main execution block
if __name__ == "__main__":
 # Load the BillSum dataset (ca_test split)
 print("Loading the BillSum dataset...")
 dataset = load_dataset("billsum", split="ca_test")
 print(f"Loaded dataset with {len(dataset)} examples.")

 # Use only the first 100 examples
 dataset = dataset.select(range(100))
 print(f"Selected first {len(dataset)} examples for evaluation.")

 # Define the variants and norm types

```

```

variants = ["baseModel", "noNorm", "AttnOnly", "FFNonly"]
norm_types = ["LN", "RMSN"]

Initialize a list to store results
results = []

Loop over norm types and variants
for norm_type in norm_types:
 for variant in variants:
 model_name = f"shng2025/GPT-Valkyrie_{norm_type}-124m_{variant}_Billsum"
 print(f"\nProcessing model: {model_name}")

 # Evaluate the model
 scores = evaluate_model(
 model_name,
 dataset,
 max_input_length=874,
 max_summary_length=150
)

 if scores:
 # Add model information to the scores
 scores['model_name'] = model_name
 scores['norm_type'] = norm_type
 scores['variant'] = variant

 # Append to the results
 results.append(scores)
 else:
 print(f"Skipping model {model_name} due to errors.")

Save the results to a CSV file
df = pd.DataFrame(results)
df.to_csv('evaluation_results.csv', index=False)
print("\nEvaluation results saved to evaluation_results.csv")

```

### G.3 Summarisation (GPT-4o Metrics)

```

import torch
from transformers import GPT2TokenizerFast, GPT2LMHeadModel
from datasets import load_dataset
from tqdm import tqdm
import pandas as pd

Function to load model and tokenizer
def load_model_and_tokenizer(model_name):
 """
 Loads the GPT-2 model and tokenizer from the specified HuggingFace repository.
 """
 tokenizer = GPT2TokenizerFast.from_pretrained(model_name)
 model = GPT2LMHeadModel.from_pretrained(model_name)

```

```

tokenizer.pad_token = tokenizer.eos_token # Ensure pad_token is set
model.config.pad_token_id = tokenizer.eos_token_id # Set pad_token_id in model config
return model, tokenizer

Function to generate summary
def generate_summary(model, tokenizer, text, max_input_length=874, max_summary_length=150):
 """
 Generates a summary for the given text using the fine-tuned model.
 """
 # Prepare the prompt
 prompt = f"summarize: {text} summary:"

 # Tokenize the prompt without padding
 inputs = tokenizer(
 prompt,
 return_tensors="pt",
 max_length=max_input_length,
 truncation=True,
 padding=False # Do not pad here
)

 input_ids = inputs.input_ids.to(model.device)
 attention_mask = inputs.attention_mask.to(model.device)

 # Record the length of the prompt
 prompt_length = input_ids.size(1)

 # Generate summary
 with torch.no_grad():
 outputs = model.generate(
 input_ids=input_ids,
 attention_mask=attention_mask,
 max_new_tokens=max_summary_length,
 num_beams=5, # Beam search for better quality
 no_repeat_ngram_size=3, # Prevent repetition
 early_stopping=True,
 do_sample=False, # Disable sampling for determinism
 pad_token_id=tokenizer.eos_token_id,
 eos_token_id=tokenizer.eos_token_id,
)

 # Extract generated token IDs
 generated_ids = outputs[0]

 # Extract summary token IDs (tokens beyond the prompt)
 summary_ids = generated_ids[prompt_length:]

 # Decode the summary tokens
 generated_summary = tokenizer.decode(summary_ids, skip_special_tokens=True).strip()

 # Truncate the generated summary to max_summary_length tokens
 generated_summary_tokens = tokenizer.tokenize(generated_summary)
 if len(generated_summary_tokens) > max_summary_length:

```

```

generated_summary_tokens = generated_summary_tokens[:max_summary_length]
generated_summary = tokenizer.convert_tokens_to_string(generated_summary_tokens)

Decode the full generated output for manual inspection (optional)
full_output = tokenizer.decode(generated_ids, skip_special_tokens=True)

return generated_summary, full_output

Main execution block
if __name__ == "__main__":
 # Load the BillSum dataset (ca_test split)
 print("Loading the BillSum dataset...")
 dataset = load_dataset("billsum", split="ca_test")
 print(f"Loaded dataset with {len(dataset)} examples.")

 # Use only the first 100 examples
 dataset = dataset.select(range(100))
 print(f"Selected first {len(dataset)} examples for evaluation.")

 # Define the variants and norm types
 variants = ["baseModel", "noNorm", "AttnOnly", "FFNonly"]
 norm_types = ["LN", "RMSN"]

 # Loop over norm types and variants
 for norm_type in norm_types:
 for variant in variants:
 model_name = f"shng2025/GPT-Valkyrie_{norm_type}-124m_{variant}_Billsum"
 print(f"\nProcessing model: {model_name}")

 # Load model and tokenizer
 try:
 model, tokenizer = load_model_and_tokenizer(model_name)
 except Exception as e:
 print(f"Error loading model {model_name}: {e}")
 continue

 device = 'cuda' if torch.cuda.is_available() else 'cpu'
 model.to(device)
 model.eval()

 # Initialize a list to store results for this model
 data = []

 for i, example in enumerate(tqdm(dataset, desc=f"Generating summaries for {model_name}")):
 original_text = example['text']
 original_reference = example['summary']

 # Truncate the input text
 max_input_tokens = 874 # Adjust based on your context window
 text_tokens = tokenizer.tokenize(original_text)
 if len(text_tokens) > max_input_tokens:
 text_tokens_truncated = text_tokens[:max_input_tokens]


```

```

 truncated_text = tokenizer.convert_tokens_to_string(text_tokens_truncated)
 else:
 truncated_text = original_text

 # Truncate the reference summary
 max_summary_tokens = 150 # Adjust based on your context window
 reference_tokens = tokenizer.tokenize(original_reference)
 if len(reference_tokens) > max_summary_tokens:
 reference_tokens_truncated = reference_tokens[:max_summary_tokens]
 truncated_reference =
tokenizer.convert_tokens_to_string(reference_tokens_truncated)
 else:
 truncated_reference = original_reference

 generated_summary, _ = generate_summary(model, tokenizer, truncated_text,
max_input_length=874, max_summary_length=150)

 data.append({
 'model_name': model_name,
 'norm_type': norm_type,
 'variant': variant,
 'input_text': original_text,
 'truncated_input': truncated_text,
 'reference_summary': original_reference,
 'truncated_reference_summary': truncated_reference,
 'generated_summary': generated_summary
 })

Save the data for this model to a separate CSV file
df = pd.DataFrame(data)
Create a filename based on norm_type and variant
filename = f"{norm_type}_{variant}_evaluation_data.csv"
df.to_csv(filename, index=False)
print(f"Data for model {model_name} saved to {filename}")

```

## G.4 Question Answering (Traditional Metrics)

```

from transformers import AutoModelForQuestionAnswering, AutoTokenizer
from datasets import load_dataset
from evaluate import load as load_metric
import torch
import numpy as np
import pandas as pd
from tqdm import tqdm

def load_model_and_tokenizer(model_name):
 tokenizer = AutoTokenizer.from_pretrained(model_name)
 model = AutoModelForQuestionAnswering.from_pretrained(model_name)
 return model, tokenizer

def evaluate_model(model_name, dataset, max_samples=100):
 model, tokenizer = load_model_and_tokenizer(model_name)

```

```

model.eval()
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
model.to(device)
all_predictions = []
all_references = []

for i, example in enumerate(tqdm(dataset, desc=f"Evaluating {model_name}):"
 if i >= max_samples:
 break

 inputs = tokenizer(
 example["question"],
 example["context"],
 return_tensors="pt",
 truncation=True,
 max_length=512,
 padding=True
).to(device)

 with torch.no_grad():
 outputs = model(**inputs)

 start_logits = outputs.start_logits[0].cpu().numpy()
 end_logits = outputs.end_logits[0].cpu().numpy()

 # Find the tokens with the highest `start` and `end` scores
 start_index = np.argmax(start_logits)
 end_index = np.argmax(end_logits)

 # Compute the score of the "no answer" option
 no_answer_score = start_logits[0] + end_logits[0]
 best_answer_score = start_logits[start_index] + end_logits[end_index]

 if no_answer_score > best_answer_score or end_index < start_index:
 answer = ""
 else:
 answer_tokens = inputs["input_ids"][0][start_index:end_index+1]
 answer = tokenizer.decode(answer_tokens, skip_special_tokens=True)

 # Collect predictions and references
 prediction = {"id": example["id"], "prediction_text": answer, "no_answer_probability": float(1 / (1 + np.exp(best_answer_score - no_answer_score)))}
 reference = {"id": example["id"], "answers": example["answers"]}

 all_predictions.append(prediction)
 all_references.append(reference)

squad_metric = load_metric("squad_v2")
results = squad_metric.compute(predictions=all_predictions, references=all_references)
return results

def main():
 variants = ["baseModel", "noNorm", "AttnOnly", "FFNonly"]

```

```

norm_types = ["LN", "RMSN"]
results = []
Load dataset
dataset = load_dataset("squad_v2", split="validation")
Limit to first 100 examples
dataset = dataset.select(range(100))
for norm_type in norm_types:
 for variant in variants:
 model_name = f"shng2025/GPT-Valkyrie_{norm_type}-124m_{variant}_SQuAD"
 print(f"\nEvaluating model {model_name}")
 try:
 metrics = evaluate_model(model_name, dataset, max_samples=100)
 metrics['model_name'] = model_name
 metrics['norm_type'] = norm_type
 metrics['variant'] = variant
 results.append(metrics)
 except Exception as e:
 print(f"Error evaluating model {model_name}: {e}")
 continue
Save results to CSV
df = pd.DataFrame(results)
df.to_csv('squad_evaluation_results.csv', index=False)
print("\nEvaluation results saved to squad_evaluation_results.csv")

if __name__ == "__main__":
 main()

```

## G.5 Question Answering (GPT-4o Metrics)

```

from transformers import AutoModelForQuestionAnswering, AutoTokenizer
from datasets import load_dataset
from evaluate import load as load_metric
import torch
import numpy as np
import pandas as pd
from tqdm import tqdm

def load_model_and_tokenizer(model_name):
 tokenizer = AutoTokenizer.from_pretrained(model_name)
 model = AutoModelForQuestionAnswering.from_pretrained(model_name)
 return model, tokenizer

def evaluate_model(model_name, dataset, norm_type, max_samples=100):
 model, tokenizer = load_model_and_tokenizer(model_name)
 model.eval()
 device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
 model.to(device)
 all_predictions = []
 all_references = []
 data_for_gpt4 = []

 for i, example in enumerate(tqdm(dataset, desc=f"Evaluating {model_name}")):

```

```

if i >= max_samples:
 break

question = example["question"]
context = example["context"]
reference_answers = example["answers"]["text"]

inputs = tokenizer(
 question,
 context,
 return_tensors="pt",
 truncation=True,
 max_length=512,
 padding=True
).to(device)

with torch.no_grad():
 outputs = model(**inputs)

start_logits = outputs.start_logits[0].cpu().numpy()
end_logits = outputs.end_logits[0].cpu().numpy()

Find the tokens with the highest `start` and `end` scores
start_index = np.argmax(start_logits)
end_index = np.argmax(end_logits)

Compute the score of the "no answer" option
no_answer_score = start_logits[0] + end_logits[0]
best_answer_score = start_logits[start_index] + end_logits[end_index]

if no_answer_score > best_answer_score or end_index < start_index:
 answer = ""
else:
 answer_tokens = inputs["input_ids"][0][start_index:end_index+1]
 answer = tokenizer.decode(answer_tokens, skip_special_tokens=True)

Collect predictions and references
prediction = {"id": example["id"], "prediction_text": answer, "no_answer_probability": float(1 / (1 + np.exp(best_answer_score - no_answer_score)))}
reference = {"id": example["id"], "answers": example["answers"]}

all_predictions.append(prediction)
all_references.append(reference)

Collect data for GPT-4 evaluation
data_for_gpt4.append({
 "model_name": model_name, # Add the model_name here
 "question": question,
 "context": context,
 "reference_answers": reference_answers,
 "generated_answer": answer
})

```

```

Save data for GPT-4 evaluation
df = pd.DataFrame(data_for_gpt4)
Create a filename based on norm_type and variant
norm_type = model_name.split("_")[2].split("-")[0]
variant = model_name.split("__")[1]
filename = f"{norm_type}_{variant}_gpt4_evaluation_data.csv"
df.to_csv(filename, index=False)
print(f"Data for GPT-4 evaluation saved to {filename}")

Compute evaluation metrics
squad_metric = load_metric("squad_v2")
results = squad_metric.compute(predictions=all_predictions, references=all_references)
return results

def main():
 variants = ["baseModel", "noNorm", "AttnOnly", "FFNonly"]
 norm_types = ["LN", "RMSN"]
 results = []
 # Load dataset
 dataset = load_dataset("squad_v2", split="validation")
 # Limit to first 100 examples
 dataset = dataset.select(range(100))
 for norm_type in norm_types:
 for variant in variants:
 model_name = f"shng2025/GPT-Valkyrie_{norm_type}-124m_{variant}_SQuAD"
 print(f"\nEvaluating model {model_name}")
 try:
 metrics = evaluate_model(model_name, dataset, norm_type, max_samples=100)
 metrics['model_name'] = model_name
 metrics['norm_type'] = norm_type
 metrics['variant'] = variant
 results.append(metrics)
 except Exception as e:
 print(f"Error evaluating model {model_name}: {e}")
 continue
 # Save results to CSV
 df = pd.DataFrame(results)
 df.to_csv('squad_evaluation_results.csv', index=False)
 print("\nEvaluation results saved to squad_evaluation_results.csv")

if __name__ == "__main__":
 main()

```

## Appendix H - GPT4o API

This section includes the code used to pass the CSV files generated from Appendix G into GPT-4o through OpenAI's API (OpenAI, 2024). The prompts used to guide GPT-4o as our evaluator are shown in the code blocks below.

### H.1 TextGen

```
import pandas as pd
```

```

import openai
import os
import time
from tqdm import tqdm
from google.colab import userdata
import re

Set your OpenAI API key
openai.api_key = userdata.get('OPENAI_API_KEY')

def get_gpt4_evaluation_text(model_name, norm_type, variant, prompt, generated_text):
 # Construct the prompt for GPT-4o evaluation
 evaluation_prompt = f"""
As an expert evaluator, your task is to assess the quality of a generated text based on the provided prompt.
Please focus on the following criteria:

1. **Creativity**: How original and imaginative is the text?
2. **Coherence**: Is the text logically consistent and does it flow well?
3. **Fluency**: Is the text well-written with proper grammar and style?
4. **Relevance**: Does the text stay on topic and relate to the prompt?
5. **Engagement**: Is the text interesting and engaging to read?

For each criterion, provide:
- **Score**: A number from 1 to 5 (1 is poor, 5 is excellent).
- **Explanation**: A brief justification for the score.

After evaluating each criterion, provide:
- **Overall Score**: The average of the five scores.
- **Overall Feedback**: A short overall feedback.

Additionally, provide a short paragraph commenting on the following columns:
- **Model Name**: {model_name}
- **Norm Type**: {norm_type}
- **Variant**: {variant}
- **Prompt**: [Prompt is provided above.]
- **Generated Text**: [Generated text is provided above.]

Please present your evaluation in the following structured format:
```
Model Name: {model_name}

Creativity Score: [1-5]
Creativity Explanation: [Your explanation]

Coherence Score: [1-5]
Coherence Explanation: [Your explanation]

Fluency Score: [1-5]
Fluency Explanation: [Your explanation]

Relevance Score: [1-5]
Relevance Explanation: [Your explanation]

Engagement Score: [1-5]
Engagement Explanation: [Your explanation]

Overall Score: [Average score]
Overall Feedback: [Your feedback]

Comments on Columns:
[Your short paragraph commenting on each column]
```

```

```

```
---  

**Prompt:**  

{prompt}  

---  

**Generated Text:**  

{generated_text}  

---  

"""
# Call the OpenAI API using the ChatCompletion endpoint
try:
    response = openai.ChatCompletion.create(
        model="gpt-4o", # Ensure gpt-4o is available to you
        messages=[
            {"role": "user", "content": evaluation_prompt}
        ],
        max_tokens=700, # As per your requirement
        temperature=0.0, # For deterministic output
    )
    evaluation = response['choices'][0]['message']['content']
    return evaluation
except openai.error.RateLimitError as e:
    print(f"Rate limit error: {e}")
    retry_after = int(e.headers.get("Retry-After", 5))
    print(f"Retrying after {retry_after} seconds...")
    time.sleep(retry_after)
    # Retry the request
    return get_gpt4_evaluation_text(model_name, norm_type, variant, prompt, generated_text)
except Exception as e:
    print(f"Error during OpenAI API call: {e}")
    return None

def parse_evaluation_text(evaluation_text):
    """
    Parses GPT-4o's evaluation text and extracts scores, explanations, and comments.
    """
    patterns = {
        'Creativity Score': r'Creativity Score:\s*(\d)',
        'Creativity Explanation': r'Creativity Explanation:\s*(.*?)\n\n',
        'Coherence Score': r'Coherence Score:\s*(\d)',
        'Coherence Explanation': r'Coherence Explanation:\s*(.*?)\n\n',
        'Fluency Score': r'Fluency Score:\s*(\d)',
        'Fluency Explanation': r'Fluency Explanation:\s*(.*?)\n\n',
        'Relevance Score': r'Relevance Score:\s*(\d)',
        'Relevance Explanation': r'Relevance Explanation:\s*(.*?)\n\n',
        'Engagement Score': r'Engagement Score:\s*(\d)',
        'Engagement Explanation': r'Engagement Explanation:\s*(.*?)\n\n',
        'Overall Score': r'Overall Score:\s*([\d\.]+)',
        'Overall Feedback': r'Overall Feedback:\s*(.*?)\n\n',
        'Comments on Columns': r'Comments on Columns:\s*(.*)', # Captures the paragraph
    }

    result = {}
    for key, pattern in patterns.items():
        match = re.search(pattern, evaluation_text, re.DOTALL)
        if match:
            result[key] = match.group(1).strip()

```

```

    else:
        result[key] = None
    return result

def evaluate_generated_texts():
    # Directory containing generated text CSV files
    input_dir = "generated_texts"
    output_dir = "evaluations"
    os.makedirs(output_dir, exist_ok=True)

    # List CSV files in the input directory
    csv_files = [f for f in os.listdir(input_dir) if f.endswith('.csv')]

    for csv_file in csv_files:
        input_filepath = os.path.join(input_dir, csv_file)
        print(f"Processing file: {input_filepath}")

        # Read the CSV file
        df = pd.read_csv(input_filepath)

        # Limit to the first 25 rows to manage costs
        df_limited = df.head(25)
        evaluations = []

        for idx, row in tqdm(df_limited.iterrows(), total=df_limited.shape[0], desc=f"Evaluating texts in {csv_file}"):
            model_name = row['model_name']
            prompt = row['prompt']
            generated_text = row['generated_text']

            # Extract norm_type and variant from model_name
            norm_type, variant = extract_norm_type_variant(model_name)

            # Optionally truncate generated_text if too long
            max_text_length = 1000 # Adjust as needed
            if len(generated_text.split()) > max_text_length:
                generated_text = ' '.join(generated_text.split()[:max_text_length]) + "..."

            # Get GPT-4o evaluation
            evaluation_text = get_gpt4_evaluation_text(
                model_name,
                norm_type,
                variant,
                prompt,
                generated_text
            )

            if evaluation_text:
                parsed = parse_evaluation_text(evaluation_text)
                parsed['model_name'] = model_name
                parsed['norm_type'] = norm_type
                parsed['variant'] = variant
                evaluations.append(parsed)
                time.sleep(1) # To respect API rate limits
            else:
                evaluations.append({
                    'model_name': model_name,
                    'norm_type': norm_type,
                    'variant': variant,
                    'Creativity Score': None,
                    'Creativity Explanation': None,
                    'Coherence Score': None,
                })

```

```

'Coherence Explanation': None,
'Fluency Score': None,
'Fluency Explanation': None,
'Relevance Score': None,
'Relevance Explanation': None,
'Engagement Score': None,
'Engagement Explanation': None,
'Overall Score': None,
'Overall Feedback': None,
'Comments on Columns': 'Error or Empty Response'
})

# Save evaluations to a new CSV file
eval_df = pd.DataFrame(evaluations)
output_filename = f"{norm_type}_{variant}_evaluations.csv"
output_filepath = os.path.join(output_dir, output_filename)
eval_df.to_csv(output_filepath, index=False)
print(f"Saved evaluations to {output_filepath}\n")

def extract_norm_type_variant(model_name):
    """
    Extracts norm_type and variant from the model_name string.
    """
    try:
        # Assuming the model_name format is: shng2025/GPT-Valkyrie_<norm_type>-124m__<variant>__
        base_name = model_name.split('/')[-1]
        parts = base_name.split('__')
        config_part = parts[0] # e.g., 'GPT-Valkyrie_RMSN-124m'
        variant = parts[1] if len(parts) > 1 else 'Unknown'
        norm_type = config_part.split('_')[1].split('-')[0] # e.g., 'RMSN'
        return norm_type, variant
    except Exception as e:
        print(f"Error extracting norm_type and variant from model_name '{model_name}': {e}")
        return "Unknown", "Unknown"

if __name__ == "__main__":
    evaluate_generated_texts()

```

H.2 Summarisation (Billsum)

```
import pandas as pd
import openai
import os
import time
from tqdm import tqdm
from google.colab import userdata

# Set your OpenAI API key
openai.api_key = userdata.get('OPENAI_API_KEY')
```

```
def get_gpt4_evaluation_summary(model_name, norm_type, variant, truncated_input, generated_summary):
```

```
    # Construct the prompt for GPT-4 evaluation
```

```
    prompt = f"""
```

As an expert evaluator, your task is to assess the quality of a generated summary based on the provided truncated input text. Please focus on the following criteria:

1. **Relevance**: Does the summary capture the main points of the truncated input?
2. **Conciseness**: Is the summary succinct without unnecessary details?
3. **Fluency**: Is the summary well-written with proper grammar and style?
4. **Accuracy**: Does the summary accurately represent the content of the truncated input without errors?
5. **Coherence**: Is the summary logically organized and easy to understand?

For each criterion, provide:

- **Score**: A number from 1 to 5 (where 1 is poor and 5 is excellent).
- **Explanation**: A brief justification for the score.

After evaluating each criterion, provide:

- **Overall Score**: The average of the five scores.
- **Overall Feedback**: A short overall feedback.

Additionally, provide a short paragraph commenting on the following columns:

- **Model Name**: {model_name}
- **Norm Type**: {norm_type}
- **Variant**: {variant}
- **Truncated Input**: [Truncated input is provided above.]
- **Generated Summary**: [Generated summary is provided above.]

Please present your evaluation in the following structured format:

Model Name: {model_name}

Relevance Score: [1-5]

Relevance Explanation: [Your explanation]

Conciseness Score: [1-5]

Conciseness Explanation: [Your explanation]

Fluency Score: [1-5]

Fluency Explanation: [Your explanation]

Accuracy Score: [1-5]

Accuracy Explanation: [Your explanation]

Coherence Score: [1-5]

Coherence Explanation: [Your explanation]

Overall Score: [Average score]

Overall Feedback: [Your feedback]

```

Comments on Columns:
[Your short paragraph commenting on each column]
```

Truncated Input:

{truncated_input}

Generated Summary:

{generated_summary}

"""
Call the OpenAI API using the ChatCompletion endpoint
try:
 response = openai.ChatCompletion.create(
 model="gpt-4o",
 messages=[
 {"role": "user", "content": prompt}
],
 max_tokens=700, # As per your requirement
 temperature=0.0, # For deterministic output
)
 evaluation = response['choices'][0]['message']['content']
 return evaluation
except openai.error.RateLimitError as e:
 print(f"Rate limit error: {e}")
 retry_after = int(e.headers.get("Retry-After", 5))
 print(f"Retrying after {retry_after} seconds...")
 time.sleep(retry_after)
 # Retry the request
 return get_gpt4_evaluation_summary(model_name, norm_type, variant, truncated_input, generated_summary)
except Exception as e:
 print(f"Error during OpenAI API call: {e}")
 return None

def parse_evaluation_summary(evaluation_text):
 """
 Parses GPT-4's evaluation text and extracts scores, explanations, and comments.
 """
 patterns = {
 'Relevance Score': r'Relevance Score:\s*(\d)',
 'Relevance Explanation': r'Relevance Explanation:\s*(.*?)\n\n',
 'Conciseness Score': r'Conciseness Score:\s*(\d)',
 'Conciseness Explanation': r'Conciseness Explanation:\s*(.*?)\n\n',
 'Fluency Score': r'Fluency Score:\s*(\d)',
 'Fluency Explanation': r'Fluency Explanation:\s*(.*?)\n\n',
 'Accuracy Score': r'Accuracy Score:\s*(\d)',
 'Accuracy Explanation': r'Accuracy Explanation:\s*(.*?)\n\n',
 'Coherence Score': r'Coherence Score:\s*(\d)',
 'Coherence Explanation': r'Coherence Explanation:\s*(.*?)\n\n',
 'Overall Score': r'Overall Score:\s*([\d.]+)',
 'Overall Feedback': r'Overall Feedback:\s*(.*?)\n\n',
 'Comments on Columns': r'Comments on Columns:\s*(.*')', # Captures the paragraph
 }

 result = {}
 for key, pattern in patterns.items():

```

```

match = re.search(pattern, evaluation_text, re.DOTALL)
if match:
 result[key] = match.group(1).strip()
else:
 result[key] = None
return result

def evaluate_summaries():
 variants = ["baseModel", "noNorm", "AttnOnly", "FFNonly"]
 norm_types = ["LN", "RMSN"]

 for norm_type in norm_types:
 for variant in variants:
 filename = f"./Summarization/Original/{norm_type}_{variant}_evaluation_data_modified.csv"
 print(f"Processing file: {filename}")

 if not os.path.exists(filename):
 print(f"File {filename} does not exist. Skipping.")
 continue

 # Read the CSV file
 df = pd.read_csv(filename)
 # Limit to the first 25 rows to manage costs
 df_limited = df.head(25)
 evaluations = []

 for idx, row in tqdm(df_limited.iterrows(), total=df_limited.shape[0], desc=f"Evaluating Summaries
for {filename}"):
 model_name = row['model_name']
 norm_type = row['norm_type']
 variant = row['variant']
 truncated_input = row['truncated_input']
 generated_summary = row['generated_summary']

 # Optionally truncate input if too long
 max_input_length = 1000 # Adjust as needed
 if len(truncated_input.split()) > max_input_length:
 truncated_input = ' '.join(truncated_input.split()[:max_input_length]) + "..."

 # Get GPT-4 evaluation
 evaluation_text = get_gpt4_evaluation_summary(
 model_name,
 norm_type,
 variant,
 truncated_input,
 generated_summary
)

 if evaluation_text:
 parsed = parse_evaluation_summary(evaluation_text)
 parsed['model_name'] = model_name
 parsed['norm_type'] = norm_type
 parsed['variant'] = variant
 evaluations.append(parsed)
 time.sleep(1) # To respect API rate limits
 else:
 evaluations.append({
 'model_name': model_name,
 'norm_type': norm_type,
 'variant': variant,
 'Relevance Score': None,
 'Relevance Explanation': None,

```

```

 'Conciseness Score': None,
 'Conciseness Explanation': None,
 'Fluency Score': None,
 'Fluency Explanation': None,
 'Accuracy Score': None,
 'Accuracy Explanation': None,
 'Coherence Score': None,
 'Coherence Explanation': None,
 'Overall Score': None,
 'Overall Feedback': None,
 'Comments on Columns': 'Error or Empty Response'
 })

Save evaluations to a new CSV file
eval_df = pd.DataFrame(evaluations)
parsed_eval_filename =
f"./Summarization/{norm_type}_{variant}_gpt4_summary_parsed_evaluations.csv"
eval_df.to_csv(parsed_eval_filename, index=False)
print(f"Saved parsed evaluations to {parsed_eval_filename}")

if __name__ == "__main__":
 evaluate_summaries()

```

### H.3 Question Answering (SQuAD)

```

import pandas as pd
import openai
import os
import time
from tqdm import tqdm
from google.colab import userdata
import re

Set your OpenAI API key
openai.api_key = userdata.get('OPENAI_API_KEY')

def get_gpt4_evaluation_qa(model_name, norm_type, variant, question, context, reference_answers,
generated_answer):
 # Construct the prompt with instructions for commenting on each column
 prompt = f"""
As an expert evaluator, your task is to assess the quality of an answer generated by a question-answering
system based on the provided context. Please focus on the following criteria:

1. **Correctness**: Is the answer correct based on the context?
2. **Completeness**: Does the answer fully address the question?
3. **Relevance**: Is the answer relevant to the question and context?
4. **Fluency**: Is the answer well-written with proper grammar and style?
5. **Conciseness**: Is the answer concise and to the point?

For each criterion, provide:
- **Score**: A number from 1 to 5 (where 1 is poor and 5 is excellent).
- **Explanation**: A brief justification for the score.

After evaluating each criterion, provide:
- **Overall Score**: The average of the five scores.
- **Overall Feedback**: A short overall feedback.

Additionally, provide a short paragraph commenting on the following columns:
- **Model Name**: {model_name}

```

```

- **Norm Type**: {norm_type}
- **Variant**: {variant}
- **Question**: {question}
- **Context**: {context}
- **Reference Answers**: {reference_answers}
- **Generated Answer**: {generated_answer}

Please present your evaluation in the following structured format:

```
Model Name: {model_name}

Correctness Score: [1-5]
Correctness Explanation: [Your explanation]

Completeness Score: [1-5]
Completeness Explanation: [Your explanation]

Relevance Score: [1-5]
Relevance Explanation: [Your explanation]

Fluency Score: [1-5]
Fluency Explanation: [Your explanation]

Conciseness Score: [1-5]
Conciseness Explanation: [Your explanation]

Overall Score: [Average score]
Overall Feedback: [Your feedback]

Comments on Columns:
[Your short paragraph commenting on each column]
```

```
-- 
**Model Name:** {model_name}

-- 
**Question:** {question}

-- 
**Context:** {context}

-- 
**Reference Answers:** {reference_answers}

-- 
**Generated Answer:** {generated_answer}

```
 # Call the OpenAI API using the ChatCompletion endpoint
 try:
 response = openai.ChatCompletion.create(
 model="gpt-4o",
 messages=[
 {"role": "user", "content": prompt}

```

```

],
 max_tokens=700, # Adjusted to accommodate additional comments
 temperature=0.0, # For deterministic output
)
 evaluation = response['choices'][0]['message']['content']
 return evaluation
except openai.error.RateLimitError as e:
 print(f"Rate limit error: {e}")
 retry_after = int(e.headers.get("Retry-After", 5))
 print(f"Retrying after {retry_after} seconds...")
 time.sleep(retry_after)
 # Retry the request
 return get_gpt4_evaluation_qa(model_name, norm_type, variant, question, context, reference_answers,
generated_answer)
except Exception as e:
 print(f"Error during OpenAI API call: {e}")
 return None

def parse_evaluation_qa(evaluation_text):
 """
 Parses GPT-4's evaluation text and extracts scores, explanations, and comments.
 """
 patterns = {
 'Correctness Score': r'Correctness Score:\s*(\d)',
 'Correctness Explanation': r'Correctness Explanation:\s*(.*?)\n\n',
 'Completeness Score': r'Completeness Score:\s*(\d)',
 'Completeness Explanation': r'Completeness Explanation:\s*(.*?)\n\n',
 'Relevance Score': r'Relevance Score:\s*(\d)',
 'Relevance Explanation': r'Relevance Explanation:\s*(.*?)\n\n',
 'Fluency Score': r'Fluency Score:\s*(\d)',
 'Fluency Explanation': r'Fluency Explanation:\s*(.*?)\n\n',
 'Conciseness Score': r'Conciseness Score:\s*(\d)',
 'Conciseness Explanation': r'Conciseness Explanation:\s*(.*?)\n\n',
 'Overall Score': r'Overall Score:\s*(\d)\.?',
 'Overall Feedback': r'Overall Feedback:\s*(.*?)\n\n',
 'Comments on Columns': r'Comments on Columns:\s*(.*?)" # Captures the paragraph
 }

 result = {}
 for key, pattern in patterns.items():
 match = re.search(pattern, evaluation_text, re.DOTALL)
 if match:
 result[key] = match.group(1).strip()
 else:
 result[key] = None
 return result

def evaluate_qa():
 variants = ["baseModel", "noNorm", "AttnOnly", "FFNonly"]
 norm_types = ["LN", "RMSN"]
 results = []

 for norm_type in norm_types:
 for variant in variants:
 filename = f"./modified_QA-source/{norm_type}_{variant}_gpt4_evaluation_data_modified.csv"
 print(f"Processing file: {filename}")

 if not os.path.exists(filename):
 print(f"File {filename} does not exist. Skipping.")
 continue

```

```

Read the CSV file
df = pd.read_csv(filename)
Limit to the first 25 rows to manage costs
df_limited = df.head(25) # <-- no limits placed on QA eval
evaluations = []

for idx, row in tqdm(df_limited.iterrows(), total=df_limited.shape[0], desc=f"Evaluating QA for {filename}"):
 model_name = row['model_name']
 norm_type = row['norm_type']
 variant = row['variant']
 question = row['question']
 context = row['context']
 reference_answers = row['reference_answers']
 generated_answer = row['generated_answer']

 # Convert reference_answers from string representation to list if necessary
 if isinstance(reference_answers, str):
 try:
 reference_answers = eval(reference_answers)
 if not isinstance(reference_answers, list):
 reference_answers = [reference_answers]
 except:
 reference_answers = [reference_answers]
 elif not isinstance(reference_answers, list):
 reference_answers = [reference_answers]

 # Optionally truncate context if too long
 max_context_length = 2000 # Adjust as needed
 if len(context.split()) > max_context_length:
 context = ' '.join(context.split()[:max_context_length]) + "..."

 # Get GPT-4 evaluation
 evaluation_text = get_gpt4_evaluation_qa(model_name, norm_type, variant, question, context,
 reference_answers, generated_answer)

 if evaluation_text:
 parsed = parse_evaluation_qa(evaluation_text)
 parsed['model_name'] = model_name
 parsed['norm_type'] = norm_type
 parsed['variant'] = variant
 evaluations.append(parsed)
 time.sleep(1) # To respect API rate limits
 else:
 evaluations.append({
 'model_name': model_name,
 'norm_type': norm_type,
 'variant': variant,
 'Correctness Score': None,
 'Correctness Explanation': None,
 'Completeness Score': None,
 'Completeness Explanation': None,
 'Relevance Score': None,
 'Relevance Explanation': None,
 'Fluency Score': None,
 'Fluency Explanation': None,
 'Conciseness Score': None,
 'Conciseness Explanation': None,
 'Overall Score': None,
 'Overall Feedback': None,
 'Comments on Columns': 'Error or Empty Response'
 })

```

```
)

Save evaluations to a new CSV file
eval_df = pd.DataFrame(evaluations)
parsed_eval_filename = f"{norm_type}_{variant}_gpt4_qa_parsed_evaluations.csv"
eval_df.to_csv(parsed_eval_filename, index=False)
print(f"Saved parsed evaluations to {parsed_eval_filename}")

if __name__ == "__main__":
 evaluate_qa()
```

## Appendix I - Text Generation GPT-4o CSV Files

Note: We are displaying only 20 out of the 200 prompts, as including the full CSV file in this document would extend it to over 600 pages. This large volume of data highlights the effectiveness of GPT-4o for evaluation purposes. The complete CSV files are available in our GitHub repository.<sup>66</sup>

| model_name | t                                                                  | task                                  | prompt                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | generated_text                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | O<br>v<br>e<br>r<br>a<br>l<br>I<br>S<br>c<br>o<br>r<br>e | C<br>r<br>e<br>a<br>t<br>i<br>v<br>i<br>t<br>e<br>y<br>S<br>c<br>o<br>r<br>e | C<br>o<br>h<br>e<br>r<br>e<br>n<br>c<br>e<br>S<br>c<br>o<br>r<br>e | F<br>l<br>u<br>e<br>c<br>y<br>S<br>c<br>o<br>r<br>e | R<br>e<br>l<br>e<br>v<br>a<br>n<br>c<br>e<br>S<br>c<br>o<br>r<br>e | E<br>n<br>g<br>a<br>m<br>e<br>n<br>t<br>S<br>c<br>o<br>r<br>e | Overall Feedback                                                                                                                                                                                                                                                     | Comments on Columns                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|------------|--------------------------------------------------------------------|---------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|------------------------------------------------------------------------------|--------------------------------------------------------------------|-----------------------------------------------------|--------------------------------------------------------------------|---------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| LN_noNorm  | T<br>e<br>x<br>t<br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | Once upon a time in a distant galaxy, | Once upon a time in a distant galaxy, Phase HL7- is-the for —.—♦. we or is are and the sub-?♦ otherwise even./?                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | Once upon a time in a distant galaxy, Phase HL7- is-the for —.—♦. we or is are and the sub-?♦ otherwise even./?                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 1<br>.0                                                  | 1                                                                            | 1                                                                  | 1                                                   | 1                                                                  | 1                                                             | The generated text is of very poor quality. It fails to meet any of the criteria for a good piece of writing. It is incoherent, irrelevant, and lacks any form of creativity or engagement. The text appears to be a malfunction or error in the generation process. | - **Model Name**: The model name suggests it is a variant of GPT, but the output does not reflect the capabilities typically associated with such models.<br>- **Norm Type**: LN (Layer Normalization) is a common technique in neural networks, but its effectiveness is not evident in this output.<br>- **Variant**: The "noNorm" variant might be contributing to the poor quality of the text, as normalization techniques often help in stabilizing and improving the performance of models.<br>- **Prompt**: The prompt is clear and sets up a potential for a creative and engaging story, but the generated text does not follow through.<br>- **Generated Text**: The generated text is essentially gibberish, with no meaningful content or structure. It does not fulfill the expectations set by the prompt. |
| LN_noNorm  | T<br>e<br>x<br>t<br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | The secret to happiness is            | The secret to happiness is tort- ....♦ ()..¶.¶7.. or E♦, SB:- and Un— otherwise♦, inf- F? (D-based:1. infin. sub- ...M..."-e. management—♦..4M.S.♦-,♦(F...InM. inf♦-♦5-based.Figure-1♦10-C-3-1AThis♦♦Do.4-♦-2-D-. development♦A-Figure--10-5--5-18--based♦-----/2 ♦♦-NC-E--S.64.4-....7-D. by.... ...1- to:-based(6WeM (191/4-FigureB-S-S-(Di---♦A,-F and♦-3-E-5--E♦As-1B-6--.InS181525 exchange-F♦Av-2-H3B-12-P♦4--B---B.5.C?8-(11A. 3-2D--5-6-8-K3.l(3-P8--2A-C.♦-A6-E-C9-B♦-5-This 7M-12B--1 liter-1-N12Fe2-19-I10Phase-8121340st-2Learning6S- (E6-In-Sch-In♦-As-S-205-1♦-By7-2When permitting5.-S-S-In-St.-B1OverP-8WeS.-Di FTA Phase functionB3-The-1318As-S(1♦-s2S5--s-2♦IfMulti-1.♦ (KDiLength7M-S-A♦The-4As-C.1-Do3E-5-N2 | The secret to happiness is tort- ....♦ ()..¶.¶7.. or E♦, SB:- and Un— otherwise♦, inf- F? (D-based:1. infin. sub- ...M..."-e. management—♦..4M.S.♦-,♦(F...InM. inf♦-♦5-based.Figure-1♦10-C-3-1AThis♦♦Do.4-♦-2-D-. development♦A-Figure--10-5--5-18--based♦-----/2 ♦♦-NC-E--S.64.4-....7-D. by.... ...1- to:-based(6WeM (191/4-FigureB-S-S-(Di---♦A,-F and♦-3-E-5--E♦As-1B-6--.InS181525 exchange-F♦Av-2-H3B-12-P♦4--B---B.5.C?8-(11A. 3-2D--5-6-8-K3.l(3-P8--2A-C.♦-A6-E-C9-B♦-5-This 7M-12B--1 liter-1-N12Fe2-19-I10Phase-8121340st-2Learning6S- (E6-In-Sch-In♦-As-S-205-1♦-By7-2When permitting5.-S-S-In-St.-B1OverP-8WeS.-Di FTA Phase functionB3-The-1318As-S(1♦-s2S5--s-2♦IfMulti-1.♦ (KDiLength7M-S-A♦The-4As-C.1-Do3E-5-N2 | 1<br>.0                                                  | 1                                                                            | 1                                                                  | 1                                                   | 1                                                                  | 1                                                             | The generated text is of very poor quality. It does not exhibit any creativity, coherence, fluency, relevance, or engagement. It appears to be a random string of characters and symbols rather than a meaningful response to the prompt.                            | - **Model Name**: The model name suggests it is a variant of GPT with some specific modifications, but the output does not reflect any advanced capabilities.<br>- **Norm Type**: LN (Layer Normalization) is a common technique in neural networks, but its effectiveness is not evident in this output.<br>- **Variant**: The "noNorm" variant might indicate that certain normalization techniques were not applied, which could explain the poor quality of the text.<br>- **Prompt**: The prompt is clear and straightforward, asking for the secret to happiness.                                                                                                                                                                                                                                                   |

<sup>66</sup> <https://github.com/Ice-Citron/Extended-Essay-Appendix.git> by (Student, 2024)

|           |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |         |   |   |   |   |   |                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|-----------|------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           |                                                                        | debA-4c/S-1◆-A--What191341S-F12 material--61(k blockk◆How processesD,CA(3S-By-7st5CC--2-C-(b-B--TheA◆StS --4A1-A-As1-1.1-MPA-If20-InD2.-When2 and◆1-6-3-cKt2)3.4/Shas6st◆We-C-5--B6-2.◆-As-1-AsA4.1)Ks1.8-C9)K2 (4-8b,B5--5b,M◆When1.10 components code.3/C1.1 andWhen-C◆-3)◆3:B13).sWe(-5.A1.-                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |         |   |   |   |   |   | - **Generated Text**: The generated text is nonsensical and does not provide any meaningful information or insight related to the prompt.                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| LN_noNorm | T<br>e<br>x<br>t<br><br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | In the midst of the bustling city,? andS. e exchange exempted, and unE11., U22A- design- eds◆, K◆. to at enk., 1- are are- and a un◆D- so in sub-...In- E... for, otherwise, en1-◆-D- B-/).10--Figure20.◆-(◆-4....DCK-.K-◆◆-S -?-3◆D-◆-D-D-5..S-◆2-12-D.◆3-◆◆2-ItS-1-D ◆S-◆15-10S.31-I-D).WeB-3(M.M--F◆When code.-8-5..2A-S-..-2---◆In-B--The--1(1912-◆4-IB◆( In-In(111-◆When--1◆5.--18N.-S-2WeAreA---2M.6-3I nf,.10 dup-926.If fruit8-1-F-We7-1Fe2SK8-Washington pre-l9-2In-1WhenDiagrap-12-C(6R(Fe3-5DFe◆4D- Do2Read-5A◆When.-St.-2A-724-124023b-122022-3I f1/St19-2E-3When10A8-9:2A1531-As10-10-WeAre 1 0-There-A◆◆When◆-There-As-A--When-11-There ◆A4--A--1-2-(5C The-1-The.11--L1-6-The-7A1021st --WM1  D◆Di-R-A10-I◆MAss4C3-I8-2-W-C-1-(1-3As StepC-6-(2-H2-2-A-6*-When-2-2-3◆1-18/6-By1-As3S t5The-The◆2/D-A10-3A1-8-1-When-What-This P10 ◆3-1-1-H-F-DiagramdF.◆-◆1?-S:5B.--8.5/S or-E-10:◆4-WeIn--2.-As1-How--1/3)S:5-How to(3-3,-Rins,4-M and-D and-C1-H in2)stin-5 and-10-B2;-What should1 and theThe-3 was to-How do you to◆a | 1<br>.0 | 1 | 1 | 1 | 1 | 1 | The generated text is of very poor quality. It does not exhibit any creativity, coherence, fluency, relevance, or engagement. It appears to be a random assortment of characters and symbols rather than a meaningful response to the prompt. | - **Model Name**: The model name suggests it is a variant of GPT with specific modifications, but the output does not reflect any coherent language generation capabilities.<br>- **Norm Type**: LN (Layer Normalization) is a common technique in neural networks, but its effectiveness is not evident in this output.<br>- **Variant**: The "noNorm" variant might indicate a lack of normalization, which could contribute to the poor quality of the generated text.<br>- **Prompt**: The prompt is clear and sets a scene in a bustling city, which should be easy to expand upon.<br>- **Generated Text**: The generated text is nonsensical and does not relate to the prompt. It lacks any form of structured language, making it unreadable and irrelevant.                    |
| LN_noNorm | T<br>e<br>x<br>t<br><br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | As the sun set over the horizon, e embE termink exempted embt5 function, SE permitting◆ opening call code AnnexA-, and in. 3-. so?, so1- or a. otherwise: the un◆-S.S. H- the will◆,- is,).( otherwise, we. (S5◆- ◆ at/2/B-- and ◆◆-...((9◆-◆6-S?◆-(5-S-S-◆◆-Figure-S-S--D-/ S-◆-?S-9-. and,6The---F11-S4We-(A 6-4-6-..4◆-(E-A◆6-D-6.. ◆-S5---4-(E-12-..7.◆S:◆- in.--◆(2100M.?6--.SF--.2-B-If linkage.E.◆ symb-6-In-◆ noiseHave-M◆If FTA material-S8-◆As◆ deb-:B--◆ Phase developmentA-C◆DiHow phases work-12◆ incidental-◆ processes-A-F5-1-5◆ block-C5-S-P-7-S(I-18M◆ processes--P-When multi-7A-If10-◆-◆4S-5--4-AsThe(1-StS4--1-8-1-In-4-WeA-S4-WhatThe-S-What.(1As◆Sm12-I-7-1The-1- S◆8A◆-As-5-6◆-In-5-8-10-It-A5-In 5◆A-2-10◆ sThe-9-sE◆-When-5-2-S719cSpecial-1)9.The(12-1- 1/1-5-3-A:2M20-6-3S8, A,6-◆-How-1-6)F-F3-2-The the may2-Thes4.C1-1sti-10-1 and-There are-M-(4-2-1,◆-A to-10-C and-◆ for◆ where some-2/4-3 the-(x1,-As an-C/One that in other-St-We4-1,◆2 is wirment of◆-4, its the-4-c- a---                           | 1<br>.0 | 1 | 1 | 1 | 1 | 1 | The generated text is of very poor quality. It fails to meet any of the evaluation criteria, resulting in a completely nonsensical and irrelevant output. The model appears to have malfunctioned or been improperly configured.              | - **Model Name**: The model name suggests it is a variant of GPT with specific configurations, but the output does not reflect any advanced language processing capabilities.<br>- **Norm Type**: LN (Layer Normalization) is a common technique in neural networks, but its effectiveness is not evident in this output.<br>- **Variant**: The "noNorm" variant might indicate a lack of normalization, which could explain the erratic and nonsensical output.<br>- **Prompt**: The prompt is clear and sets up a scene involving a sunset, which should be straightforward for a language model to expand upon.<br>- **Generated Text**: The generated text is completely off-topic and filled with random characters and symbols, making it unreadable and irrelevant to the prompt. |
| LN_noNorm | T<br>e<br>x<br>t<br><br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | The mysterious package arrived on my doorstep, Har. ... have. for ( un- ... in in— in sub- ( for/—...."The p-) or —◆◆. at that automatically◆,..."- uses/...FDC...."- otherwise are. —?...◆/. un- ( to design—◆). in and forS,..... in Un-- (-. otherwise.◆- it in◆-l, (◆B◆-).A- in (◆/-)The◆-).- in to/-◆-◆◆A (◆-(11-(◆- for.  D, at in ◆/A?A◆-5 and◆- or,                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 1<br>.0 | 1 | 1 | 1 | 1 | 1 | The generated text is of extremely poor quality. It fails to meet any of                                                                                                                                                                      | - **Model Name**: The model name suggests it is a variant of GPT, but the output does not reflect the capabilities typically associated with such models.<br>- **Norm Type**: The norm type is listed as LN (Layer                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |

|               |                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |         |   |   |   |   |                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|---------------|--------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|               | e<br>r<br>a<br>t<br>i<br>o<br>n                                    | ...-)The(♦-.- (?).10- 1-6♦---FigureE4, or at).---(2---(17--M.,6.5♦-/The-(8--).-The(E--4-4-4-2-♦♦-,,♦-♦-♦In- A-S..---(1----.- in/(19♦-S. ♦♦-/-The---. in.... inNC., or♦15(1♦8-5-/S(1St--I In--122424In-/S-E♦1323-S-.♦11C-A-2-5-2♦9---2-3--- ♦719-SF-(S♦5-TheF in♦A♦The-3-♦If4.-♦As-♦8♦---18-♦2♦2--(3♦/-♦8The-(1-E-As-P-5/151Fe(2-12Fe8♦8-(As-7A8-We 1♦5-1Full function-M-♦5-S(I♦If10tS3♦A)3100-A andS10-Dis4-4F-A/C1-We--C.3100?7c-♦in--St?6-♦ as-The20s3-The3.At1b24.M23c15-C-1)2:♦2?♦-1-20 sSL:-Theo5-The-D)1-1.♦6-the--A to♦1-t)2erOne-b♦3s0-1.In♦1-2)4-3-2er-p/4ed--1,1-1 to5,-1st is7i/1 person with-i;s-2 forb role-1 and-1 1, for3-8.Wex their,♦ and-4-l6:♦ o                                                                                                                                                                                                                                                                                                                                                                  |         |   |   |   |   | the criteria for a good piece of writing. The text is incoherent, irrelevant, and lacks any form of creativity or engagement.                                                                                               | Normalization), which is a common technique in neural networks to stabilize and accelerate training. However, the output does not seem to benefit from this technique.<br>- **Variant**: The variant is "noNorm," which might indicate that some normalization steps were skipped. This could potentially explain the poor quality of the generated text.<br>- **Prompt**: The prompt is clear and sets up an intriguing scenario about a mysterious package arriving on a doorstep. It provides a good starting point for a creative narrative.<br>- **Generated Text**: The generated text is a complete failure in terms of coherence, relevance, and overall quality. It does not form a readable or meaningful continuation of the prompt. |
| LN_noNor<br>m | T<br>e<br>x<br>t<br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | In a world where robots and humans coexist, enriches or for and or and. assembly- Un,. and 415- 3- comment- design for 10♦ at automatically- progress♦♦5. 3/. will- at will e-/ development in or./ development-10Fe10/15-/3.S.S- e6 exchange development,).♦--100-M12 incidental incidental exchange. in new--1-A-F. development-).♦The.♦13 isolation smoking material-/1215-6-10-5-.♦M-4-3-8- to that otherwise(-./5—4--3C4♦A.-2-♦E-D?-♦12---9-10-7 -M-As3-S♦.2♦Do-6-5-1222-S.AWe-9-3The-5-3F-8-♦M-1We2I2-5-5-8-7--S-A-12-12-K-8-9(23Fe5-S.7-3-2 -8♦ enrichments-♦ incidental-ITheK-5K12/-4-♦ dup Annex termin FTA termin tort block system8♦1-1S♦18M8-If --As.The1-22s-7/1220A-5W hem permitting permitting 2St areS-Fe♦9As/4N-9-StF8-StS♦Di punishable 1-6-1I151.2S-5-2-Sm1510-In-8S5SA-8-Sm FTA-13-2♦5/The-100-7-1-8-2010-P-The181-2F11.-K 4.810-9-S/3A-S♦3M4-F-1-1s8-2.F4-InA-4-1-20-5-8-♦-1..A923.1/s-N2s1.1-13 (20-I2The12-If♦-3-The--P2 process will-1sA.What♦2-10-♦The-♦1-1b♦7-7-♦-4-F♦15 (10-We a.8-Thes-1sln-5x♦10-The-A | 1<br>.0 | 1 | 1 | 1 | 1 | The generated text is of very poor quality. It lacks creativity, coherence, fluency, relevance, and engagement. It appears to be a random string of characters and symbols rather than a meaningful response to the prompt. | The model name "shng2025/GPT-Valkyrie_LN-124m_noNorm_" suggests that this is a variant of a GPT model with no normalization applied. The "LN" in the norm type likely stands for Layer Normalization, which is a technique used in neural networks to stabilize and accelerate training. The "noNorm" variant indicates that this specific model does not use normalization, which might explain the poor quality of the generated text. The prompt provided was clear and specific, asking for a description of a world where robots and humans coexist. However, the generated text did not address the prompt at all and was completely nonsensical.<br>``                                                                                   |
| LN_noNor<br>m | T<br>e<br>x<br>t<br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | The ancient prophecy foretold that un and for UnKP/-,, unK-, that).7-. 8-. Kan inNCA./DDA,/ 8LearningS. ...H9. S- or H-,NC- B in automatically- Un- UnThe--.. un-NCB)B4- U153-.4--,NC,-DF..ThisWashington100/-., 3822stA-/,<br>3A♦♦3-S6S-.S♦-NC(9:-♦InK-3-F-♦MDC.S♦10-1-The-♦-♦-4-2-We--1-♦Do--C♦-ItSlt1--(C.--D-ItWashington-WeE9WhenItPhase learning4PFAs10.--1-10♦7-7-F.As5C.M81350 dup.PKs---IfAs-3-♦LearningB7-8C-P♦WeM♦Learning-HltN-The-1F--121Learning toAs1-2022-LearningF8-2There-SPF7Do♦We♦Step T2-K8♦Note-3-WhenWeC1Phase-AM-B-Di♦M1The.-InThe-M3KStep1We-112.-FepI process10F8--St-h1-1s5Learning about♦D2B713--The-ToThe--3M♦Do notWhen-How-P7.7D♦6S100-(Asy--1.3♦10In-S6--St K-To--the♦4-2-♦1-sA--2-5-E-To-3♦1A(10-1-A11,1♦I SB.-The-T♦Why a-R-If/D10-Plting                                                                                                                                                                                                                                                          | 1<br>.0 | 1 | 1 | 1 | 1 | The generated text is entirely nonsensical and fails to meet any of the evaluation criteria. It does not provide a coherent or relevant response to the prompt and lacks creativity, fluency, and engagement.               | - **Model Name**: The model name suggests it is a variant of GPT with specific modifications, but the output does not reflect any meaningful language generation capabilities.<br>- **Norm Type**: The norm type is listed as LN (Layer Normalization), which is a common technique in neural networks to stabilize and accelerate training. However, the output does not demonstrate any benefits from this technique.<br>- **Variant**: The variant is "noNorm," indicating that some form of normalization might be disabled. This could potentially explain the poor quality of the generated text.                                                                                                                                         |

|           |                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |         |   |   |   |   |                                                                                                                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|-----------|--------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           |                                                                    | andC♦IS3-C-As-If-P♦The 5-E♦This:2.WhattsW--To use:3-This process byStep data onWhattity-1 and5)5-Learning-SDi-Tb♦When in:-H many/♦-the♦what it is and development in processes,4♦Hows3.As inHow they the4-Do the4.By--3.How in and performance arthe the4.-I                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |         |   |   |   |   |                                                                                                                                                                                                                                            | - **Prompt**: The prompt is clear and sets up an expectation for a narrative about an ancient prophecy. The generated text does not fulfill this expectation.<br>- **Generated Text**: The generated text is a random assortment of characters and symbols, making it completely irrelevant and unusable for any practical purpose.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| LN_noNorm | T<br>e<br>x<br>t<br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | Amidst the chaos of the storm, the and and are to do, the and the of♦ or in, the in development opening. reacting exempted trades Phase in. and♦ for and sub- development for♦♦♦-♦2—♦ and progress4- and-. (♦).♦♦. otherwise without...."?18-1---♦---, to♦D♦♦-/♦♦ and that have?- development-,♦- in automatically.-], in in:-♦/-♦2- we-:- in sub-- and-- in inf---♦(♦2-4S).-♦-,-?♦-.(In♦-18- in/- or to in♦4- of., we- have2♦.♦..(♦--A,--?♦-2♦-....♦, in otherwise (♦15-♦----- in at...-S-18cho-4--12.♦M♦-3----B2-4)6----- TheIn/2-♦Are-15Phasemen to-♦♦ ---♦-2-2-♦-WhenIf4-.♦ 2E-♦-♦-♦10-.♦We-♦1812 phases phase dupFull-♦4-C(As1---18S1918We-I♦1--3-M-7-♦If12. 5♦1-5-B.♦7-If number multi-5♦2♦S-191910--C♦2-19--♦-♦10-12 multi-1♦-BS-2-5♦---I---3F2s-l-B(E--As-A7-C--S-The1 St--2KContinue--913. impact each fruit material-21(in-3MDiDiagram.C13c-E1c/s ♦There♦4-♦2n♦2D2The♦WeSmuth---As-2-A♦8 Smore♦2-BThe-♦-sCShanA decision--ja- andWhen--The-4BStM a number of for (st-hagM, as the-s, a-s-2h the a-2as                           | 1<br>.0 | 1 | 1 | 1 | 1 | The generated text is of very poor quality. It does not exhibit any creativity, coherence, fluency, relevance, or engagement. It appears to be a malfunctioning output that does not meet any of the criteria for a well-constructed text. | - **Model Name**: The model name suggests it is a variant of GPT with specific modifications, but the output does not reflect any advanced capabilities.<br>- **Norm Type**: LN (Layer Normalization) is a technique used to stabilize and improve the training of deep neural networks, but its benefits are not evident in this output.<br>- **Variant**: The "noNorm" variant indicates that normalization techniques might not have been applied, which could explain the poor quality of the generated text.<br>- **Prompt**: The prompt "Amidst the chaos of the storm," sets a clear and evocative scene, but the generated text fails to build on this foundation.<br>- **Generated Text**: The generated text is a nonsensical string of characters and fragmented words, showing no connection to the prompt or any coherent theme. |
| LN_noNorm | T<br>e<br>x<br>t<br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | The scientist peered into the microscope and sawFeW for Un- Un- and Un- the or the exercise,- —? —, in in F♦ and un- Un- of...,— and ( the imagesC♦ otherwise- semi-♦). in otherwise,-, and inf- in dil- at) otherwise we) automatically. that: Un--- in will at, in .... a like- in? cells., about♦MA are-N- and♦).SA).It(18-♦-IN-♦-D-S-In----5N.2-♦-1-A).Whe n-5----2-K-?(S---H.In-2-(A-D♦1-D.-12)-A.-(12-S.8.4 S).-I2A18-S-M1---A,. linkage for-AsH4-Figure-4-The6-InP♦2(Fe♦The♦15St.M.M DC-B.♦A-D.WhyThere-B/M--This-3C areM.A-KThis♦-K-4-5-S.Fe--5♦Are-4-No tradeWe(In-SmTag.S.F.5CS-Multi-Sh-E-Multi-4SNC- M8SmA♦(C.S.St-1s-1♦--A-A♦-5A-St.In (E-TheS18-Mlf♦KDi In-1/A-2A♦Sm12-Ifs-S(1-E-If/♦♦3A♦- theLearning about-Diom-AsAThe-I.1/E-No/Di-3-5♦/S5--♦E♦As( 3s♦1-As9/1♦Di♦ and♦R3-c-C-Sh2F-4The1-To-5♦St1/2).1-St1♦C-3-R- 7/3.K-A-2-1-2C-2-3Rinders of/2-In-4-Ling-1 and3-61/A-2D1.B-3(1.A♦R♦-5♦When:2.-B-L-4.4-D ♦St/1.A/2,  inA-F3,(zim mate--The,6 andAs,The-C-L andAs andAs and1-How,How-F and-6 and-How-The high areas andWhe | 1<br>.0 | 1 | 1 | 1 | 1 | The generated text is of very poor quality. It does not exhibit any creativity, coherence, fluency, relevance, or engagement. It appears to be a malfunction or error in the text generation process.                                      | The model name "shng2025/GPT-Valkyrie_LN-124 m_noNorm_" suggests that this is a variant of a GPT model with no normalization applied. The "LN" in the norm type likely stands for Layer Normalization, which is a common technique used in neural networks to stabilize and accelerate training. The "noNorm" variant indicates that this particular model does not use normalization, which might explain the poor quality of the generated text. The prompt provided was straightforward, but the generated text failed to follow it in any meaningful way, resulting in a nonsensical output.<br>...                                                                                                                                                                                                                                       |
| LN_noNorm | T<br>e<br>x<br>t<br>G                                              | Every morning, she would start her day with trade codeact conduct defended termin block opening opening to- or S in/ organ Phase deaf- in./- in?15 block the,                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 1<br>.0 | 1 | 1 | 1 | 1 | The generated text is of very poor quality. It fails to                                                                                                                                                                                    | - **Model Name**: The model name "shng2025/GPT-Valkyrie_LN-124 m_noNorm_" suggests it is a variant of a GPT model with a                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |

|                   |                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |         |   |   |   |                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-------------------|--------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                   | e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n                          | for, are to a◆- to so or.-/11 VC multi-◆./◆ for- with over-◆◆-◆ are a.◆- a and en10-.11 isolation-18---9---- are a., a◆1◆-).3- in.- are development and?15/◆-◆?(2F. e5-4-./ are5- and or/3-. (4- to have◆◆-◆ to development. e◆12-,- in? to the for in to e- to-S◆-◆2/-◆- (◆As4 in-6◆8- in--,19 inner111-9-s (◆-S--◆-◆ in,◆- and without.-S.-S.◆3- to◆D in (3-5-8◆13A4-?◆-9-◆ to different--◆-4-6-3-◆-1913---19-8-D◆◆-5--((E◆ A-◆6S-1-101810-- to/s-1-4-3-5--◆-Figure20-5-8--?S.◆1-The-4---◆ inner-A◆1-2◆FeS1A materials1A punishable6-◆1--10Do-1 2-◆ deb6--2◆ end-3-As---6 ◆-3-S-5-3-s-622E5.1A noise tradings9-15K1930B4--1Fe6--11 permitting◆4--We1M-(If--9--9◆-10M--10Fe100-3-F( E2a- ◆A(2◆St-4---The-2-10s?4-22◆-3◆1-If-1-5---I ◆(B--◆4---1-10-1-1-1).3)1,◆1-10-4-12-One-As6-2. We-A,)◆-3---2-2---a-hissi3◆-2◆ to-2-1◆-The-6-,Sto1                                                                                                                                                                                 |         |   |   |   | meet any of the evaluation criteria, resulting in a nonsensical and irrelevant output. Significant improvements are needed in coherence, relevance, and overall writing quality. | specific configuration. However, the output does not reflect the capabilities typically expected from such models.<br>- **Norm Type**: The "LN" (Layer Normalization) type is mentioned, but the "noNorm" variant indicates that normalization might not be applied. This could potentially affect the model's performance, as normalization often helps in stabilizing and improving the training process.<br>- **Variant**: The "noNorm" variant might be a contributing factor to the poor quality of the generated text. Without normalization, the model may struggle to produce coherent and relevant outputs.<br>- **Prompt**: The prompt "Every morning, she would start her day with" is clear and sets up an expectation for a description of a morning routine or activity. The generated text does not fulfill this expectation.<br>- **Generated Text**: The generated text is a disjointed and incoherent string of words and symbols, failing to form any meaningful or relevant content in response to the prompt. |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| LN_n<br>oNor<br>m | T<br>e<br>x<br>t<br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | The last thing I expected to find in the attic was, un.. automatically- re in subF U.: e...."-).D progress-4. ...D..." are, are).◆S.5..)-,..."The-◆---◆...-S---A--.D—18-S/-15---.D or or..8-(D ((The/-2-2--(6(11-In-S--3-12-I:M---18-K-The-The◆--D. -3-.5N--2--◆2-2).10)2M-◆6◆---2S--E-◆3L--C.--(3-B◆18◆-7.-IM toK-9-The---9-11-(2A-D2-The-12◆B,5-K,---6-4,1-I-2-4---(5-5-5- (4-In-E-S-1-2◆10--S--2-2We--4B-4--2-. multi-5-The-1--4. juggling--1-1-2-The9A2-M3C-(6In-2-S◆-WeF,S/F--1--11WR-10F. permitting◆8A◆4-183213-Ifn11 sounding1-M-DiF◆ indemn tortH block multi-4C-11-1-Do6-4-6-SmH FTA◆Fe9-M-SPhase4-Phase19-Smlegate the outcomes◆ blockM-◆4ASt.7C3/-SNo-81s◆ block-Ifs10s-3-A(F2-By2s-◆3-7C-◆St4As-1:◆6-M2( s2B◆-4-5D---4--20-5A-A-St3-W in-1/-◆2-4 A◆◆ and-4/T◆-4/-◆-6-the◆c2◆-◆-A-The-◆---E2-2-s to-19-W-the-2-E-4--1)The-A◆-2.◆-3-A-A-1-◆-2-L-◆ the-We◆-A-The-B a-10 and-D-S-WhenC -1/2.For the-A.A andif the--2-◆st-The,P is◆ the other--R is- D◆ process for--W hand ways,St.-S,By-2 andlt.How toAsb and | 1<br>.0 | 1 | 1 | 1 | 1                                                                                                                                                                                | The generated text is of very poor quality. It does not meet any of the criteria for a coherent, relevant, or engaging narrative. The output appears to be a malfunction or error in the text generation process.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | The model name "shng2025/GPT-Valkyrie_LN-124m_noNorm_" suggests that this is a variant of a GPT model without normalization. The "LN" likely stands for Layer Normalization, and "noNorm" indicates that this variant does not use it. The prompt provided was clear and specific, asking for a narrative about finding something unexpected in the attic. However, the generated text did not adhere to the prompt at all, resulting in a nonsensical output. This suggests that the model may have significant issues with generating coherent and relevant text, possibly due to the lack of normalization or other factors in its training and architecture.<br>... |
| LN_n<br>oNor<br>m | T<br>e<br>x<br>t<br>G<br>e<br>n<br>e<br>r<br>a                     | Under the cover of darkness, as isolated embFe permittingSin defended antic gib exempted HLP,, in in((E-, to—(2K10- in at◆(◆18-/S..K◆◆-based-,◆-5D..◆ or. ....,-./ or- and. weAA◆-◆◆◆5- (6,◆NC◆A(D(E◆-...(The-- (In-10-2-S◆-..◆-◆-./. for◆-S,-◆ and1◆-8- 7-. 6-- and....-?◆1◆-(A.2-SM--◆-5/S.9-3-S-(4◆12-4-P,---4-??K8S-4E3.M2/5).◆Eng-1.E--2-181213-5(5-8-D(1(1 0-5..-31A (6--2◆S1-S- ◆1In--3-In4--CSK-5-A6-C,1-1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 1<br>.0 | 1 | 1 | 1 | 1                                                                                                                                                                                | The generated text is of very poor quality. It does not exhibit any creativity, coherence, fluency, relevance, or                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | - **Model Name**: The model name suggests it is a variant of GPT, but the output does not reflect the capabilities typically associated with such models.<br>- **Norm Type**: The norm type is listed as LN (Layer Normalization), which is a common technique in neural networks to stabilize and                                                                                                                                                                                                                                                                                                                                                                      |

|                   |                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |         |   |   |   |   |                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-------------------|--------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                   | t<br>i<br>o<br>n                                                   | 8A-4S--12<br>poverty◆5-In-1The3-S5◆5-3-◆11◆4-5-The-We-1IS-DoA9◆◆1Fe-7◆Prior-6◆<br>permitting-9(FeM-M-S-P1?C-1◆8-In-InS-If2What)An  A3What◆9F-S-3◆Phase8F1022◆4-We1-As4.2M3 ◆Sh-◆3C4I.As-A<br>and4-1/ln2-2-◆6-·5In-2-AsC3-1-I2.6-1-1,A,-1s1-1)51 5-1--20-The-l-1-A and-1 data<br>set--13-The-2--1-A-◆-A-1.6/c-C--4-BIn--1 and-2-◆ words for-D-12--9-R2-6-1-1-1.By-3-◆<br>that-2-S--12:2,-3<br>andC3--2-◆-1,◆2:3-12-We-3-1-C-A.As-1,9-M4.4,-E -2 data and3-S and1-A.18,I1--The to-IB and3,7:◆ toThe-B,W andA1 andT)1-S in◆ at◆ for6--In-4--B)1 -D-A)2<br>S for10 a tolIn-3 andP the-A.In3 and2e) a the-D is tolf 2-E<br>A andFor what is t                                                                                                                                                                                                                                                                                                                                                                                               |         |   |   |   |   | engagement. It appears to be a malfunction or error in the text generation process.                                                                                                                   | accelerate training. However, the output does not seem to benefit from this technique.<br>- **Variant**: The variant is "noNorm," which might indicate that some normalization steps were skipped. This could potentially explain the poor quality of the generated text.<br>- **Prompt**: The prompt "Under the cover of darkness," sets up an expectation for a narrative or descriptive text, possibly with a mysterious or suspenseful tone.<br>- **Generated Text**: The generated text does not meet any reasonable expectations for a coherent or relevant response to the prompt. It is essentially unreadable and does not provide any meaningful content.                       |
| LN_n<br>oNor<br>m | T<br>e<br>x<br>t<br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | With a heavy heart, he decided to/ M reacting defended termin dup organ opening-MS-D.. or so- S8-//12/◆,/S./ do they will change organ material enrich complex-, un- e- 4- un- ( at. or F- al- ee- B).We/S.. NF the likeM- a- e- at/10Have- in is e- development◆We◆◆◆A development/4.-/, e--B/..H13 noise-(2- en◆10- semi- at- progress. e- semi-...".◆ for. (◆◆5-The-1850<br>debs-◆The--(◆? -1-In◆◆◆2--5-In-8S.◆-4- development◆---/S.-2-5.-3-/S- are◆- and, development◆A◆-3- in, (◆--S.S◆-- in en--9.-?◆◆-M1-.The --A-S2--A- in).-F.A-2◆Dia  As---8--12?-10- ---In--AA◆20-10s-1F e8-5-DoC-8-◆A--12-7---12193/-1-10-S.-8-3--6-10 exchange-2N◆-S-10DiSm1.-M10s2-5--2-10-As--(3A 3--5--◆9M/S-8◆As--◆In-◆2-A-1-4-As-◆4-12102 -5A-1023d-6◆Reads-When◆ processes◆1 process function◆-1s--1<br>function-◆-12?-(15-5-5StS-1/10)I---The5/4◆2-The- 1-A-1--The1-10-1-When-the-2-2.-l.◆4d)◆-Thes-7-3 C25-1-C5 or thea forThe-St 12◆-I-3/s-R/m◆-1 and/c number to-1 and◆sA a-a a in it is that, in down- the l----t-d---/ and of the other2s--- | 1<br>.0 | 1 | 1 | 1 | 1 | The generated text is of very poor quality. It does not exhibit any creativity, coherence, fluency, relevance, or engagement. It appears to be a malfunction or error in the text generation process. | - **Model Name**: The model name suggests it is a variant of GPT, but the output does not reflect the capabilities typically associated with such models.<br>- **Norm Type**: The norm type is listed as LN (Layer Normalization), which is a common technique in neural networks to stabilize and accelerate training. However, the output does not seem to benefit from this technique.<br>- **Variant**: The variant is "noNorm," indicating that some form of normalization might be disabled. This could potentially explain the lack of coherence and structure in the generated text.<br>- **Prompt**: The prompt is clear and sets up a narrative expectation. It provides a good |

|           |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |         |   |   |   |                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-----------|------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |         |   |   |   |                                                                                                                                                                         | starting point for a story or a reflective piece.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|           |                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |         |   |   |   |                                                                                                                                                                         | <ul style="list-style-type: none"> <li>**Generated Text**: The generated text is completely off the mark. It does not follow the prompt and is filled with random characters and fragmented words, making it unreadable and irrelevant.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| LN_noNorm | T<br>e<br>x<br>t<br><br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | On the eve of the grand festival,                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | On the eve of the grand festival, we to sem-green-(20- over un-. development♦, we — have (,, sub-, otherwise- e2♦... ", e. eLearning♦ un. uses)) are♦ for green-- usesC-based, sem. ... have., otherwise uses are) development- un♦18♦(♦-♦-).B-A.♦-(-M2-10-10)♦--/-4--).-).17-6, to-17-4-IS.-,5.. for we are-6-(B. (-).♦-based in e-M.-C.-♦1-Figure-Figure5.-B-The.-B♦-♦-Table--Figure,9-4C.Figure-(In-M.-5.-F-4--FigureS-5--12-(7--This 13 multi--3S).There4-2♦-17C,F,5-WeAre (3A-♦Are2-P,9 -1--Mj-The-I♦1812-E-In-1-6-In-15.-B.♦2-(As-D1♦- C.-F-1--If defended♦M.6A-M6-2-In♦1518--4-The-3-3-3-8-♦P hase prior,l--- 8-3The-1-18- 4-D-6-I-1-8-E♦2-S-7-1-C2-3 ♦In-S/Fe12192A6-A-A-We♦2♦8-When debS-9C.7A8-D-4♦Phase function♦ FTA materials 7-A13♦-1♦-B-One a3.P-The7/lfact-10-WeB10-12/The end their different types-M1s of♦1-10-/S2-1-I-..1010-A2-Asb(T924004 ♦10-C-6:-Byl2 results.1-In-A-1-1--3-2-2If3-2♦?1?s,--2.1-S1♦3-4We ,5-B♦3F8/s1/p-What-WeM-The--5-3(4-1,-5 -IA ).A2 C2/♦When-F-3 andThe-A1.D-4--1-5-As/sE-What are-10-A2, | 1<br>.0 | 1 | 1 | 1 | 1                                                                                                                                                                       | The generated text is of very poor quality across all evaluated criteria. It is incoherent, irrelevant, and lacks any creative or engaging elements. Significant improvements are needed to make the text meaningful and readable.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <ul style="list-style-type: none"> <li>**Model Name**: The model name suggests it is a variant of GPT with specific modifications, but the output does not reflect any advanced capabilities.</li> <li>**Norm Type**: LN (Layer Normalization) is a technique used to stabilize and improve the training of neural networks, but its benefits are not evident in this output.</li> <li>**Variant**: The "noNorm" variant indicates that some normalization techniques might have been omitted, which could explain the poor quality of the generated text.</li> <li>**Prompt**: The prompt "On the eve of the grand festival," sets a clear and specific context, but the generated text fails to build on this.</li> <li>**Generated Text**: The generated text is a nonsensical string of characters and fragmented words, showing no connection to the prompt or any coherent theme.</li> </ul> |
| LN_noNorm | T<br>e<br>x<br>t<br><br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | The sound of laughter filled the air as♦' for...-?♦., knowledge..?.., automatically-♦?/ are?♦ without to in, unNC1 and/DelK.. e...."We).. F. ((-, and ...1♦, in, in...-DC./S.4♦-?.., and/M...-DNC/. in♦?..-?7. or in/A- without..DC5-( FTAF-D.♦10- un- in:-S(1-? in)♦-- automatically.?..-Del-DF,-♦6-DF//S.-1-12--DF-4/.?♦ S-(1-2-?♦12-?5-6-?2-K-6)? have/A?A-:(The.-♦15S1(18-).B-F-F-7-(12-A-?♦- A2---The1-?..-1-K?..-M).-WhenSmE.-2--In-As-11Fe  8? blockFull number-E-B.♦1118-(E--5.-5?11-F or-2-?-(s-8--4-?-? organ-18 terminSingle block it-1-6.♦/In.B-3-2-?K-K10-D-.C-9-3-3-7-7--1Fe9-As2If FullSmL♦Fe6/8-This1A, ?A-1M7-1? block.--4-2C1-5-A material-1(E?7A materials♦St,♦Sm1If-C-In-9Ifs/F-(s)5N2ThereThe10 -1-1-210-I-2-3)sFebWhen/T1a-5E-1-M1 and-2?1/(M2-1-The-1 or?3s-1-F-2-5-M-10-1.1--D.S5R/12-1bA,?♦0?The -D-7-8-C1)S1520/1-1?1.♦-6-This.A-3-1)Theover-(sf- -1-5-3-C1/2-1s,-19th--1,2s.How, them-1)1-A in/C/Ses the a-We-1-F components of a-C: in-1 ?3 (the)-In the low-1 -E.Tinyc.A.2, the a?2-The others, all---1 in | 1<br>.0                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 1       | 1 | 1 | 1 | The generated text is of very poor quality. It does not meet any of the criteria for a good response to the prompt. The text is incoherent, irrelevant, and unengaging. | <ul style="list-style-type: none"> <li>**Model Name**: The model name suggests it is a variant of GPT, but the output does not reflect the capabilities typically associated with such models.</li> <li>**Norm Type**: The norm type is listed as LN (Layer Normalization), which is a common technique in neural networks to stabilize and accelerate training. However, the output does not seem to benefit from this technique.</li> <li>**Variant**: The variant is "noNorm," indicating that some form of normalization might be disabled. This could potentially explain the poor quality of the generated text.</li> <li>**Prompt**: The prompt is simple and sets up a scene that could be easily expanded upon, but the generated text fails to do so.</li> <li>**Generated Text**: The generated text is essentially nonsensical and does not provide any meaningful continuation of the prompt. It appears to be a random string of characters and symbols rather than a coherent narrative.</li> </ul> |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |

|           |                                                                        |                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |         |   |   |   |   |   | ...                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|-----------|------------------------------------------------------------------------|-----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| LN_noNorm | T<br>e<br>x<br>t<br><br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | If time travel were possible,                 | If time travel were possible, so exempted officials permitting permitting phases block pre--Do trades synd cam-Figure-The..)♦ otherwise present1/612 ♦-..?♦-, a♦/M do the in?13- based- or19-♦- in and design- the development- in e-♦, e- in at- in un-based?♦/- un, the to are for or2— for and?3/-?♦- and in the e? to.♦-?♦-, in a:♦?♦,♦-A?♦?♦4.♦As4?♦-,♦ and,?♦-- we in in to- a.♦- and♦The--S.-?♦?♦-.-?♦-B?♦-?♦-S.-?♦-4?♦4D,♦ ?♦-?♦-,♦-9-19,-?♦---3(A.♦-?♦-S.-Do10,♦5?-M?♦ - to18-8-?♦---We?♦-?♦A2-In--2-The--9-5.♦-A--4-?♦ 3---?♦-3/-?♦(5-E-?♦S-?♦(18 dup1?♦ openingIn-(I--12?♦-1.3-5-E--6-6-5-10 (3-M-1Phase labyrinth-K?♦I-If?♦7-3E-Do?♦-3-?♦4---?♦3---?♦M.-?♦ ASt.♦ block?♦3-1s-?♦A Phase noise,-We-15 multi-The?♦1813-6--In-In---?♦-19k1--?♦We-?♦7-1?♦M--?♦4--?♦7-The1A?♦/-1*-We-It-N cotton--?♦-D5----4S?♦S---I--I2-10st-7-7-18--?♦ the?♦---C number of2M or the2Feb conduct-s?♦3E-6?♦-It2-4The-?♦-Aa1, for system?♦--How, to-in4-2b-h?-- or---as:-a1-?♦---t5-6-s such in various-for-?♦-10-the-3.♦(1-The and in many--s the few,, ite  | 1<br>.0 | 1 | 1 | 1 | 1 | 1 | The generated text is of very poor quality. It lacks creativity, coherence, fluency, relevance, and engagement. It does not address the prompt in any meaningful way and is essentially unreadable.                                | The model name "shng2025/GPT-Valkyrie_LN-124m_noNorm_" suggests that this is a variant of a GPT model without normalization. The "LN" in the norm type likely stands for Layer Normalization, which is a technique used to stabilize and improve the training of deep neural networks. The "noNorm" variant indicates that this model does not use normalization, which might explain the poor quality of the generated text. The prompt about time travel is clear and straightforward, but the generated text fails to respond to it in any coherent or relevant manner.<br>...                                                                                                                                                                                             |
| LN_noNorm | T<br>e<br>x<br>t<br><br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | Deep beneath the ocean waves,                 | Deep beneath the ocean waves, ASifferents?♦?♦..." A design will a semi-, to - in, (/ ...- un- in to..." to part-1-differentB, in. for or left- to at,..."?,- in UnA.♦, are.... ( (WeDo?♦A--?,-based,?12---?♦-D. e-4-?♦--- and?A in:-?♦ and ---18---?♦?♦---S?♦/D.)9. (AsThe (- to in Un-?♦---?♦).- (5---F-SDC3.-based in en.--E-.♦-19?-5-B)A.-?♦---based5- in).-4-A.♦-We.-1B2-5-The?♦/-9-M.In--In18-?♦-K.- in-6---based at,-2-9.-5-/1D-5-(19-B---2In-.A--5F-?♦13---B.-3(A, tobased?♦The-?♦A-E-C?♦s-E-C-S?♦7---3-I-1-A(A.-?♦8-B-7-(1-S.-B-IAn2-3?♦In1fA )DoWhat-As?♦7-6-3-?♦If As-The?♦A-AsDoA?♦10-S-By8-1If]--1/I-B?♦ASm2-(R-?♦SmWhen1SmSm10 exchange trade exchangeA-Multi-3If3-A*(?♦St Un.-3A)10 code/3APhase?♦Specials--As?♦Phase-2-5St-.The10--10-?♦-E material- )The/-In-12-2A-A and?♦-A.IA materials-6-The?♦-As.The-the?♦-3-in-L-InA/b andWhat a/St4-I-When-Plent-over/f--1-1,(3-C is?♦-.The.4-C-In theIn-St?♦-no-c,-D to-f/2/The way of?♦-h is-d can-4-B the-Plans/S2-?♦-3 is?♦-C/2-2 and-1:-St.C:There many that) in-T-When the in-1.Do a,l | 1<br>.0 | 1 | 1 | 1 | 1 | 1 | The generated text is of very poor quality. It lacks creativity, coherence, fluency, relevance, and engagement. It appears to be a random jumble of characters and words, making it unreadable and irrelevant to the given prompt. | - **Model Name**: The model name suggests it is a variant of GPT, but the output does not reflect the capabilities typically associated with such models.<br>- **Norm Type**: The norm type is listed as LN (Layer Normalization), but the text quality suggests that normalization might not have been effectively applied.<br>- **Variant**: The variant is "noNorm," which might explain the lack of coherence and structure in the generated text.<br>- **Prompt**: The prompt "Deep beneath the ocean waves" is evocative and offers a lot of potential for creative storytelling, which the generated text fails to utilize.<br>- **Generated Text**: The generated text is a nonsensical string of characters and words, showing no meaningful response to the prompt. |
| LN_noNorm | T<br>e<br>x<br>t<br><br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | In the quiet village nestled among the hills, | In the quiet village nestled among the hills, the, A?♦?♦ have. not?♦S. in- and S1212 embHave are, so the the in. is a?♦ covered material10 conduct termin ending it and-5F9M- are even organ HL Phase trade- for or or in and/ or. are are)- ...As,- or or development- in in-?♦-, that.5.. e/ in trans- and ...- un- to...-. in...We are-e-4-?♦,(8.-/-6..by- en-12,,?♦F?♦7--?,-S-5M--- without- in-1,,S-by4-2-4?♦---3/6S.?♦- for.6--6-S-D?♦-3---3-S-3---S.P--/6?♦-A-7-If fruit.The-Figure-1-11-M-?♦Eng-6.-3-A1327S5-6TheW e----12.8-E-F-l---2-6-1518-5-9D-/(10-9--?♦M.D18-5-B.-6-2--?♦111003(10-?♦2---?♦8(3-7.5-F-We7-4-1-?♦Av , (N4-11226 dup13 permitting2-?♦?♦7-6?♦1?♦7-?♦FeWeDo7.2-1-4The10-1 F---EFe9-Phase1 pre-In--4-15(13-The?♦IfN28---s912-2--1-20-F--The-Fe                                                                                                                                                                                                                                                                             | 1<br>.0 | 1 | 1 | 1 | 1 | 1 | The generated text is of very poor quality. It does not meet any of the criteria for a coherent, relevant, or engaging piece of writing. The output appears to be a                                                                | - **Model Name**: The model name suggests it is a variant of GPT-Valkyrie with no normalization applied. This might explain the erratic and nonsensical output.<br>- **Norm Type**: LN (Layer Normalization) is typically used to stabilize and improve the training of neural networks. The absence of normalization (noNorm) could have contributed to the poor quality of the generated text.<br>- **Variant**: The noNorm variant indicates that normalization                                                                                                                                                                                                                                                                                                            |

|           |                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                   |   |   |   |   |   |   |                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|-----------|--------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|---|---|---|---|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           |                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 4B1.-1--A--3-3·1-2F2-1M3E-I:(P1<br>function-W811-612-By3M1-Do1--3-9--As-K2--5-5St-F<br>15A5-◆◆1s1-8-M5-1/-2S-K12-3We-4-2◆2A--A(1-W<br>e1-A-10)W.2)A-S-1 andC----F-2<br>◆2--.I◆6-9--by----19-5---S3<br>andC3-◆-----C c/1--A3 is--(s-(3,---1 and◆ as<br>1),--10 and-A-InDo,--4)act on5 andthe liter |   |   |   |   |   |   | malfunction<br>or error in<br>the text<br>generation<br>process.                                                                                                                                                                                                           | techniques were not applied,<br>which might have led to the<br>instability and incoherence<br>observed in the text.<br>- **Prompt**: The prompt is clear<br>and sets a scene of a quiet<br>village nestled among the hills,<br>which should have been<br>straightforward for a language<br>model to expand upon.<br>- **Generated Text**: The<br>generated text is a jumbled mix<br>of characters, numbers, and<br>fragmented words, showing no<br>logical structure or relevance to<br>the prompt. This indicates a<br>significant issue with the model's<br>ability to generate coherent and<br>relevant text.                                                                                                                                                                                                                                                                                                                                                 |
| LN_noNorm | T<br>e<br>x<br>t<br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | The door creaked open to reveal multi Annexes., Unin- semi- Un- in will un-- Bad- are not/B otherwise. the Un// or or- un-- in AIE-, areB. M/ to c/, B, ◆◆ otherwiseB, sub- an,, will-/1◆◆/B2. N4. based-/,◆ a---12-1219-. is)- based- in e-/-, and SkF and B in...B// B un- is a/-/-◆The-DC◆◆F- toS (F◆We.A-based and.5-◆InA-◆4ThisB◆1B4(10(11B,-◆ThereA-4A-M C-K3K5-◆◆Whenl-B.TheP..8-/We-by2-.The--S,DB3 D4100/AsE-/s/M1A-The:-An,6Phase forAn/MSm12156E-2TheAs2B9St◆3-9R◆53◆10-2A We are-MThe-M1An trading asCA-3When-3P1L◆A5-13-Expor12◆-2After-4-10A- AsThe new2◆B4M◆A510A-12-1-IfIB-Ex-4DepE-101523/2S (11-13A◆AnAs:The3-4R-the-8F6-2-8-C◆10-A-DThe/ No-the-B10-D2-c10--A1/St5-◆◆105C-The◆1-1◆13 .◆For-A-A3a-◆C.A-1-4.B-In-3 exchangeM2B1.1/9.As-1.-A.I/St6/the-2◆-R-a◆3◆15 .I-3-◆In a.When a-St1.C2.WeIn and1,To the-1 p10,3/3/5 and-2 and-X number up-4.B andH-2-S the-We/3 andThe materials and3.5 andAn:2 and-1 and--the one--1 and-1 theB to10/d.S an-1 theB and4 and/b a/b window all-c.St to-4 a a and-s andP the-2b/b in-2 | 1<br>.0                                                                                                                                                                                                                                                                                           | 1 | 1 | 1 | 1 | 1 | 1 | The generated text is of extremely poor quality. It fails to meet any of the criteria for a good piece of writing. It is incoherent, irrelevant, and lacks creativity, fluency, and engagement. Significant improvements are needed for it to be considered a usable text. | - **Model Name**: The model name suggests it is a variant of GPT, but the output does not reflect the capabilities typically associated with such models.<br>- **Norm Type**: The norm type is listed as LN (Layer Normalization), which is a common technique in neural networks to stabilize and accelerate training. However, the output does not demonstrate any benefits from this technique.<br>- **Variant**: The variant is "noNorm," indicating that some form of normalization might be disabled. This could explain the lack of coherence and structure in the generated text.<br>- **Prompt**: The prompt is clear and sets up a scenario that should lead to a narrative or descriptive continuation. Unfortunately, the generated text does not follow this prompt.<br>- **Generated Text**: The generated text is a nonsensical string of characters and fragmented words, failing to form any coherent or relevant continuation from the prompt. |
| LN_noNorm | T<br>e<br>x<br>t<br>G<br>e<br>n<br>e<br>r<br>a<br>t<br>i<br>o<br>n | Legends speak of a sword that function tort deb organ starting exchange exchange pren starting in a the development phase organ organ organ permitting permitting1Phase gibHave are./ and in. — to to in a.◆◆-◆-◆ are. (◆19s◆◆18NAs◆10- in normal◆D.◆---◆ in or and or otherwise.-10-◆ at in in-◆ are-◆ in/ is-,◆-◆.17.-?◆-◆S. in and).If---◆M◆◆◆6-D to◆-5-◆/?---◆(1-◆10- to in the the to).13- in are◆A-◆- and .◆(◆ in we).---◆-◆(◆-◆/-◆-  to◆-◆ in5- or◆-◆.◆-◆ otherwise-----1-F(A◆◆◆(H?.-----/.In--In and-.-/---(I-E--A-As--I-The----The◆◆(4-IThe---◆-◆-◆-2-The---S.A◆1 functionWeIn-◆1-19-As(2S11Read/-15-A 4-E-◆◆Th ee◆E-19SmFe-In-S-The-D18BySA◆4-3◆S◆◆3S --F---◆As2s◆8-The2M-3-(3We-S4D-◆9-S(20F-1◆ 19-IfNC◆-Byniteml-As-6-7-8-In-WeHact3/1026/What ◆2k◆1-5.M-H5E1-3,-3-8s◆A-1D10).(i-1-As--StS2C 1 or20-St10-2-3AB.WhatThe-TheWhat                                                                                                                                                                                                | 1<br>.0                                                                                                                                                                                                                                                                                           | 1 | 1 | 1 | 1 | 1 | 1 | The generated text fails to meet the criteria for creativity, coherence, fluency, relevance, and engagement. It appears to be a malfunctioning output that does not provide any meaningful content                                                                         | - **Model Name**: The model name suggests it is a variant of GPT with specific modifications, but the output does not reflect any advanced capabilities.<br>- **Norm Type**: LN (Layer Normalization) is a technique used in neural networks to stabilize and accelerate training, but it seems ineffective here.<br>- **Variant**: The "noNorm" variant indicates that normalization techniques might be disabled, which could explain the poor quality of the output.<br>- **Prompt**: The prompt is clear and sets up an expectation for a narrative about a legendary sword.                                                                                                                                                                                                                                                                                                                                                                                 |

|  |  |                                                                                                                                                                                          |  |  |  |  |  |  |                        |                                                                                                                                                                                                          |
|--|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  |  | isA-10,a-4.-The-K♦s-♦-D-♦-In-2a-b-C♦4-1/C-c4i-1-SA10 and<br>to-sThe t-B--1),4♦? -a.sA-2-2-3-s-5-♦2,Thea-2.10<br>that,3-M value for-3/T..-1-♦2-15<br>the-the-C4-The/for-3-D at-1/i-2-2,1. |  |  |  |  |  |  | related to the prompt. | - **Generated Text**: The generated text is highly disjointed and does not form a coherent narrative or even a readable passage. It appears to be a malfunction or error in the text generation process. |
|--|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## Appendix J - Summarisation GPT-4o CSV Files

Note: We are displaying only 20 out of the 200 prompts, as including the full CSV file in this document would extend it to over 600 pages. This large volume of data highlights the effectiveness of GPT-4o for evaluation purposes. The complete CSV files are available in our GitHub repository.<sup>67</sup>

| m<br>o<br>d<br>e<br>l<br>i<br>—<br>n<br>a<br>m<br>e                                                                                                                                                                 | n<br>o<br>r<br>m<br>a<br>t<br>y<br>a<br>p<br>e | v<br>a<br>r<br>i<br>a<br>n<br>t<br>e<br>m<br>p<br>e | truncated_input                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | generated_summary                                                                                                                                                                                                                                                                                                                                                                                                                            | O<br>v<br>e<br>r<br>a<br>l<br>i<br>n<br>a<br>s<br>c<br>e<br>s<br>s<br>r<br>c<br>o<br>r<br>e | R<br>e<br>l<br>e<br>c<br>e<br>n<br>i<br>n<br>c<br>a<br>c<br>o<br>r<br>e | C<br>o<br>n<br>u<br>n<br>c<br>a<br>y<br>S<br>c<br>e<br>n<br>S<br>c<br>e<br>c<br>o<br>s<br>r<br>e<br>c<br>o<br>r<br>e | F<br>u<br>r<br>e<br>n<br>c<br>a<br>c<br>y<br>S<br>c<br>e<br>c<br>o<br>c<br>o<br>r<br>e | A<br>c<br>c<br>a<br>n<br>S<br>y<br>S<br>e<br>c<br>o<br>c<br>o<br>r<br>e | C<br>o<br>h<br>e<br>r<br>e<br>n<br>c<br>e<br>n<br>c<br>o<br>c<br>o<br>r<br>e | Overall Feedback                                                                                                                                                                                                                     | Comments on Columns                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------|-----------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|-------------------------------------------------------------------------|------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>I<br>k<br>y<br>r<br>i<br>e<br>—<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>—<br>—<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>—<br>—<br>B<br>i | L<br>N<br>t<br>t                               | A                                                   | The people of the State of California do enact as follows:<br><br>SECTION 1.<br>The Legislature finds and declares all of the following:<br>(a) (1) Since 1899 congressionally chartered veterans' organizations have provided a valuable service to our nation's returning service members. These organizations help preserve the memories and incidents of the great hostilities fought by our nation, and preserve and strengthen comradeship among members.<br>(2) These veterans' organizations also own and manage various properties including lodges, posts, and fraternal halls. These properties act as a safe haven where veterans of all ages and their families can gather together to find camaraderie and fellowship, share stories, and seek support from people who understand their unique experiences. This aids in the healing process for these returning veterans, and ensures their health and happiness.<br>(b) As a result of congressional chartering of these veterans' organizations, the United States Internal Revenue Service created a special tax exemption for these organizations under Section 501(c)(19) of the Internal Revenue Code.<br>(c) Section 501(c)(19) of the Internal Revenue Code and related federal regulations provide for the exemption for posts or organizations of war veterans, or an auxiliary unit or society of, or a trust or foundation for, any such post or organization that, among other attributes, carries on programs to perpetuate the memory of deceased veterans and members of the Armed Forces and to comfort their survivors, conducts programs for religious, charitable, scientific, literary, or educational purposes, sponsors or participates in activities of a patriotic nature, and provides social and recreational activities for their members.<br>(d) Section 215.1 of the Revenue and Taxation Code stipulates that all buildings, support and so much of the real property on which the buildings are situated as may be required for the convenient use and occupation of the buildings, used exclusively for charitable purposes, owned by a veterans' organization that has been chartered by the Congress of the United States, organized and operated for charitable purposes, when the same are used solely and exclusively for the purpose of the organization, if not conducted for profit and no part of the net | Directs the Secretary of Veterans Affairs (VA) to establish a program to provide financial incentives to veterans and their family members to assist in the preservation and enhancement of their memories. Authorizes appropriations. Requires the Secretary to make grants to nonprofit organizations to carry out programs to preserve and enhance their memory. Requires such programs to be carried out in a manner consistent with the | 2<br>.0                                                                                     | 1                                                                       | 3                                                                                                                    | 3                                                                                      | 1                                                                       | 2                                                                            | The generated summary fails to capture the main points of the input text and introduces irrelevant information. While it is somewhat fluent and concise, its lack of relevance and accuracy significantly detracts from its quality. | - **Model Name**: The model used is shng2025/GPT-Valkyrie_LN-124m_AttnOnly_Billsum, which appears to be a specialized model for summarizing legislative texts.<br>- **Norm Type**: The normalization type used is LN (Layer Normalization), which is a common technique in neural networks to stabilize and accelerate training.<br>- **Variant**: The variant is AttnOnly, indicating that the model likely uses attention mechanisms exclusively, which should theoretically help in focusing on relevant parts of the input text.<br>- **Truncated Input**: The truncated input provides detailed |

<sup>67</sup> <https://github.com/Ice-Citron/Extended-Essay-Appendix.git> by (Student, 2024)

|                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                      |          |          |          |          |          |  |  |  |  |  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|----------|----------|----------|----------|----------|--|--|--|--|--|
|                                                                                                                                                                                                              | <p>earnings of which ensures to the benefit of any private individual or member thereof, are exempt from taxation.</p> <p>(e) The Chief Counsel of the State Board of Equalization concluded, based on a 1979 appellate court decision, that only parts of American Legion halls are exempt from property taxation and that other parts, such as billiard rooms, card rooms, and similar areas, are not exempt.</p> <p>(f) In a 1994 memorandum, the State Board of Equalization's legal division further concluded that the areas normally considered eligible for exemptions are the office areas used to counsel veterans and the area used to store veterans' records, but that the meeting hall and bar found in most of the facilities are not considered used for charitable purposes.</p> <p>(g) Tax-exempt status is intended to provide economic incentive and support to veterans' organizations to provide for the social welfare of the community of current and former military personnel.</p> <p>(h) The State Board of Equalization's constriction of the tax exemption has resulted in an onerous tax burden on California veteran service organizations posts or halls, hinders the posts' ability to provide facilities for veterans, and threatens the economic viability of many local organizations.</p> <p>(i) The charitable activities of a veteran service organizations post or hall are much more than the counseling of veterans. The requirements listed for qualification for the federal tax exemption clearly dictate a need for more than just an office.</p> <p>(j) Programs to perpetuate the memory of deceased veterans and members of the Armed Forces and to comfort their survivors require the use of facilities for funerals and receptions.</p> <p>(k) Programs for religious, charitable, scientific, literary, or educational purposes require space for more than 50 attendees.</p> <p>(l) Activities of a patriotic nature need facilities to accommodate hundreds of people.</p> <p>(m) Social and recreational activities for members require precisely those areas considered "not used for charitable purposes" by the State Board of Equalization.</p> <p>(n) The State Board of Equalization's interpretation of the Revenue and Taxation Code reflects a lack of understanding of the purpose and programs of the veterans service</p> | <p>requirements of this Act. Provides for the establishment and operation of centers of excellence for the study, study, and study of the role of veterans in the development and implementation of programs to commemorate the service of their predecessors.</p> <p>Prohibits the Secretary from making a grant to a nonprofit organization unless the organization demonstrates to the satisfaction of the Secretary that the organization has a</p> |                      |          |          |          |          |          |  |  |  |  |  |
| <p style="text-align: center;">s<br/>h<br/>n<br/>g<br/>2<br/>0<br/>2<br/>5<br/>/<br/>G<br/>P<br/>T<br/>-<br/>V<br/>a<br/>l<br/>k<br/>y<br/>r<br/>i<br/>e<br/>-<br/>L<br/>N<br/>-<br/>1<br/>2<br/>4<br/>m</p> | <p>The people of the State of California do enact as follows:</p> <p><b>SECTION 1.</b></p> <p>Section 1170.02 is added to the Penal Code, to read:</p> <p><b>1170.02.</b></p> <p>A prisoner is not eligible for resentence or recall pursuant to subdivision (e) of Section 1170 if he or she was convicted of first-degree murder if the victim was a peace officer, as defined in Section 830.1, 830.2, 830.3, 830.31, 830.32, 830.33, 830.34, 830.35, 830.36, 830.37, 830.4, 830.5, 830.6, 830.10, 830.11, or 830.12, who was killed while engaged in the performance of his or her duties, and the individual knew, or reasonably should have known, that the victim was a peace officer engaged in the performance of his or her duties, or the victim was a peace officer or a former peace officer under any of the above-enumerated sections, and was intentionally killed in retaliation for the performance of his or her official duties.</p> <p><b>SEC. 2.</b></p> <p>Section 3550 of the Penal Code is amended to read:</p> <p><b>3550.</b></p> <p>(a) Notwithstanding any other law, except as provided in subdivision (b), if the head physician of an institution in which a prisoner is incarcerated determines, as provided in this section, that the prisoner is permanently medically incapacitated with a medical condition that renders him or her permanently unable to perform activities of basic daily living, and results in the prisoner requiring 24-hour care, and that incapacitation did not exist at the time of sentencing, the prisoner shall be granted medical parole if</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <p>Amends the Violent Crime Control and Law Enforcement Act of 1968 to revise the definition of "prisoner" to include a prisoner who: (1) is a prisoner of war; (2) is incarcerated for a period of not more than one year; and (3) has been convicted of a first degree murder.</p> <p>Requires the Director of the Bureau of Prisons</p>                                                                                                              | <p>2<br/>.<br/>4</p> | <p>1</p> | <p>3</p> | <p>4</p> | <p>1</p> | <p>3</p> |  |  |  |  |  |

|                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                             |         |   |   |   |   |                                                                                                                                                |                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-----------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| -<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>I<br>s<br>u<br>m                                                        | <p>the Board of Parole Hearings determines that the conditions under which he or she would be released would not reasonably pose a threat to public safety.</p> <p>(b) This section does not alter or diminish the rights conferred under the Victims' Bill of Rights Act of 2008 (Marsy's Law). Subdivision (a) does not apply to any of the following:</p> <p>(1) A prisoner sentenced to death or life in prison without possibility of parole.</p> <p>(2) A prisoner who is serving a sentence for which parole, pursuant to subdivision (a), is prohibited by any initiative statute.</p> <p>(3) A prisoner who was convicted of first-degree murder if the victim was a peace officer, as defined in Section 830.1, 830.2, 830.3, 830.31, 830.32, 830.33, 830.34, 830.35, 830.36, 830.37, 830.4, 830.5, 830.6, 830.10, 830.11, or 830.12, who was killed while engaged in the performance of his or her duties, and the individual knew, or reasonably should have known, that the victim was a peace officer engaged in the performance of his or her duties, or the victim was a peace officer or a former peace officer under any of the above-enumerated sections, and was intentionally killed in retaliation for the performance of his or her official duties.</p> <p>(c) When a physician employed by the Department of Corrections and Rehabilitation who is the primary care provider for a prisoner identifies a prisoner that he or she believes meets the medical criteria for medical parole specified in subdivision (a), the primary care physician shall recommend to the head physician of the institution where the prisoner is located that the prisoner be referred to the Board of Parole Hearings for consideration for medical parole. Within 30 days of receiving that recommendation, if the head physician of the institution concurs in the recommendation of the primary care physician, he or she shall refer the matter to the Board of Parole Hearings using a standardized form and format developed by the department, and if the head physician of the institution does not concur in the recommendation, he or she shall provide the primary care physician with a written explanation of the reasons for denying the referral.</p> <p>(d) Notwithstanding any other provisions of this section, the prisoner or his or her family member or designee may independently request consideration</p> | <p>(BOP) to report to the Congress on the status of prisoners in the United States who are incarcerated for at least one year and who are released from prison within one year after the date of the enactment of this Act.</p> <p>Requires the BOP to submit to Congress a plan for the release of prisoners who are not prisoners of war and who have not been released for more than</p> |         |   |   |   |   | ct information, leading to poor relevance and accuracy. The summary needs significant improvement to align with the content of the input text. | which should theoretically help in focusing on relevant parts of the input text.                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                              |
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>l<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m | <p>A The people of the State of California do enact as follows:</p> <p><b>SECTION 1.</b></p> <p>O The Legislature finds and declares all of the following:</p> <p>(a) California has long been known as the land of opportunity, the republic of the future. But for too many of its residents the future is receding. Inequality continues to rise — even though California has one of the most progressive tax structures in the nation.</p> <p>(b) Something more is needed; a new philosophy of governance that focuses on the overall progressive outcome that can be achieved through modernizing our tax system and investing in the means of upward mobility, above all job creating infrastructure and public higher education for our increasingly youthful population.</p> <p>(c) Beyond these foundations, building and sustaining a middle class means new jobs with good wages. Small businesses, like plumbing contractors, auto repair shops, and restaurants that account for over 90 percent of the state's businesses and well over a third of all jobs, are a key rung on the ladder of upward mobility. They need a tax policy that will enable them to grow and add employees.</p> <p>(d) California's two trillion dollar economy has shifted from being mainly agricultural and manufacturing in the 1950s and 1960s, when the framework of today's tax system was set, to one based on information and services, which now accounts for 80 percent of all economic activities in the state. To achieve a future as promising as California's past, we need a tax system that is based on this real economy of the 21st century while ensuring</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <p>Directs the Governor of California to establish a commission to study and report to the Congress on: (1) the future of California; (2) the role of education in the economy; and (3) the impact of education on the environment.</p> <p>Requires the Commission to report to specified congressional committees on its findings and</p>                                                  | 2<br>.8 | 2 | 4 | 4 | 1 | 3                                                                                                                                              | The generated summary is fluent and concise but fails significantly in relevance and accuracy. It does not capture the main points of the input text and introduces unrelated | The model name "shng2025/GPT-Valkyrie_LN-124m_AttnOnly_Billsum" suggests a specialized model for summarizing legislative documents. The "Norm Type" being "LN" indicates Layer Normalization, which is a technique used to stabilize and improve the training of deep neural networks. The "Variant" being "AttnOnly" implies that the model uses attention mechanisms exclusively, which should ideally help in focusing on |

|                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |         |   |   |   |   |   |                                                                                                                                     |                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                        |
|-------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|---|-------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| -<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>I<br>s<br>u<br>m                     | <p>that new revenue is invested in strengthening the ladder of mobility for all our residents.</p> <p>(e) California of the 1950s and 1960s was governed with an eye towards the future and was renowned for the opportunities that it created for its residents. California's water system was born during that era and transformed the desert into fertile agricultural land that not only fed Californians but the world. California also constructed its freeway system to more rapidly and safely move people and goods through the state as California became the gateway to the Pacific Rim. California's higher education system was the envy of all, reaching new heights as the University of California and the California State University grew by six and eight campuses respectively between 1958 and 1965. California's investment in infrastructure and education paid off as agriculture, aerospace, and then technology boomed and drove California into the 21st century as the fifth largest economy in the world. As businesses thrived, they created an abundance of middle class jobs that enabled Californians to capitalize on new opportunities to better the standard of living for themselves and their families.</p> <p>(f) As California's economy thrived, however, its eye on the future wavered. By the late 1970s, state and local finances became intertwined; the state increasingly used its funds to support traditionally local operations and both state and local governments pulled back on the types of investments needed to help businesses and residents succeed. Today, Californians live with the investments made more than three generations ago. Fifty-five percent of our local streets need to be repaired or replaced. While the state's water system received some funding in 2014, more is needed to meet the state's demands.</p> <p>(g) On a local level, 70 percent of Los Angeles' water infrastructure is composed of cast-iron pipes, most of which was laid during the early half of the 20th century.</p> <p>(h) Our financial commitment to kindergarten and grades 1 to 12, inclusive, education has waned. Average Daily Attendance grew anemically by 0.06 percent annually between 2007 and 2011. By 2011, California ranked 43rd in per pupil spending and California's ADA was \$2,580 less than the United States average — the largest gap in 40 years.</p> <p>(i) California's commitment to higher education has also receded. In addition to opening professional and economic doorways for students, California's higher education system is one of our most important economic engines. With almost 60 faculty and researchers who have won the Nobel prize, the University of California has over 3,200 active patents and contributes \$33 billion to the California economy annually. The California State University generates an additional \$17 billion in economic activity and supports 150,000 jobs</p> | <p>recommendations.</p> <p>Authorizes appropriations.</p> <p>Amends the Higher Education Act of 1965 to authorize the Secretary of Education to award competitive grants to institutions of higher education (IHEs) for the purpose of improving the quality of education provided to students in IHEs.</p> <p>Authorizes the Secretary to award grants to eligible institutions to improve the education of such students.</p> <p>Requires the Secretary, in awarding such grants, to give priority to institutions</p> |         |   |   |   |   |   |                                                                                                                                     | information, leading to a low overall score.                                                                                                                                                                                                                                                                  | relevant parts of the input text. However, in this case, the generated summary does not align well with the provided truncated input, indicating potential issues with the model's ability to accurately capture and summarize the key points of the input text.<br>`` |
| s<br>h<br>n<br>g<br>2<br>0<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>I<br>k<br>y<br>r<br>i<br>e<br>-<br>L | <p>The people of the State of California do enact as follows:</p> <p><b>SECTION 1.</b></p> <p>Section 75220 of the Public Resources Code is amended to read: 75220.</p> <p>(a) The Transit and Intercity Rail Capital Program is hereby created to fund transformative capital improvements, as defined in subdivision (d), that will modernize California's intercity, commuter, and urban rail systems and bus and ferry transit systems to achieve all of the following policy objectives:</p> <ul style="list-style-type: none"> <li>(1) Reduce emissions of greenhouse gases.</li> <li>(2) Expand and improve transit service to increase ridership.</li> <li>(3) Integrate the rail service of the state's various rail operators, including integration with the high-speed rail system.</li> <li>(4) Improve transit safety.</li> </ul> <p>(b) The Transportation Agency shall evaluate applications consistent with the criteria set forth in this part and approve a multiyear program of projects for funding pursuant to Section 75224, which may be revised as necessary.</p> <p>(c) The California Transportation Commission shall allocate funding to applicants pursuant to the program of projects approved by the Transportation Agency.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <p>Directs the Transportation Administration (TSA) to:</p> <ul style="list-style-type: none"> <li>(1) develop and implement a multimodal transportation strategy to reduce greenhouse gases; and</li> <li>(2) submit to Congress a plan to implement such strategy.</li> </ul> <p>Authorizes appropriations.</p>                                                                                                                                                                                                         | 2<br>.6 | 2 | 3 | 4 | 1 | 3 | The generated summary fails to accurately represent the content of the truncated input. While it is fluent and somewhat concise, it | - **Model Name**: The model used is shng2025/GPT-V alkyrie_LN-124m __AttnOnly_Bill sum, which appears to be a specialized model for summarizing legislative texts.<br>- **Norm Type**: LN (Layer Normalization) is used, which is a common technique in neural networks to stabilize and accelerate training. |                                                                                                                                                                                                                                                                        |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |             |   |   |   |   |   |                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|---|---|---|---|---|-----------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| N<br>-<br>1<br>2<br>4<br>m<br>-<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>i<br>l<br>s<br>u<br>m                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <p>(d) "Transformative capital improvement" means a rail, bus, or ferry transit project that will significantly reduce vehicle miles traveled, congestion, and greenhouse gas emissions by creating a new transit system, increasing the capacity of an existing transit system, or otherwise significantly increasing the ridership of a transit system.</p> <p><b>SEC. 2.</b></p> <p>Section 75221 of the Public Resources Code is amended to read: 75221.</p> <p>(a) Projects eligible for funding under the program include, but are not limited to, all of the following:</p> <p>(1) Rail capital projects, including acquisition of rail cars and locomotives, that expand, enhance, and improve existing rail systems and connectivity to existing and future transit systems, including the high-speed rail system.</p> <p>(2) Intercity, commuter, and urban rail projects that increase service levels, improve reliability, or decrease travel times, including infrastructure access payments to host railroads in lieu of capital investments.</p> <p>(3) Rail, bus, and ferry integration implementation, including integrated ticketing and scheduling systems, shared-use corridors, related planning efforts, and other service integration initiatives.</p> <p>(4) Bus rapid transit and other bus and ferry transit investments to increase ridership and reduce greenhouse gas emissions.</p> <p>(b) In order to be eligible for funding under the program, a project shall demonstrate that it will achieve a reduction in emissions of greenhouse gases. In selecting projects for funding, the Transportation Agency shall consider the extent to which a project reduces emissions of greenhouse gases.</p> <p>(c) The program shall have a programmatic goal of providing at least 25 percent of available funding to projects benefiting disadvantaged communities, consistent with the objectives of Chapter 830 of the Statutes of 2012.</p> <p>(d) In evaluating grant applications for funding, the Transportation Agency shall consider all of the following:</p> <p>(1) The cobenefits of projects that support the implementation of sustainable communities strategies through one or more of the following:</p> <p>(A) Reducing vehicle miles traveled from automobiles and the number of automobile trips through growth in transit ridership.</p> <p>(B) Promoting housing development in the vicinity of rail stations and major transit centers.</p> <p>(C) Expanding existing rail and public transit systems.</p> <p>(D) Enhancing the connectivity, integration, and coordination of the state's various transit systems, including, but not limited to, regional and local transit systems and the high-speed rail system.</p> <p>(E) Implementing clean vehicle technology.</p> <p>(F) Promoting active transportation.</p> <p>(G) Improving public health.</p> <p>(2) The project priorities developed through the collaboration of two or more rail operators and any memoranda of understanding between state agencies and local or regional rail operators.</p> <p>(3) Geographic equity.</p> <p>(4) Consistency with an adopted sustainable communities strategy or, if a sustainable strategy is not required for a region by law, a regional plan that includes policies and programs to reduce emissions of greenhouse gases.</p> <p>(5) The extent to which a project has supplemental funding committed to it from other nonstate sources.</p> <p>(6) The extent to which the project will increase transit ridership.</p> | <p>Requires the Secretary of Transportation to award competitive grants to eligible entities to carry out projects for the construction, operation, maintenance, and operation of new rail and bus facilities.</p> <p>Requires such entities to submit to specified congressional committees a plan for implementing the strategy. Requires such plan to include a description of the goals and objectives of the strategy, including an assessment of the impact of such strategy on public health and the environment, transportation infrastructure, and transportation systems. Requires the Secretary to award grants to entities to implement the strategy</p> |             |   |   |   |   |   | introduces inaccuracies and irrelevant details, reducing its overall relevance and coherence. | <p>- **Variant**: AttnOnly indicates that the model likely uses attention mechanisms exclusively, which should theoretically help in focusing on relevant parts of the input text.</p> <p>- **Truncated Input**: The input provides detailed information about amendments to the Public Resources Code in California, focusing on funding and criteria for transit projects.</p> <p>- **Generated Summary**: The summary inaccurately represents the input, misidentifying agencies and introducing elements not present in the original text. This suggests that the model may need further fine-tuning to better capture the specifics of legislative texts.</p> |
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/ <td> <p>The people of the State of California do enact as follows:</p> <p><b>SECTION 1.</b></p> <p>This act shall be known, and may be cited, as the California POLST eRegistry Pilot Act.</p> <p><b>SEC. 2.</b></p> <p>Section 4788 is added to the Probate Code, to read: 4788.</p> </td> <td> <p>Directs the Secretary of Health and Human Services (HHS), acting through the Commissioner of the</p> </td> <td>2<br/>.<br/>4</td> <td>1</td> <td>3</td> <td>4</td> <td>1</td> <td>3</td> <td>** The generated summary fails to capture the main</td> <td> <p>** - **Model Name**: The model used is shng2025/GPT-V alkyrie_LN-124m __AttnOnly_Bill sum, which appears to be a</p> </td> | <p>The people of the State of California do enact as follows:</p> <p><b>SECTION 1.</b></p> <p>This act shall be known, and may be cited, as the California POLST eRegistry Pilot Act.</p> <p><b>SEC. 2.</b></p> <p>Section 4788 is added to the Probate Code, to read: 4788.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <p>Directs the Secretary of Health and Human Services (HHS), acting through the Commissioner of the</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 2<br>.<br>4 | 1 | 3 | 4 | 1 | 3 | ** The generated summary fails to capture the main                                            | <p>** - **Model Name**: The model used is shng2025/GPT-V alkyrie_LN-124m __AttnOnly_Bill sum, which appears to be a</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |

|                                                                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |  |  |  |  |  |                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| G<br>P<br>T<br>-<br>V<br>a<br>I<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>I<br>s<br>u<br>m | <p>(a) For purposes of this section:</p> <p>(1) "Authority" means the Emergency Medical Services Authority.</p> <p>(2) "Authorized user" means a person authorized by the authority to submit information to, or to receive information from, the POLST eRegistry Pilot, including health care providers, as defined in Section 4781, and their designees.</p> <p>(3) "POLST" means a Physician Orders for Life Sustaining Treatment that fulfills the requirements, in any format, of Section 4780.</p> <p>(4) "POLST eRegistry Pilot" means the California POLST eRegistry Pilot Act established pursuant to this section to make electronic, in addition to other modes of submission and transmission, POLST information available to authorized users.</p> <p>(b) (1) The authority shall establish a pilot project, in consultation with stakeholders, to operate an electronic registry system on a pilot basis, to be known as the California POLST eRegistry Pilot, for the purpose of collecting a patient's POLST information received from a physician or physician's designee and disseminating the information to an authorized user.</p> <p>(2) The authority shall implement this section only after determining that sufficient nonstate funds are available to allow for the development of the POLST eRegistry Pilot, any related startup costs, and an evaluation of the POLST eRegistry Pilot.</p> <p>(3) The authority shall coordinate the POLST eRegistry Pilot, which shall be operated by, and as a part of, the health information exchange networks, or by an independent contractor, or by a combination thereof. The POLST eRegistry Pilot may operate in a single geographic area or multiple geographic areas and may test various methods of making POLST information available electronically. The design of the POLST eRegistry Pilot shall be sufficiently robust, based on the success of the pilot, to inform the permanent, statewide operation of a POLST eRegistry.</p> <p>(4) The authority shall adopt guidelines necessary for the operation of the POLST eRegistry Pilot. In developing these guidelines, the authority shall seek input from interested parties and hold at least one public meeting. The adoption, amendment, or repeal of the guidelines authorized by this paragraph is hereby exempted from the Administrative Procedure Act (Chapter 3.5 (commencing with Section 11340) of Part 1 of Division 3 of Title 2 of the Government Code). The guidelines shall include, but not be limited to, the following:</p> <ul style="list-style-type: none"> <li>(A) The means by which initial or subsequent POLST information may be submitted to, or withdrawn from, the POLST eRegistry Pilot, which shall include a method for electronic delivery of this information and the use of legally sufficient electronic signatures.</li> <li>(B) Appropriate and timely methods by which the information in the POLST eRegistry Pilot may be disseminated to an authorized user.</li> <li>(C) Procedures for verifying the identity of an authorized user.</li> <li>(D) Procedures to ensure the accuracy of, and to appropriately protect the confidentiality of, POLST information submitted to the POLST eRegistry Pilot.</li> <li>(E) The requirement that a patient, or, when appropriate, his or her legally recognized health care decisionmaker, receive a confirmation or a receipt that the patient's POLST information has been received by the POLST eRegistry Pilot.</li> <li>(F) The ability of a patient, or, when appropriate, his or her legally recognized health care decisionmaker, with his or her health care provider, as defined in Section 4621, to modify or withdraw POLST information on the POLST eRegistry Pilot.</li> <li>(6) (A) Prior to implementation of the POLST eRegistry Pilot, the authority shall submit a detailed plan to the Legislature that explains how the POLST eRegistry Pilot will operate.</li> <li>(B) The plan to be submitted pursuant</li> </ul> | <p>Health Resources and Services Administration (HRSA), to carry out a pilot program to collect and disseminate health care information collected from physicians and physicians.</p> <p>Requires the Secretary to establish a program for the collection and dissemination of such information. Authorizes appropriations.</p> <p>Prohibits the Secretary from making grants to any State or local government for the establishment or operation of an electronic health record system.</p> <p>Directs HHS to establish an electronic database of health information collected by physicians and health care professionals. Requires the database to be maintained by the National Institute of Standards and Technology (NIST).</p> <p>Provides for the submission</p> |  |  |  |  |  | points of the input text and introduces irrelevant information. While it is well-written and organized, it does not accurately represent the content of the input, leading to a low overall score. | <p>specialized model for summarizing legislative texts.</p> <ul style="list-style-type: none"> <li>- **Norm Type:** LN (Layer Normalization) is a common technique used in neural networks to stabilize and accelerate training.</li> <li>- **Variant:** AttnOnly indicates that the model likely uses attention mechanisms exclusively, which should theoretically help in focusing on relevant parts of the input text.</li> <li>- **Truncated Input:** The input text provides detailed information about the California POLST eRegistry Pilot Act, including definitions, the establishment of a pilot project, and guidelines for its operation.</li> <li>- **Generated Summary:** The summary incorrectly focuses on directives for the Secretary of Health and Human Services and other unrelated federal programs, failing to capture the essence of the input text.</li> </ul> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                                                                                                                                                                    |                                                                                                                                                         |   |                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |             |   |   |   |   |   |                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|---|------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|---|---|---|---|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|                                                                                                                                                                                                                                    |                                                                                                                                                         |   |                                                            | of a report to Congress on the                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |             |   |   |   |   |   |                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>I<br>k<br>y<br>r<br>i<br>e<br>—<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>—<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>—<br>B<br>i<br>I<br>l<br>s<br>u<br>m | L<br>N<br>t<br>O<br>n<br>y<br>I<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>—<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>—<br>B<br>i<br>I<br>l<br>s<br>u<br>m | A | The people of the State of California do enact as follows: | (This measure has not been amended since it was reported to the House on September 30, 2016. The summary of that version is repeated here.) (Sec. 2) This bill amends the Internal Revenue Code to: (1) allow a tax credit for the tax deduction for the taxable year, and (2) provide for a tax deduction of up to ten percent of the amount of the tax credit. (Sec. 3) Amends the Immigration and Nationality Act to require the Secretary of Homeland Security (DHS) to report annually to Congress on the progress made by the United States in reducing the number of aliens admitted for permanent resident status. Directs the Secretary to establish a program to provide grants to states. | 2<br>.<br>6 | 1 | 3 | 4 | 1 | 4 | The generated summary is fluent and coherent but fails significantly in relevance and accuracy. It discusses topics that are entirely unrelated to the provided input text, making it an ineffective summary for the given content. | - **Model Name**: The model name suggests it is a variant of GPT-Valkyrie with a focus on attention mechanisms, specifically designed for summarizing bills.<br>- **Norm Type**: LN (Layer Normalization) is a common technique used in neural networks to stabilize and accelerate training.<br>- **Variant**: AttnOnly indicates that the model uses attention mechanisms exclusively, which should theoretically help in focusing on relevant parts of the input text.<br>- **Truncated Input**: The input text is a detailed legislative amendment regarding income reporting requirements in the Government Code.<br>- **Generated Summary**: The summary provided is completely unrelated to the input text, discussing amendments to the Internal Revenue Code and the Immigration and Nationality Act instead. This indicates a significant failure in the model's ability to generate |  |

|                                                                                                                                                                                                                                         |                                      |   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |             |   |   |   |   |   |                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|---|---|---|---|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                                                                                                                         |                                      |   | (E) Greater than five hundred thousand dollars (\$500,000).<br>(3) A description of the consideration, if any, for which the income was received.<br>(4) In the case                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |             |   |   |   |   |   |                                                                                                                                                                                                                                                                               | relevant summaries for the given input.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>l<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>i<br>i<br>s<br>u<br>m | L<br>N<br>t<br>t<br>n<br>o<br>n<br>y | A | <p>The people of the State of California do enact as follows:</p> <p><b>SECTION 1.</b></p> <p>The Legislature finds and declares as follows:</p> <p>(a) More than \$40 million of funding for the training of California's primary care physicians is expiring in 2016.</p> <p>(b) Each year in California, only 368 slots are available to the thousands of medical students seeking to train in family medicine. If the funding is not replaced, 158 of those slots will be lost, creating a terrible deficit of primary care physicians in California's underserved communities.</p> <p>(c) Only 36 percent of California's active patient care physicians practice primary care. Twenty-three of California's 58 counties fall below the minimum required primary care physician to population ratio.</p> <p>(d) As of 2010, California needed an estimated additional 8,243 primary care physicians by 2030 to prevent projected shortages in the state, which is about 412 new primary care physicians per year.</p> <p>(e) More than 32 percent of California's practicing primary care physicians are 60 years of age or older – only four other states have a larger percentage of soon-to-retire physicians.</p> <p>(f) States with higher ratios of primary care physicians to population have better health outcomes, including decreased mortality from cancer, heart disease, and stroke.</p> <p>(g) The Song-Brown program provides an existing state infrastructure to support an increase in the number of primary care providers serving California's underserved populations. By investing in Song-Brown, California will realize an immediate return on investment as each primary care resident provides an average of 600 additional patient visits per physician per year during training alone.</p> <p>(h) California's long-term workforce will also grow significantly as the vast majority of physicians who train in a region stay there to practice. California leads all fifty states in the percentage of residency program graduates who stay in the state in which they are trained.</p> <p><b>SEC. 2.</b></p> <p>Notwithstanding Section 13340 of the Government Code, there is hereby continuously appropriated from the General Fund the sum of three hundred million dollars (\$300,000,000) to the Director of Statewide Health Planning and Development, for the purpose of funding new and existing graduate medical education physician residency positions, and supporting training faculty, pursuant to the Song-Brown Health Care Workforce Training Act (Article 1 (commencing with Section 128200) of Chapter 4 of Part 3 of Division 107 of the Health and Safety Code). The moneys shall be expended as follows:</p> <p>(a) The sum of one hundred million dollars (\$100,000,000) shall be expended in the 2016–17 fiscal year.</p> <p>(b) The sum of one hundred million dollars (\$100,000,000) shall be expended in the 2017–18 fiscal year.</p> <p>(c) The sum of one hundred million dollars (\$100,000,000) shall be expended in the 2018–19 fiscal year.</p> <p><b>SECTION 1.</b></p> <p>Article 7 (commencing with Section 128590) is added to Chapter 5 of Part 3 of Division 107 of the Health and Safety Code , to read:</p> <p>7.</p> <p>California Medical Residency Training Program 128590.</p> <p>As used in this article:</p> <p>(a) "Director" means the Director of Statewide Health Planning and Development.</p> | <p>Directs the Secretary of Health and Human Services (HHS) to establish within the Foundation the Advisory Panel to: (1) advise the Secretary on matters relating to the provision of health care services in California; and (2) report to the Congress on such matters. Authorizes appropriations. Requires the Secretary to make grants to eligible entities to carry out this Act.</p> <p>Requires a grantee to submit to the Secretary a plan for the establishment of a primary care residency training program. Requires the Secretary, in carrying out the program, to submit an annual report to Congress on the program. Directs HHS to establish a program to award grants to nonprofit organizations to assist in the establishment and</p> | 2<br>.<br>6 | 2 | 3 | 4 | 1 | 3 | ** The generated summary fails to accurately capture the main points of the input text, focusing instead on unrelated directives and grants. While it is fluent and somewhat concise, its relevance and accuracy are significantly lacking, leading to a lower overall score. | ** - **Model Name:** The model used is shng2025/GPT-V alkyrie_LN-124m _AttnOnly_Bill sum, which appears to be a specialized model for summarizing legislative texts. - **Norm Type:** The normalization type used is LN (Layer Normalization), which is a common technique in neural networks to stabilize and accelerate training. - **Variant:** The variant is AttnOnly, indicating that the model likely uses attention mechanisms exclusively, which should theoretically help in focusing on relevant parts of the input text. - **Truncated Input:** The input text provides detailed information about the funding and establishment of primary care residency positions in California, highlighting the need for more primary care physicians and the financial allocations to support this. - **Generated Summary:** The summary generated does not align well with the input text, introducing |

|                                                                                                                                                                                                                                              |                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                        |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                               |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
|                                                                                                                                                                                                                                              |                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | (b) "Foundation" means the Health Professions Education Foundation.<br>(c) "Fund" means the Medical Residency Training Fund.<br>(d) "Office" means the Office of Statewide Health Planning and Development.<br>(e) "Panel" means the Medical Residency Training Advisory Panel, established pursuant to Section 128591.<br>(f) "Primary care" means the medical practice areas of family medicine, general surgery, internal medicine, obstetrics and gynecology, pediatrics, psychiatry, and related specialties and subspecialties as the office deems appropriate.<br>(g) "Residency position" means a graduate medical education residency position in the field of primary care.<br>128591.<br>(a)(1)There is established within the foundation the Medical Residency Training Advisory Panel.<br>(2)The panel shall consist of 13 members | operation of such programs. Requires such grants to be |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | unrelated elements and failing to capture the key points about funding and residency positions in California. |
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>l<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>I<br>I<br>s<br>u<br>m | L<br>N<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>I<br>I<br>s<br>u<br>m | A<br>t<br><br>SECTION 1.<br>The Legislature finds and declares all of the following:<br>(a) Scientific research has demonstrated that young children living in deep poverty experience lifelong cognitive impairments limiting their ability to be prepared for, and succeed in, school.<br>(b) Academic research has documented an increase in missed days of school and an increase in visits to hospital emergency rooms by children who live in deep poverty.<br>(c) The Maximum Family Grant rule was adopted to limit the length of time a family could receive basic needs assistance, and to limit the amount of assistance a family could receive, through the Aid to Families with Dependent Children (AFDC) program before the implementation of welfare reform. At the time the rule was adopted, there was no limit on the length of time a family could receive aid, no work requirements, and the benefits provided were approximately 80 percent of the federal poverty level.<br>(d) Since the implementation of the Maximum Family Grant rule, AFDC has been replaced with the California Work Opportunity and Responsibility to Kids Act (CalWORKs), which imposes lifetime limits on aid and requires adult CalWORKs participants to meet work requirements in order to receive a maximum benefit of approximately 40 percent of the federal poverty level.<br>(e) The Maximum Family Grant rule makes poor children poorer, reducing the income of families with infants to below 30 percent of the federal poverty level.<br>(f) This act is necessary to protect infants born to families receiving CalWORKs from experiencing lifelong cognitive impairments due to the toxic stress of deep poverty and to ready those children for participation in California's public school system.<br>(g) This act is also necessary to protect the reproductive and privacy rights of all applicants for, and recipients of, aid under CalWORKs.<br>SEC. 2.<br>Section 11270.5 is added to the Welfare and Institutions Code, immediately following Section 11270, to read:<br>11270.5.<br>(a) An applicant for, or recipient of, aid under this chapter shall not be required, as a condition of eligibility, to do any of the following:<br>(1) Divulge that any member of the assistance unit is a victim of rape or incest.<br>(2) Share confidential medical records related to any member of the assistance unit's rape or incest.<br>(3) Use contraception, choose a particular method of contraception, or divulge the method of contraception that any member of the assistance unit uses.<br>(b) An applicant for, or recipient of, aid under this chapter shall not be denied aid, nor denied an increase in the maximum aid | Directs the Secretary of Health and Human Services (HHS), acting through the Director of the Centers for Disease Control and Prevention (CDC), to: (1) establish a program to provide assistance to low-income families with children who are at risk for developmental disabilities; and (2) make grants to States for the purpose of providing assistance to such families.<br><br>Requires the Administrator of the Agency for Healthcare Research and Quality (AHRQ) to establish a pilot program to test the feasibility and advisability of providing such assistance                                                                                                                                                                                                                                                                     | 2<br>.4                                                | 1 | 3 | 4 | 1 | 3 | The generated summary is fluent and concise but fails to capture the main points of the input text. It inaccurately represents the content, leading to a low relevance and accuracy score. The summary needs to be more aligned with the specific legislative changes and protections discussed in the input text. | - **Model Name**: The model name indicates that it is a variant of GPT-Valkyrie with layer normalization and attention-only mechanisms, designed for summarizing bills.<br>- **Norm Type**: LN (Layer Normalization) suggests that the model uses layer normalization techniques to stabilize and improve training.<br>- **Variant**: AttnOnly indicates that the model relies solely on attention mechanisms, which should theoretically help in focusing on relevant parts of the input text.<br>- **Truncated Input**: The input text provides detailed legislative changes regarding aid to families in deep poverty in California, focusing on the repeal of the Maximum Family Grant rule and the protection of |                                                                                                               |

|                                                                                                                                                                                                           |                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |         |   |   |   |   |   |                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                       |                                                |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------|
|                                                                                                                                                                                                           |                                           | <p>payment, for a child born into the applicant's or recipient's family during a period in which the applicant's or recipient's family was receiving aid under this chapter.</p> <p>(c) An applicant for, or recipient of, aid under this chapter shall not be entitled to an increased benefit payment for any month prior to January 1, 2016, as a result of the repeal of former Section 11450.04 (as added by Section 1 of Chapter 196 of the Statutes of 1994) or the enactment of this section.</p> <p><b>SEC. 3.</b></p> <p>Section 11450.04 of the Welfare and Institutions Code is repealed. 11450.04.</p> <p>(a) For purposes of determining the maximum aid payment specified in subdivision (a) of Section 11450 and for no other purpose, the number of needy persons in the same family shall not be increased for any child born into a family that has received aid under this chapter continuously for the 10 months prior to the birth of the child. For purposes of this section, aid shall be considered continuous unless the family does not receive aid during two consecutive months. This subdivision shall not apply to applicants for, or recipients of, aid unless notification is provided pursuant to this section.</p> <p>(b) This section shall not apply with respect to any of the following children:</p> <p>(1) Any child who was conceived as a result of an act of rape, as defined in Sections 261 and 262 of the Penal Code, if the rape was reported to a law enforcement agency, medical or mental health professional or social services agency prior to, or within three</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | <p>to individuals with developmental disabilities. Authorizes the Secretary to award grants to eligible entities to carry out the pilot program. Requires the Secretary, acting through AHRQ, to establish an advisory committee to advise the Secretary on</p>                                                                                                                                                                                                          |         |   |   |   |   |   |                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                       | reproductive and privacy rights.               |
| s<br>h<br>n<br>g<br>2<br>0<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>I<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i | L<br>N<br>t<br>t<br>O<br>n<br>I<br>y<br>- | <p>The people of the State of California do enact as follows:</p> <p><b>SECTION 1.</b></p> <p>Section 7522.02 of the Government Code is amended to read: 7522.02.</p> <p>(a) (1) Notwithstanding any other law, except as provided in this article, on and after January 1, 2013, this article shall apply to all state and local public retirement systems and to their participating employers, including the Public Employees' Retirement System, the State Teachers' Retirement System, the Legislators' Retirement System, the Judges' Retirement System, the Judges' Retirement System II, county and district retirement systems created pursuant to the County Employees Retirement Law of 1937 (Chapter 3 (commencing with Section 31450) of Part 3 of Division 4 of Title 3), independent public retirement systems, and to individual retirement plans offered by public employers. However, this article shall be subject to the Internal Revenue Code and Section 17 of Article XVI of the California Constitution. The administration of the requirements of this article shall comply with applicable provisions of the Internal Revenue Code and the Revenue and Taxation Code.</p> <p>(2) Notwithstanding paragraph (1), this article shall not apply to the entities described in Section 9 of Article IX of, and Sections 4 and 5 of Article XI of, the California Constitution, except to the extent that these entities continue to be participating employers in any retirement system governed by state statute. Accordingly, any retirement plan approved before January 1, 2013, by the voters of any entity excluded from coverage by this section shall not be affected by this article.</p> <p>(3) (A) Notwithstanding paragraph (1), this article shall not apply to a public employee whose interests are protected under Section 5333(b) of Title 49 of the United States Code until a federal district court rules that the United States Secretary of Labor, or his or her designee, erred in determining that the application of this article precludes certification under that section, or until January 1, 2016, whichever is sooner.</p> <p>(B) If a federal district court upholds the determination of the United States Secretary of Labor, or his or her designee, that application of this article precludes him or her from providing a certification under Section 5333(b) of Title 49 of the United States Code, this article shall not apply to a public employee specified in subparagraph (A).</p> | <p>Amends the Federal Employees' Retirement System (FERS) to provide for the establishment of an independent public employee retirement system.</p> <p>Requires the Secretary of the Treasury to establish a program to be known as the Balanced Budget and Emergency Deficit Control Act of 1985 (Gramm-Rudman-Hollings Act).</p> <p>Directs the Secretary to establish an Office of Management and Budget (OMB) to be headed by a Director of OMB.</p> <p>Requires</p> | 2<br>.4 | 1 | 3 | 4 | 1 | 3 | The generated summary fails to capture the main points of the truncated input and introduces unrelated topics, leading to inaccuracies. While the summary is relatively well-written and organized, its lack of relevance and accuracy significantly detract | - **Model Name**: The model name indicates that it is a variant of GPT-Valkyrie with a specific focus on attention mechanisms and is trained on the Billsum dataset. However, the performance in this instance suggests that the model may not be well-suited for summarizing legal texts accurately. | - **Norm Type**: LN (Layer Normalization) is a common technique used in neural networks to stabilize and accelerate training. While it generally improves model performance, it did not prevent the generation of an irrelevant summary in this case. | - **Variant**: The "AttnOnly" variant suggests |

|                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                          |         |   |   |   |   |   |                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|----------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|---|---------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| I<br>I<br>s<br>u<br>m                                    | (4) Notwithstanding paragraph (1), this article shall not apply to a multiemployer plan authorized by Section 302(c)(5) of the federal Taft-Hartley Act (29 U.S.C. Sec. 186(c)(5)) if the public employer began participation in that plan prior to January 1, 2013, and the plan is regulated by the federal Employee Retirement Income Security Act of 1974 (29 U.S.C. Sec. 1001 et seq.).<br>(b) The benefit plan required by this article shall apply to public employees who are new members as defined in Section 7522.04.<br>(c) (1) Individuals who were employed by any public employer before January 1, 2013, and who became employed by a subsequent public employer for the first time on or after January 1, 2013, shall be subject to the retirement plan that would have been available to employees of the subsequent employer who were first employed by the subsequent employer on or before December 31, 2012, if the individual was subject to concurrent membership for which creditable service was performed in the previous six months or reciprocity established under any of the following provisions:<br>(A) Article 5 (commencing with Section 20350) of Chapter 3 of Part 3 of Division 5 of Title 2.<br>(B) Chapter 3 (commencing with Section 31450) of Part 3 of Division 4 of Title 3.<br>(C) Any agreement between public retirement systems to provide reciprocity to members of the systems.<br>(D) Section 22115.2 of the Education Code.<br>(2) An individual who was employed before January 1, 2013, and who, without a separation from employment, changed employment positions and became subject to a different defined benefit plan in a different public retirement system offered by his or her employer shall be subject to that defined benefit plan as it would have been available to employees who were first employed on or before | OMB to report to Congress on the effectiveness of the program. Authorizes appropriations. Prohibits the use of funds appropriated to carry out this Act from being used for any purpose other than the provision of benefits under this Act. Provides for the payment of |         |   |   |   |   |   | s from its overall quality.                                         | that the model uses attention mechanisms exclusively, which are typically effective for capturing important information. However, the generated summary indicates that the attention mechanism may not have been effectively utilized.<br>- **Truncated Input**: The truncated input provides detailed amendments to the California Government Code regarding public retirement systems. It is specific and focused on state-level legislative changes.<br>- **Generated Summary**: The generated summary incorrectly discusses federal programs and fails to address the content of the truncated input. This discrepancy highlights a significant issue with the model's ability to generate relevant and accurate summaries for this type of input. |
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T | A The people of the State of California do enact as follows:<br><br>SECTION 1.<br>Chapter 4.5 (commencing with Section 14400) is added to Division 7 of the Food and Agricultural Code, to read:<br>CHAPTER 4.5. Livestock: Use of Antimicrobial Drugs 14400.<br>For purposes of this chapter, the following definitions apply:<br>(a) "Medically important antimicrobial drug" means an antimicrobial drug listed in Appendix A of the federal Food and Drug Administration's Guidance for Industry #152, including                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | (This measure has not been amended since it was passed by the Senate on December 22, 2016. Livestocks are the                                                                                                                                                            | 2<br>.4 | 1 | 3 | 4 | 1 | 3 | The generated summary fails to capture the main points of the input | - **Model Name**: The model used is shng2025/GPT-V alkyrie_LN-124m __AttnOnly_Bill sum, which appears to be a specialized model for summarizing legislative texts.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |

|                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |  |  |  |  |                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>- V<br/>a<br/>i<br/>k<br/>y<br/>r<br/>i<br/>e<br/><br/>- L<br/>N<br/>- 1<br/>2<br/>4<br/>m<br/>- - A<br/>t<br/>t<br/>n<br/>O<br/>n<br/>I<br/>y<br/>- - B<br/>i<br/>I<br/>I<br/>s<br/>u<br/>m</p> | <p>critically important, highly important, and important antimicrobial drugs, as that appendix may be amended.</p> <p>(b) "Livestock" means all animals and poultry, including aquatic and amphibian species, that are raised, kept, or used for profit. Livestock does not include bees or those species that are usually kept as pets, such as dogs, cats, and pet birds.</p> <p>(c) "Veterinary feed directive" has the same definition as in Section 558.3 of Title 21 of the Code of Federal Regulations.</p> <p><b>14401.</b></p> <p>Beginning January 1, 2018, a medically important antimicrobial drug shall not be administered to livestock unless ordered by a licensed veterinarian through a prescription or veterinary feed directive, pursuant to a veterinarian-client-patient relationship that meets the requirements of Section 2032.1 of Title 16 of the California Code of Regulations.</p> <p><b>14402.</b></p> <p>(a) Beginning January 1, 2018, a medically important antimicrobial drug may be used when, in the professional judgment of a licensed veterinarian, the medically important antimicrobial drug is any of the following:</p> <ul style="list-style-type: none"> <li>(1) Necessary to treat a disease or infection.</li> <li>(2) Necessary to control the spread of a disease or infection.</li> <li>(3) Necessary in relation to surgery or a medical procedure.</li> </ul> <p>(b) A medically important antimicrobial drug may also be used when, in the professional judgment of a licensed veterinarian, it is needed for prophylaxis to address an elevated risk of contraction of a particular disease or infection.</p> <p>(c) A person shall not administer a medically important antimicrobial drug to livestock solely for purposes of promoting weight gain or improving feed efficiency.</p> <p>(d) Unless the administration is consistent with subdivision (a), a person shall not administer a medically important antimicrobial drug in a regular pattern.</p> <p><b>14403.</b></p> <p>(a) Notwithstanding Sections 14401 and 14402 of this code and Article 15 (commencing with Section 4196) of Chapter 9 of Division 2 of the Business and Professions Code, medically important antimicrobial drugs may be sold by retailers licensed pursuant to Article 5 (commencing with Section 14321) of Chapter 4 of Division 7 with a prescription or veterinary feed directive from a licensed veterinarian.</p> <p>(b) This section shall not be construed to invalidate the requirement to obtain a prescription or veterinary feed directive to administer a medically important antimicrobial drug as required by Section 14401.</p> <p>(c) The department may promulgate regulations to implement this section.</p> <p><b>14404.</b></p> <p>(a) The department, in consultation with the Veterinary Medical Board, the State Department of Public Health, universities, and cooperative extensions, shall develop antimicrobial stewardship guidelines and best management practices for veterinarians, as well as livestock owners and their employees who are involved with administering medically important antimicrobial drugs, on the proper use of medically important antimicrobial drugs for disease treatment, control, and prevention. The guidelines shall include scientifically validated practical alternatives to the use of medically important antimicrobial drugs, including, but not limited to, the introduction of effective vaccines and good hygiene and management practices.</p> <p>(b) The department shall consult with livestock producers, licensed veterinarians, and any other relevant stakeholders on ensuring livestock timely access to treatment for producers in rural areas with limited access to veterinary care.</p> <p>(c) For purposes of this section, "antimicrobial stewardship" is a commitment to do all of the following:</p> <ul style="list-style-type: none"> <li>(1) To use medically important antimicrobial drugs only when necessary to treat, control, and, in some cases, prevent, disease.</li> <li>(2) To select the appropriate medically important antimicrobial drug and</li> </ul> | <p>backbone of the economy of the United States. The Department of Agriculture (USDA) and the Department of Health and Human Services (HHS) must jointly develop and implement a comprehensive strategy to address antimicrobial resistance in livestock. (Sec. 4) This bill amends the Federal Food, Drug, and Cosmetic Act (FFDCA) to direct the Secretary of Agriculture to: (1) develop a strategic plan for the development and implementation of a comprehensive antimicrobial strategy for livestock; and (2) submit to Congress a report on the progress of the strategy. USDA must develop a plan</p> |  |  |  |  |  | <p>text and introduces irrelevant and inaccurate information. While it is fluently written and somewhat concise, its lack of relevance and accuracy significantly detracts from its overall quality.</p> | <p>- **Norm Type**: LN (Layer Normalization) is a common technique used in neural networks to stabilize and accelerate training.</p> <p>- **Variant**: AttnOnly indicates that the model likely uses attention mechanisms, which are designed to focus on the most relevant parts of the input text.</p> <p>- **Truncated Input**: The input text provides detailed regulations on the use of medically important antimicrobial drugs in livestock in California, including definitions, conditions for use, and guidelines for antimicrobial stewardship.</p> <p>- **Generated Summary**: The summary incorrectly focuses on a federal strategy for antimicrobial resistance and amendments to the Federal Food, Drug, and Cosmetic Act, which are not mentioned in the input text. This indicates a significant mismatch between the input and the generated output.</p> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |         |   |   |   |   |   |                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>I<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>l<br>y<br>-<br>B<br>i<br>I<br>I<br>s<br>u<br>m | A<br><br>The people of the State of California do enact as follows:<br><br>SECTION 1.<br>Section 13515.28 is added to the Penal Code, to read:<br>13515.28.<br>(a) (1) The commission shall require the field training officers who provide instruction in the field training program to have at least eight hours of crisis intervention behavioral health training to better train new peace officers on how to effectively interact with persons with mental illness or intellectual disability. This course shall include classroom instruction and instructor-led active learning, such as scenario-based training, and shall be taught in segments that are at least four hours long.<br>(2) If a field training officer has completed eight hours of crisis intervention behavioral health training within the past 24 months, or if a field training officer has completed 40 hours of crisis intervention behavioral health training, the requirement described in paragraph (1) shall not apply.<br>(b) The crisis intervention behavioral health training shall address issues relating to stigma, shall be culturally relevant and appropriate, and shall include all of the following topics:<br>(1) The cause and nature of mental illnesses and intellectual disabilities.<br>(2) (A) How to identify indicators of mental illness, intellectual disability, and substance use disorders.<br>(B) How to distinguish between mental illness, intellectual disability, and substance use disorders.<br>(C) How to respond appropriately in a variety of situations involving persons with mental illness, intellectual disability, and substance use disorders.<br>(3) Conflict resolution and deescalation techniques for potentially dangerous situations.<br>(4) Appropriate language usage when interacting with potentially emotionally distressed persons.<br>(5) Community and state resources available to serve persons with mental illness or intellectual disability, and how these resources can be best utilized by law enforcement.<br>(6) The perspective of individuals or families who have experiences with persons with mental illness, intellectual disability, and substance use disorders.<br>(c) Field training officers assigned or appointed before January 1, 2017, shall complete the crisis intervention behavioral health training by June 30, 2017. Field training officers assigned or appointed on or after January 1, 2017, shall complete the crisis intervention behavioral health training within 180 days of assignment or appointment.<br>(d) This section does not prevent an agency from requiring its field training officers to complete additional hours of crisis intervention behavioral health training or requiring its field training officers to complete that training earlier than as required by this section.<br>SEC. 2.<br>Section 13515.29 is added to the Penal Code, to read:<br>13515.29.<br>(a) The commission shall establish and keep updated a field training officer course relating to competencies of the field training program and police training program that addresses how to interact with persons with mental illness or intellectual disability.<br>(b) This course shall consist of at least four hours of classroom instruction and instructor-led active learning, such as scenario-based training, shall address issues related to stigma, and shall be culturally relevant and appropriate.<br>(c) All prospective field training officers shall complete the course described in subdivisions (a) and (b) as part of the existing field training officer program.<br>(d) The commission shall implement the provisions of this section on or before August 1, 2016.<br>SEC. 3.<br>Section 13515.295 is added to the Penal Code, to read: | Amends the Criminal Justice and Delinquency Prevention Act of 1994 to direct the Attorney General to conduct a study and report to the Congress on the effectiveness of the criminal justice system in preventing and responding to crimes involving mental illness and intellectual disability in the United States.<br><br>Directs the Secretary of Justice to establish and maintain a national clearinghouse for the collection, analysis, and dissemination of information related to the prevention, response, and recovery of individuals with mental illnesses or intellectual disabilities who have been convicted of a crime of violence committed by a person who has committed such a crime.<br><br>Authorizes the Secretary to make grants to State and local law | 2<br>.4 | 1 | 3 | 4 | 1 | 3 | The generated summary is fluent and somewhat concise but fails to capture the main points of the input text. It inaccurately represents the content and lacks relevance and coherence with the provided input. | - **Model Name**: The model used is shng2025/GPT-V alkyrie_LN-124m<br>- **AttnOnly**: sum, which appears to be designed for summarizing legislative texts.<br>- **Norm Type**: LN (Layer Normalization) is a common technique used in neural networks to stabilize and accelerate training.<br>- **Variant**: AttnOnly indicates that the model likely uses attention mechanisms, which are crucial for understanding context in text.<br>- **Truncated Input**: The input text provides detailed requirements for crisis intervention behavioral health training for field training officers, including specific topics and timelines.<br>- **Generated Summary**: The summary incorrectly focuses on amendments to a different act and the establishment of a national clearinghouse, which are not mentioned in the input text. This indicates a significant mismatch between the input and the generated output. |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                                                                                                                                                                              |                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |             |   |   |   |   |   |                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |  |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|---|---|---|---|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|                                                                                                                                                                                                                                              |                                                                                   | <p>13515.295.</p> <p>(a) The commission shall, by May 1, 2016, conduct a review and evaluation of the required competencies of the field training program and police training program to identify areas where additional training is necessary to better prepare law enforcement officers to effectively address incidents involving persons with a mental illness or intellectual disability.</p> <p>(b) Upon identifying what additional training is needed, the commission shall update the training in consultation with appropriate community, local, and state organizations, and agencies that have expertise in the area of mental illness, intellectual disabilities, and substance abuse disorders, and with appropriate consumer and family advocate groups.</p> <p>(c) The training shall address issues related to stigma, shall be culturally relevant and appropriate, and shall include all of the following topics:</p> <p>(1) How to identify indicators of mental illness, intellectual disability, substance use disorders, neurological disorders, traumatic brain injury, post-traumatic stress disorder, and</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | <p>enforcement agencies to assist in the collection and analysis of information regarding the prevention and response of individuals who have committed such</p>                                                                                                                                                                                                                                                                                                                                                                                             |             |   |   |   |   |   |                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |  |
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>I<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>I<br>l<br>s<br>u<br>m | L<br>N<br>t<br>t<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>I<br>l<br>s<br>u<br>m | <p>The people of the State of California do enact as follows:</p> <p><b>SECTION 1.</b></p> <p>(a) The Legislature finds and declares all of the following:</p> <p>(1) Existing federal law (42 U.S.C. Sec. 1396p) requires state Medicaid programs to seek reimbursement from the estates of deceased Medicaid beneficiaries, or from any recipient of the decedent's property by distribution or survival, for Medicaid paid services received on or after 55 years of age, unless specific exemptions or other limitations apply.</p> <p>(2) Federal law requires states to collect for long-term services and supports for individuals 55 years of age or older, and gives states the option to collect for other health care services.</p> <p>(3) Federal law permits states to collect from the surviving spouse of a Medi-Cal beneficiary, but does not require collection upon the passing of a spouse of a deceased Medi-Cal beneficiary.</p> <p>(4) Federal law defines "estate" for purposes of estate recovery to include all real and personal property and other assets included within the individual's estate, as defined for purposes of state probate law, and permits states to have a broader definition of estate.</p> <p>(5) The State Medicaid Manual allows states to establish an undue hardship exemption from estate recovery for a homestead of "modest value," defined as a home valued at 50 percent or less of the average price of homes in the county where the homestead is located as of the date of the beneficiary's death.</p> <p>(6) Estate recovery is unfair to low-income individuals who need Medi-Cal for basic health care coverage, is a deterrent to signing individuals up for Medi-Cal, and is counter to both state and federal efforts to enroll individuals into health care coverage.</p> <p>(7) By recovering for health care services beyond what is required by federal law, California forces low-income individuals 55 years of age or older to choose between signing up for basic health care services and passing on their home and other limited assets they possess to their children.</p> <p>(8) California's estate recovery program undermines the idea of Medi-Cal as a health care entitlement program by essentially turning Medi-Cal coverage for basic medical services into a loan program, with collection taking place at death.</p> <p>(9) Estate recovery unfairly places part of the burden of financing the cost of health care in Medi-Cal on the estates of deceased Medi-Cal beneficiaries with limited assets.</p> <p>(10) Estate recovery is inequitable as other social and health care programs, such as tax-subsidized coverage through the California Health Benefit Exchange, commonly referred to as Covered California, and the broadly financed federal Medicare program, do not have estate recovery.</p> <p>(11) California does not adequately inform individuals on how to obtain information on the amounts that will be collected from</p> | <p>Amends title XVIII (Medicare) of the Social Security Act to direct the Secretary of Health and Human Services (HHS), acting through the Director of the Centers for Medicare &amp; Medicaid Services (CMS), to: (1) establish a program for the collection of income from deceased Medicare beneficiaries ; and (2) make grants to States to provide such services to such beneficiaries .</p> <p>Directs the Secretary, acting through CMS, to provide grants to states to provide services to beneficiaries who are unable to obtain such services.</p> | 2<br>.<br>4 | 1 | 3 | 4 | 1 | 3 | The generated summary fails to capture the main points of the input text and inaccurately represents the content . While it is concise and somewhat fluent, the lack of relevance and accuracy significantly impacts its overall quality. | - **Model Name**: The model used is shng2025/GPT-V alkyrie_LN-124m __AttnOnly_Bill sum, which appears to be a specialized model for summarizing legislative texts.<br>- **Norm Type**: LN (Layer Normalization) is a common technique used in neural networks to stabilize and accelerate training.<br>- **Variant**: AttnOnly indicates that the model likely uses attention mechanisms, which are crucial for understanding and summarizing long texts.<br>- **Truncated Input**: The input text provides detailed information about California's Medi-Cal estate recovery program and the legislative intent to limit it.<br>- **Generated Summary**: The summary incorrectly |  |

|                                                                                                                                                                                                                  |                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |         |   |   |   |   |   |                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                               |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                                                                                                  |                                                                | <p>their estate, and charges individuals \$25 to find out how much Medi-Cal has spent on their behalf.</p> <p>(b) It is the intent of the Legislature, with the enactment of this act, to do all of the following:</p> <ol style="list-style-type: none"> <li>(1) Limit Medi-Cal estate recovery to only those services required to be collected for under federal law.</li> <li>(2) Limit the definition of "estate" to include only the real and personal property and other assets required to be included within the definition of "estate" under federal law.</li> <li>(3) Require the State Department of Health Care Services to implement the option in the State Medicaid Manual to waive its claim, as a substantial hardship, when the estate, subject to recovery, is a homestead of modest value.</li> <li>(4) Prohibit recovery from the surviving spouse of a deceased Medi-Cal beneficiary.</li> <li>(5) Ensure that Medi-Cal beneficiaries can easily and timely receive information about how much their estate will owe Medi-Cal when they die.</li> </ol> <p><b>SEC. 2.</b></p> <p>Section 14009.5 of the Welfare and Institutions Code is amended to read:</p> <p><b>14009.5.</b></p> <p>(a) Notwithstanding any other provision of this chapter, the department shall claim against the estate of the decedent, or against any recipient of the property of that decedent by distribution an amount equal</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | <p>Requires such grants to be used only for the purpose of assisting such beneficiaries in obtaining information about the amount of their estate. Requires the Secretary to make such grants available to eligible individuals who are not eligible for such services</p>                                                                                                                                                                                                                                                                                          |         |   |   |   |   |   |                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | focuses on Medicare and grants, which are not relevant to the input text about Medi-Cal estate recovery in California.<br>``' |
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>l<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>-A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-B<br>i<br>I | L<br>N<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-B<br>i<br>I | <p>The people of the State of California do enact as follows:</p> <p><b>SECTION 1.</b></p> <p>Section 5205.5 of the Vehicle Code is amended to read:</p> <p><b>5205.5.</b></p> <p>(a) For purposes of implementing Section 21655.9, the department shall make available for issuance, for a fee determined by the department to be sufficient to reimburse the department for the actual costs incurred pursuant to this section, distinctive decals, labels, and other identifiers that clearly distinguish the following vehicles from other vehicles:</p> <ol style="list-style-type: none"> <li>(1) A vehicle that meets California's super ultra-low emission vehicle (SULEV) standard for exhaust emissions and the federal inherently low-emission vehicle (ILEV) evaporative emission standard, as defined in Part 88 (commencing with Section 88.101-94) of Title 40 of the Code of Federal Regulations.</li> <li>(2) A vehicle that was produced during the 2004 model-year or earlier and meets California's ultra-low emission vehicle (ULEV) standard for exhaust emissions and the federal ILEV standard.</li> <li>(3) A vehicle that meets California's enhanced advanced technology partial zero-emission vehicle (enhanced AT PZEV) standard or transitional zero-emission vehicle (TZEV) standard.</li> </ol> <p>(b) The department shall include a summary of the provisions of this section on each motor vehicle registration renewal notice, or on a separate insert, if space is available and the summary can be included without incurring additional printing or postage costs.</p> <p>(c) The Department of Transportation shall remove individual HOV lanes, or portions of those lanes, during periods of peak congestion from the access provisions provided in subdivision (a), following a finding by the Department of Transportation as follows:</p> <ol style="list-style-type: none"> <li>(1) The lane, or portion thereof, exceeds a level of service C, as discussed in subdivision (b) of Section 65089 of the Government Code.</li> <li>(2) The operation or projected operation of the vehicles described in subdivision (a) in these lanes, or portions thereof, will significantly increase congestion.</li> <li>(3) The finding shall also demonstrate the infeasibility of alleviating the congestion by other means, including, but not limited to, reducing the use of the lane by noneligible vehicles or further increasing vehicle occupancy.</li> </ol> | <p>Amends the Motor Vehicle Code to require the Secretary of Transportation (DOT) to promulgate regulations requiring the Secretary to provide for the issuance of decals and labels on motor vehicles manufactured after 2004.</p> <p>Directs the Secretary, in accordance with this Act, to:</p> <ul style="list-style-type: none"> <li>(1)(A) issue no fewer than 85,000 decals or labels for a motor vehicle manufactured before 2004; and</li> <li>(B) provide that the Secretary shall publish a list of all vehicle types identified in this Act.</li> </ul> | 2<br>.8 | 2 | 3 | 4 | 2 | 3 | The generated summary captures some elements of the input but fails to accurately represent the content. It introduces errors and irrelevant details, which affect its relevance, accuracy, and coherence. Improvements are needed to ensure the summary | - **Model Name**: The model used is shng2025/GPT-V alkyrie_LN-124m_AttnOnly_Bill sum, which appears to be a specialized model for summarizing legislative texts.<br>- **Norm Type**: LN (Layer Normalization) is used, which is a common technique in neural networks to stabilize and accelerate training.<br>- **Variant**: AttnOnly indicates that the model likely uses attention mechanisms exclusively, which should help in focusing on relevant parts of the input text.<br>- **Truncated Input**: The input text is a legislative amendment detailing the issuance of decals and labels for certain types of vehicles in California, |                                                                                                                               |

|                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                              |             |   |   |   |   |   |                                                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|---|---|---|---|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| I<br>s<br>u<br>m                                                                                                                                                            | <p>(d) The State Air Resources Board shall publish and maintain a listing of all vehicles eligible for participation in the programs described in this section. The board shall provide that listing to the department.</p> <p>(e) (1) For purposes of subdivision (a), the Department of the California Highway Patrol and the department, in consultation with the Department of Transportation, shall design and specify the placement of the decal, label, or other identifier on the vehicle. Each decal, label, or other identifier issued for a vehicle shall display a unique number, which number shall be printed on, or affixed to, the vehicle registration.</p> <p>(2) Decals, labels, or other identifiers designed pursuant to this subdivision for a vehicle described in paragraph (3) of subdivision (a) shall be distinguishable from the decals, labels, or other identifiers that are designed for vehicles described in paragraphs (1) and (2) of subdivision (a).</p> <p>(f) (1) Except as provided in paragraph (2), for purposes of paragraph (3) of subdivision (a), the department shall issue no more than</p> <p><u>85,000</u></p> <p>distinctive decals, labels, or other identifiers that clearly distinguish a vehicle specified in paragraph (3) of subdivision (a).</p> <p>(2) The department may issue a decal, label, or other identifier for a vehicle that satisfies all of the following conditions:</p> <p>(A) The vehicle is of a type identified in paragraph (3) of subdivision (a).</p> <p>(B) The owner of the vehicle is the owner of a vehicle for which a decal, label, or other identifier described in paragraph (1) was previously issued and that vehicle for which the decal, label, or other identifier was previously issued is determined by the department, on the basis of satisfactory proof submitted by the owner to the department, to be a nonrepairable vehicle or a total loss salvage vehicle.</p> <p>(C) The owner of the vehicle</p> | <p>Authorizes DOT to issue such regulations only if the Secretary determines that: (A) the vehicle meets certain standards for the emission of nitrogen oxides (NOx) and the Federal Energy Regulatory Commission (FERC) is in compliance with such standards</p>                                                                                                                            |             |   |   |   |   |   | accurately reflects the input text.                                                                                                                                                                                    | including specific standards and conditions.<br>- **Generated Summary**: The summary attempts to condense the legislative text but introduces inaccuracies and irrelevant details, such as the incorrect assignment of responsibilities and the mention of FERC.<br>``                                                                                                                                                                                                                     |
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>I<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>-<br>A<br>t<br>t<br>n | <p>The people of the State of California do enact as follows:</p> <p><b>SECTION 1.</b></p> <p>This act is known, and may be cited as, The 2024 Olympic Games and Paralympic Games Act.</p> <p><b>SEC. 2.</b></p> <p>For purposes of this act:</p> <p>(a) "Applicant committee agreement" means agreements to be entered into between the Organizing Committee for the Olympic Games (OCOG) and the United States Olympic Committee (USOC) if, and upon, the USOC's selection of the City of Los Angeles or the City and County of San Francisco as the official United States candidate city.</p> <p>(b) "Bid committee agreement" means agreements entered into between the OCOG and the USOC governing the OCOG and the bid process.</p> <p>(c) "Endorsing municipality" means the City of Los Angeles or the City and County of San Francisco which has authorized a bid by a OCOG for selection of the municipality as the site of the Olympic Games and Paralympic Games.</p> <p>(d) "Games" means the 2024 Olympic Games.</p> <p>(e) "Games support contract" means a joinder undertaking, a joinder agreement, or a similar contract executed by the Governor and containing terms permitted or required by this act.</p> <p>(f) "Joinder agreement" means an agreement entered into by:</p> <p>(1) The Governor, on behalf of this state, and a site selection organization setting out representations and assurances by the state in connection with the selection of a site in this state for the location of the games.</p> <p>(2) The endorsing municipality and a site selection organization setting out representations and assurances by the endorsing municipality in connection with the selection of a site in this state for the location of the games.</p> <p>(g) "Joinder undertaking" means an agreement entered into by:</p>                                                                                                                                                    | <p>(Sec. 3)<br/>Directs the Governor of California and the Secretary of the Treasury to: (1) enter into an agreement with the Secretary to establish a fund to be used for the construction, operation, and maintenance of Olympic facilities in California; and (2) submit to Congress a plan for the establishment and operation of such facilities.</p> <p>Authorizes appropriations.</p> | 2<br>.<br>6 | 2 | 3 | 4 | 1 | 3 | The generated summary fails to accurately represent the content of the truncated input. While it is fluent and somewhat concise, the relevance and accuracy are significantly lacking, leading to a low overall score. | - **Model Name**: The model used is shng2025/GPT-V alkyrie_LN-124m __AttnOnly_Bill sum, which appears to be a specialized model for summarizing legislative documents.<br>- **Norm Type**: LN (Layer Normalization) is a common technique used in neural networks to stabilize and accelerate training.<br>- **Variant**: AttnOnly indicates that the model likely uses attention mechanisms exclusively, which should theoretically help in focusing on relevant parts of the input text. |

|                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                     |             |   |   |   |   |   |                                                                                                                                 |                                                                                                                                                                                                                                                                                                                 |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|-------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|---|---|---|---|---|---------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| O<br>n<br>I<br>y<br>—<br>B<br>i<br>I<br>s<br>u<br>m                                                   | <p>(1) The Governor, on behalf of this state, and a site selection organization that the state will execute a joinder agreement in the event that the site selection organization selects a site in this state for the games.</p> <p>(2) The endorsing municipality and a site selection organization that the endorsing municipality will execute a joinder agreement in the event that the site selection organization selects a site in this state for the games.</p> <p>(h) "OCOG" means a nonprofit corporation, or its successor in interest, that:</p> <p>(1) Has been authorized by the endorsing municipality to pursue an application and bid on the applicant's behalf to a site selection organization for selection as the site for the games.</p> <p>(2) With the authorization of the endorsing municipality, has executed the bid committee agreement with a site selection organization regarding a bid to host the games.</p> <p>(i) "Site selection organization" means the United States Olympic Committee, the International Olympic Committee, the International Paralympic Committee, all three or some combination, as applicable.</p> <p><b>SEC. 3.</b></p> <p>The Legislature finds and declares all of the following:</p> <p>(a) The purpose of this act is to provide assurances required by a site selection organization sponsoring the games.</p> <p>(b) Hosting the games in California is expected to generate billions of dollars for the state's economy. The endorsing municipality has developed a self-sufficient bid for financing games that is based on realistic and conservative revenue scenarios and has budgeted sufficient funds to reimburse security and other service costs provided by local regional governments during the games.</p> <p>(c) The endorsing municipality plans to host an environmentally responsible games; has committed to sports and recreational opportunities for young people throughout each area by planning to generate a legacy for youth programs and other sports purposes in California with excess revenues from the games; and plans to develop and implement a unique and broad-based, statewide cultural program.</p> <p>(d) The endorsing municipality has involved athletes, sports professionals, environmentalists, business and financial experts, nonprofit organizations, youth service leaders, and individuals who represent the entire diversity of area in its bid and board of directors.</p> <p>(e) The USOC requires that all bid states, bid cities, and bid committees execute certain agreements including the joinder undertaking, which joinder undertaking must be executed on or before ____.</p> <p>(f) The endorsing municipality expects that</p> | <p>Directs the Secretary, upon request of the sponsoring municipality, to enter into a contract with the California State Olympic Committee to construct, operate, and maintain an Olympic facility in California. Requires the agreement to include assurances that the proposed facility will be constructed, operated, and maintained in a manner that is consistent with the goals of this Act and the Olympic Commission's mission. Author</p> |             |   |   |   |   |   |                                                                                                                                 |                                                                                                                                                                                                                                                                                                                 |  | - **Truncated Input**: The input provides detailed definitions and agreements related to the 2024 Olympic Games and Paralympic Games Act, focusing on the roles of various entities and the financial and environmental commitments.<br>- **Generated Summary**: The summary introduces incorrect elements such as directives to the Governor and the Secretary of the Treasury, which are not present in the input, leading to a significant misrepresentation of the original text. |
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>l<br>k<br>y<br>r<br>i<br>e | <p>A The people of the State of California do enact as follows:</p> <p>SECTION 1.</p> <p>O Article 2.6 (commencing with Section 66010.96) is added to Chapter 2 of Part 40 of Division 5 of Title 3 of the Education Code, to read:</p> <p>I Article 2.6. Office of Higher Education Performance and Accountability</p> <p>66010.96.</p> <p>(a) The Office of Higher Education Performance and Accountability is hereby established as the statewide postsecondary education coordination and planning entity. The office shall be established in state government within the Governor's office, and shall be under the direct control of an executive director.</p> <p>(b) The Governor shall appoint the Executive Director of the Office of Higher Education Performance and Accountability, who shall perform all duties, exercise all powers, assume and discharge all responsibilities, and carry out and effect all purposes vested by law in the office, including contracting for professional or consulting services in connection with the work</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | <p>Amends the Elementary and Secondary Education Act of 1965 to direct the Secretary of Education to establish within the Department of Education (ED) an Office of Excellence in Educational Research (OER).</p>                                                                                                                                                                                                                                   | 2<br>.<br>4 | 1 | 3 | 4 | 1 | 3 | The generated summary fails to accurately represent the content of the truncated input. While it is relatively well-written and | The model name "shng2025/GPT-Valkyrie_LN-124m_AttnOnly_Billsum" suggests that it is a specialized model for summarizing bills, but the generated summary does not reflect this specialization. The norm type "LN" (Layer Normalization) and variant "AttnOnly" (Attention Only) indicate specific architectural |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |

|                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                    |             |   |   |   |   |   |                          |                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|----------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------|-------------|---|---|---|---|---|--------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>B<br>i<br>I<br>I<br>s<br>u<br>m | <p>of the office. The appointment of the executive director shall be subject to confirmation by the affirmative vote of a majority of the membership of the Senate. The executive director shall appoint persons to any staff positions the Governor may authorize.</p> <p>(c) The Governor may appoint the executive director at a salary that shall be fixed pursuant to Section 12001 of the Government Code.</p> <p>(d) (1) An advisory board is hereby established for the purpose of examining and making recommendations to the office regarding the functions and operations of the office and reviewing and commenting on any recommendations made by the office to the Governor and the Legislature.</p> <p>(2) The advisory board shall consist of the Chair of the Senate Committee on Education and the Chair of the Assembly Committee on Higher Education, who shall serve as ex officio members, and six public members with experience in postsecondary education, appointed to terms of four years as follows:</p> <p>(A) Three members of the advisory board shall be appointed by the Senate Committee on Rules.</p> <p>(B) Three members of the advisory board shall be appointed by the Speaker of the Assembly.</p> <p>(3) The office shall actively seek input from, and consult with, the advisory board regarding its functions, operations and recommendations, and provide the advisory board with sufficient time to review and comment.</p> <p>(4) Advisory board meetings shall be subject to the Bagley-Keene Open Meeting Act (Article 9 (commencing with Section 11120) of Chapter 1 of Part 1 of Division 3 of Title 2 of the Government Code). Advisory board materials shall be posted on the Internet.</p> <p>(5) The advisory board shall meet at least quarterly, and shall appoint one of its members to represent the board for purposes of communicating with the Legislature.</p> <p>(6) The advisory board shall be responsible for developing an independent annual report on the condition of higher education in California.</p> <p>(7) The advisory board shall be responsible for issuing an annual review of the performance of the Executive Director of the Office of Higher Education Performance and Accountability.</p> <p>(8) Members of the advisory board shall serve without compensation, but shall receive reimbursement for actual and necessary expenses incurred in connection with the performance of their duties as board members.</p> <p>(e) The office shall consult with the higher education segments and stakeholders, as appropriate, in the conduct of its duties and responsibilities. For purposes of this subsection, higher education segments shall have the same meaning as in Section 66010.95, and higher education stakeholders shall include, but not necessarily be limited to, postsecondary faculty and students, K-12 representatives, and representatives of the business community.</p> <p><b>66010.962.</b></p> <p>The Office of Higher Education Performance and Accountability shall exist for the purpose of advising the Governor, the Legislature, and other appropriate governmental officials and institutions of postsecondary education. The office shall have the following functions and responsibilities in its capacity as the statewide postsecondary education planning and coordinating agency and advisor to the Legislature and the Governor:</p> <p>(a) It shall, through its use of information and its analytic capacity, inform the identification and periodic revision of state goals and priorities for higher education in a manner that is consistent with the goals outlined in Section 66010.91 and takes into consideration the metrics outlined in Sections 89295 and 92675. It shall, biennially, interpret and evaluate both statewide and institutional performance in relation to these goals</p> | <p>Directs the OER to:</p> <p>(1)(A) study and report to specified congressional committees on the state of California; and (2)(B) submit to the President and the Congress a report on its findings and recommendations.</p> <p>Authorizes appropriations.</p> <p>Requires the Office to report annually on its activities and responsibilities to the governor, the legislature, and the governor and the legislature.</p> <p>Authorizes appropriations for FY2011-FY2015.</p> <p>Authorises appropriations for the Office.</p> <p>Requires the Director of ED to make grants to state educational agencies (LEAs)</p> |                                                    |             |   |   |   |   |   |                          | organized, it is not relevant to the input text and contains significant inaccuracies. The summary needs to be more aligned with the actual content of the input to be useful. | choices, but these do not seem to have contributed to an accurate summary in this case. The truncated input provided detailed information about the establishment and responsibilities of the Office of Higher Education and Accountability in California, which was not captured in the generated summary. The generated summary incorrectly focused on amendments to a different act and the establishment of a different office, leading to a low relevance and accuracy score. ... |
| s<br>h<br>n<br>g                                                                                                     | L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>B<br>i<br>I<br>I<br>s<br>u<br>m                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | <p>The people of the State of California do enact as follows:</p> <p><b>SECTION 1.</b></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | (Sec. 2)<br>Amends the<br>Federal<br>Election<br>- | 2<br>.<br>4 | 1 | 3 | 4 | 1 | 3 | ** The generated summary | **<br>- **Model Name:** The model used is                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |

|                                                                                                                                                                                                           |                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                      |                                                                                                                     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| 2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>I<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>B<br>i<br>I<br>s<br>u<br>m | O<br>n<br>I<br>l<br>y<br>-<br>B<br>i<br>I<br>s<br>u<br>m | <p>Section 89003 is added to the Government Code, to read:</p> <p><b>89003.</b></p> <p>(a) A mass mailing shall not be sent within the 90 days preceding an election by or on behalf of a candidate whose name will appear on the ballot at that election for a city, county, or special district elective office.</p> <p>(b) For purposes of this section, “mass mailing” means a mass mailing, as defined by Section 82041.5, that meets the criteria of subdivision (a) of Section 18901 of Title 2 of the California Code of Regulations and, pursuant to subdivision (b) of Section 18901 of Title 2 of the California Code of Regulations, is not prohibited by Section 89001.</p> <p><b>SEC. 2.</b></p> <p>No reimbursement is required by this act pursuant to Section 6 of Article XIII B of the California Constitution because the only costs that may be incurred by a local agency or school district will be incurred because this act creates a new crime or infraction, eliminates a crime or infraction, or changes the penalty for a crime or infraction, within the meaning of Section 17556 of the Government Code, or changes the definition of a crime within the meaning of Section 6 of Article XIII B of the California Constitution.</p> <p><b>SEC. 3.</b></p> <p>The Legislature finds and declares that this bill furthers the purposes of the Political Reform Act of 1974 within the meaning of subdivision (a) of Section 81012 of the Government Code.</p> <p><b>SECTION 1.</b></p> <p>Section 14000 of the Unemployment Insurance Code is amended to read:</p> <p><b>14000.</b></p> <p>(a) The Legislature finds and declares that, in order for California to remain prosperous and globally competitive, it needs to have a well-educated and highly skilled workforce.</p> <p>(b) The Legislature finds and declares that the following principles shall guide the state's workforce investment system:</p> <p>(1) Workforce investment programs and services shall be responsive to the needs of employers, workers, and students by accomplishing the following:</p> <p>(A) Preparing California's students and workers with the skills necessary to successfully compete in the global economy.</p> <p>(B) Producing greater numbers of individuals who obtain industry-recognized certificates and degrees in competitive and emerging industry sectors and filling critical labor market skills gaps.</p> <p>(C) Adapting to rapidly changing local and regional labor markets as specific workforce skill requirements change over time.</p> <p>(D) Preparing workers for good-paying jobs that foster economic security and upward mobility.</p> <p>(2) State and local workforce investment boards are encouraged to collaborate with other public and private institutions, including businesses, unions, nonprofit organizations, kindergarten and grades 1 to 12, inclusive, career technical education programs, adult career technical education and basic skills programs, community college career technical education and basic skills programs, entrepreneurship training programs, where appropriate, the California Community Colleges Economic and Workforce Development Program, and the Employment Training Panel, to better align resources across workforce education and training service delivery systems and build a well-articulated workforce investment system by accomplishing the following:</p> <p>(A) Adopting local and regional training and education strategies that build on the strengths and fill the gaps in the education and workforce development pipeline in order to address the needs of job seekers, workers, and employers within regional labor markets by supporting sector strategies and career pathways.</p> <p>(B) Building partnerships, aligning strategies, and leveraging resources across education, social services, and workforce training delivery systems to build a career pipeline and fill critical skills gaps.</p> | <p>Campaign Act of 1971 to repeal the prohibition against mass mailing.</p> <p>Amends the Internal Revenue Code to: (1) allow a tax deduction for the cost of a tax-exempt mass mailing; and (2) increase the tax deduction to \$5,000 per year for a taxpayer who elects to receive a tax refund under this Act.</p> <p>Provides for the imposition of an excise tax on the sale or exchange of goods or services.</p> <p>Requires the Secretary of the Treasury to study and report to Congress on the feasibility and advisability of establishing a tax credit for mass mailing services.</p> <p>Requires the Secretary to report annually to the Congress on its progress in implementing this Act and to report to the</p> | <p>ry is fluent and concise but fails significantly in relevance and accuracy. It does not capture any of the main points from the input text and instead discusses unrelated federal legislation and tax codes.</p> | shng2025/GPT-V alkyrie_LN-124m __AttnOnly_Bill sum, which appears to be designed for summarizing legislative texts. |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|

|                                                                                                                                                                                                                                              |                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |         |   |   |   |   |   |                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|                                                                                                                                                                                                                                              |                                                                                        | (3)Workforce investment programs and services shall be data driven and evidence based when setting priorities, investing resources, and adopting practices.<br>(4)Workforce investment programs and services shall develop strong partnerships with the private sector, ensuring industry involvement in needs assessment, planning, and program evaluation.<br>(A)Workforce investment programs and services shall encourage industry involvement by developing strong partnerships with an industry's employers and the unions that represent the industry's workers.<br>(B)Workforce investment programs and services may consider the needs of employers and businesses of all sizes, including large, medium, small, and microenterprises, when setting                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |         |   |   |   |   |   |                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>l<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>I<br>I<br>s<br>u<br>m | L<br>N<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>I<br>I<br>s<br>u<br>m | A<br>The people of the State of California do enact as follows:<br><br>SECTION 1.<br>Article 3 (commencing with Section 115810) is added to Chapter 4 of Part 10 of Division 104 of the Health and Safety Code, to read:<br>Article 3. The Consideration of Alternatives for Artificial Turf Infill Act of 2016<br>115810.<br>The Legislature finds and declares all of the following:<br>(a) Thousands of schools, parks, and local governments have installed artificial turf fields throughout the state. It has allowed them to use fields year round, save water, and save money, among other benefits.<br>(b) Not all artificial turf fields are made from the same materials. While most artificial turf fields use less expensive crumb rubber infill from groundup used car and truck tires, many companies now offer artificial turf infill alternatives made from coconut fibers, rice husks, cork, sand, or virgin crumb rubber. Organic alternative infills can help reduce synthetic turf field temperatures on hot days by as much as 30 degrees compared to crumb rubber infill from used tires.<br>(c) The average artificial turf field uses approximately 20,000 groundup used tires to make crumb rubber infill. Tires contain many chemicals including, but not limited to: 4-t-octylphenol, acetone, arsenic, barium, benzene, benzothiazole, butylated hydroxyanisole, cadmium, carbon black, chloroethane, chromium, latex, lead, manganese, mercury, methyl ethyl ketone, methyl isobutyl ketone, n-hexadecane, naphthalene, nickel, nylon, phenol, phthalates, polycyclic aromatic hydrocarbons, and zinc.<br>(d) In 2008, then Attorney General Jerry Brown sued the nation's largest makers and installers of artificial turf fields for excessive lead levels after testing by the Center for Environmental Health found high concentrations of lead in their products.<br>(e) In 2009, the Los Angeles Unified School District banned turf fields containing infill from waste tire crumb rubber and instead chose alternative infills for their artificial turf fields.<br>(f) In 2010, then Attorney General Jerry Brown settled the case with the nation's largest makers and installers of artificial turf fields requiring them to reformulate their products to reduce lead levels and established the nation's first enforceable standards applicable to lead in artificial turf.<br>(g) The Office of Environmental Health Hazard Assessment's 2010 study on used tire crumb rubber in artificial turf fields reviewed chemical concentrations in the air above the fields and found that eight chemicals appear on the California Proposition 65 list of chemicals known to the state to cause cancer. Exposure via inhalation to five of these chemicals (benzene, formaldehyde, naphthalene, nitromethane, and styrene) gave increased lifetime cancer risks that exceeded one in one million. According to the study, the highest risk was from nitromethane, which could cause about nine cancer cases in a hypothetical population of one million soccer players. The study also found that two additional identified chemicals (toluene and benzene) appear on the | Amends the Federal Insecticide, Fungicide, and Rodenticide Act to require the Secretary of Health and Human Services (HHS) to promulgate regulations requiring the Secretary to:<br>(1) promulgate such regulations; and (2) report to Congress on the results of such regulations.<br><br>Directs the Secretary, acting through the Director of the Centers for Disease Control and Prevention (CDC), to conduct a study to determine whether there is a significant increase in the use of synthetic turf fields and whether there are significant health risks associated with exposure to such fields. | 2<br>.4 | 1 | 3 | 4 | 1 | 3 | The generated summary fails to capture the main points of the input text and introduces unrelated information. While it is concise and generally well-written, its lack of relevance and accuracy significantly detracts from its overall quality. | The model name "shng2025/GPT-Valkyrie_LN-124m_AttnOnly_Billsum" suggests a specialized model for summarizing legislative texts. The "LN" norm type indicates the use of Layer Normalization, which is a common technique in neural networks to stabilize and accelerate training. The "AttnOnly" variant implies that the model uses attention mechanisms, likely focusing on specific parts of the input text to generate the summary. However, in this case, the model failed to accurately capture the relevant details from the truncated input, leading to a summary that does not align with the provided content.<br>`` |  |

|                                                                                                                                                                                                                                              |                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |         |   |   |   |   |   |                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |  |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|                                                                                                                                                                                                                                              |                                                                                                                                     | <p>California Proposition 65 list as developmental/reproductive poisons.</p> <p>(h) At least 10 studies since 2007, including those by the United States Consumer Product Safety Commission and the United States Environmental Protection Agency, have found potentially harmful lead levels in turf fibers and in rubber crumbs.</p> <p>(i) A 2011 study titled, "An Evaluation of Potential Exposures to Lead and Other Metals as the Result of Aerosolized Particulate Matter from Artificial Turf Playing Fields" concluded that artificial turf can deteriorate to form dust containing lead at levels that may pose a risk to children.</p> <p>(j) A 2012 study published in the scientific journal Chemosphere titled, "Hazardous organic chemicals in rubber recycled tire playgrounds and pavers", showed the high content of toxic chemicals in these recycled materials and found that "uses of recycled rubber tires, especially those targeting play areas and other facilities for children, should be a matter of regulatory concern."</p> <p>(k)</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Requires the study to include a review of the scientific literature on the potential health effects of exposure to synthetic turf and whether such increased use of such fields is likely to result in a                                                                                                                                                                                                                                                                                                                                         |         |   |   |   |   |   |                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |  |
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>I<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>I<br>l<br>s<br>u<br>m | L<br>N<br>t<br>t<br>O<br>n<br>I<br>y<br>-<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>I<br>l<br>s<br>u<br>m | <p>The people of the State of California do enact as follows:</p> <p><b>SECTION 1.</b></p> <p>Section 980 of the Military and Veterans Code is amended to read: 980.</p> <p>(a) As used in this chapter, "veteran" means any of the following:</p> <p>(1) Any citizen of the United States who served in the active military, naval, or air service of the United States on or after April 6, 1917, and prior to November 12, 1918, and who received an honorable discharge or was released from active duty under honorable conditions.</p> <p>(2) Any person who did all of the following:</p> <p>(A) Served in the active military, naval, or air service of the United States for a period of not less than 90 consecutive days or was discharged from the service due to a service-connected disability within that 90-day period.</p> <p>(B) Received an honorable discharge or was released from active duty under honorable conditions.</p> <p>(C) Performed any portion of that service during any of the following periods:</p> <p>(i) On or after December 7, 1941, and prior to January 1, 1947, including, but not limited to, members of the Philippine Commonwealth Army, the Regular Scouts ("Old Scouts"), and the Special Philippine Scouts ("New Scouts").</p> <p>(ii) On or after June 27, 1950, and prior to February 1, 1955.</p> <p>(iii) On or after February 28, 1961, and prior to August 5, 1964, in the case of a veteran who served in the Republic of Vietnam during that period.</p> <p>(iv) On or after August 5, 1964, and prior to May 8, 1975.</p> <p>(v) On or after August 2, 1990, to and including the date on which the territories in and around the Arabian Peninsula cease to be designated as a place where the armed forces of the United States are engaged in combat, as described in Executive Order 12744 of the President of the United States. It is the intent of the Legislature, in enacting this clause, that the benefits provided by this chapter shall be available to all veterans who were on active duty in the armed forces of the United States or who were called to active duty in the reserves or National Guard during the pendency of the deployment of forces for Operation Desert Shield or Desert Storm, which resulted in Executive Order 12744, irrespective of whether these veterans served overseas or in the United States.</p> <p>(vi) At any time, in a campaign or expedition for service in which a medal has been authorized by the government of the United States, regardless of the number of days served on active duty.</p> <p>(vii) At any time in Somalia, or in direct support of the troops in Somalia, including, but not limited to, persons stationed on ships of the United States armed forces conducting support activities offshore in the vicinity of Somalia, during Operation Restore Hope, regardless of the number of days served.</p> | (Sec. 2) This bill amends the military and veterans code to: (1) revise provisions regarding veterans' benefits; (2) require the Department of Defense (DOD) to establish a program for veterans' health care; and (3) provide for the establishment of a Veterans Health Administration (VHA) to provide health care to veterans. The VA must establish a Veterans' Health Administration to administer the VHA. The VHA must also establish an Office of Veterans' Affairs (VA) to coordinate the VA's activities with other federal agencies. | 2<br>.4 | 1 | 3 | 4 | 1 | 3 | The generated summary fails to capture the main points of the truncated input and introduces unrelated information. While it is fluent and somewhat concise, its lack of relevance and accuracy significantly detracts from its overall quality. | - **Model Name**: The model used is shng2025/GPT-V alkyrie_LN-124m __AttnOnly_Bill sum, which appears to be a specialized model for summarizing legislative texts.<br>- **Norm Type**: LN (Layer Normalization) is a common technique used in neural networks to stabilize and accelerate training.<br>- **Variant**: AttnOnly indicates that the model likely uses attention mechanisms, which are designed to focus on the most relevant parts of the input text.<br>- **Truncated Input**: The input text is a legal document defining who qualifies as a veteran under California law, listing specific criteria and periods of service.<br>- **Generated Summary**: The summary inaccurately describes |  |

|  |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                    |  |  |  |  |  |  |  |  |  |                                                                                                                                                                                                                                                                         |
|--|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|--|--|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  |  | <p>(3) Any member of the reserves or National Guard who does all the following:</p> <p>(A) Is called to, and released from, active duty or active service, regardless of the number of days served.</p> <p>(B) Is called during any period when a presidential executive order specifies the United States is engaged in combat or homeland defense.</p> <p>(C) Has received an honorable discharge or was released from active duty or active service under honorable conditions.</p> <p>(4) Any person who did all of the following:</p> <p>(A) Served in the Merchant Marine Service of the United States.</p> <p>(B) Has been granted veteran status by the United States Secretary of Defense under Title IV of the GI Improvement Act of 1977 (Public Law 95-202, as amended).</p> <p>(5) Any person who qualifies under federal laws for revenue bond or unrestricted funds (26 U.S.C. Sec. 143) and did all of the following:</p> <p>(A) Served in the active military, naval, or air service of the United States for a period of not less than 90 consecutive days.</p> <p>(B) Received an honorable discharge or was released from active duty or active service under honorable conditions.</p> <p>(6) Any person who qualifies for funds made available from a qualified mortgage revenue bond issued pursuant to 26 U.</p> | <p>(Sec. 3) The Department of Veterans Affairs must establish the Veterans' Advisory Committee to advise the VA on the health care needs of veterans. The VA shall establish a</p> |  |  |  |  |  |  |  |  |  | amendments related to veterans' health care programs and administrative bodies, which are not mentioned in the input text. This discrepancy highlights a significant issue with the model's ability to generate relevant and accurate summaries for this type of input. |
|--|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|--|--|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                                                                                                                                                                    |                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |         |   |   |   |   |   |                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>i<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>-<br>-<br>A<br>t<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br>I<br>I<br>s | L<br>N<br>t<br>n<br>O<br>n<br>I<br>y<br>-<br>-<br>B<br>i<br> | <p>The people of the State of California do enact as follows:</p> <p><b>SECTION 1.</b></p> <p>The Legislature finds and declares all of the following:</p> <p>(a) The economic competitiveness of California is fueled by the strength of regional economies and their skilled workers. Upward social and economic mobility and increased opportunities keep the state's economy diversified and vibrant.</p> <p>(b) The pathway out of poverty for millions of California residents is the attainment of industry-valued "middle skill credentials," which is defined as a job requiring a certificate, associate's degree, or third-party credential that is less advanced than a bachelor's degree, but more advanced than a high school diploma.</p> <p>(c) Middle skill credentials serve as the gateway for a large number of careers in the state's prioritized and emergent industry sectors.</p> <p>(d) The California Community Colleges Board of Governor's Task Force on Workforce, Job Creation, and a Strong Economy, also referred to as the Strong Workforce Task Force, identified 25 policy and strategy recommendations to help close the gap on these middle skill credentials.</p> <p>(e) The recommendations built upon the foundation established by the California Community Colleges Economic and Workforce Development Program in Part 52.5 (commencing with Section 88600) of Division 7 of Title 3 of the Education Code, the Office of the Chancellor of the California Community Colleges Doing What MATTERS for Jobs and the Economy framework, and the federal Workforce Innovation and Opportunities Act (Public Law 113-128).</p> <p>(f) With the enactment of the federal Workforce Innovation and Opportunity Act (Public Law 113-128), California agencies receiving workforce-related funds have adopted the following common program strategies articulated by the California Workforce Investment Board:</p> <p>(1) Partnering in sector strategies to ensure training programs are relevant to the economy.</p> <p>(2) Building career pathways to increase access, flexibility, and facilitated navigation of training and education programs.</p> <p>(3) Utilizing "earn and learn" to increase simultaneous access to income and training for those who cannot afford full-time education.</p> <p>(4) Organizing regionally to benefit from economies of scale, recognizing gains when labor markets and industry are organized regionally.</p> | <p>Directs the Governor of California to establish a program to provide federal employers identification numbers, social security numbers, and employment development grants to eligible entities to assist them in identifying eligible entities.</p> <p>Requires the Federal Trade Commission (FTC) to report annually to the Secretary of Commerce and the Director of the National Institute of Standards and Technology (NIST) on the effectiveness of the program. Authorizes the</p> | 2<br>.4 | 1 | 3 | 4 | 1 | 3 | ** The generated summary fails to accurately capture the main points of the input text, introducing irrelevant and incorrect information. While it is relatively concise and fluent, its relevance and accuracy are significantly lacking. The coherence is acceptable but is affected by the | ** - **Model Name:** The model used is shng2025/GPT-V alkyrie_LN-124m __AttnOnly_Bill sum, which appears to be a specialized model for summarizing legislative texts. - **Norm Type:** LN (Layer Normalization) is a common technique used in neural networks to stabilize and accelerate training. - **Variant:** AttnOnly indicates that the model likely uses attention mechanisms exclusively, which should theoretically help in focusing on relevant parts of the input text. - **Truncated Input:** The input text provides detailed legislative findings and amendments related to workforce development and |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                        |         |   |   |   |   |   |                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|----------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|---|---|---|---|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| u<br>m                                                                                                                                             | (5) Providing supportive services to remove barriers to program completion and employment.<br>(6) Creating cross-system data capacity to ensure effective use of resources.<br>(7) Integrating service delivery and braiding of resources to optimize limited resources and make use of program specializations to better serve individuals.<br><b>SEC. 2.</b><br>Section 30 of the Business and Professions Code is amended to read:<br>30.<br>(a) (1) Notwithstanding any other law, any board, as defined in Section 22, and the State Bar and the Bureau of Real Estate shall, at the time of issuance of the license, require that the applicant provide its federal employer identification number, if the applicant is a partnership, or the applicant's social security number for all other applicants.<br>(2) No later than January 1, 2016, in accordance with Section 135.5, a board, as defined in Section 22, and the State Bar and the Bureau of Real Estate shall require either the individual taxpayer identification number or social security number if the applicant is an individual for purposes of this subdivision.<br>(b) A licensee failing to provide the federal employer identification number, or the individual taxpayer identification number or social security number shall be reported by the licensing board to the Franchise Tax Board. If the licensee fails to provide that information after notification pursuant to paragraph (1) of subdivision (b) of Section 19528 of the Revenue and Taxation Code, the licensee shall be subject to the penalty provided in paragraph (2) of subdivision (b) of Section 19528 of the Revenue and Taxation Code.<br>(c) In addition to the penalty specified in subdivision (b), a licensing board shall not process an application for an initial license unless the applicant provides its federal employer identification number, or individual taxpayer identification number or social security number where requested on the application.<br>(d) A licensing board shall, upon request of the Franchise Tax Board or the Employment Development Department, furnish to the board or the department, as applicable, | Secretary to award competitive grants to nonprofit organizations to assist such organizations in identifying individuals who are eligible for such grants. Requires such grants to be used for:<br>1) training programs;<br>2) job creation and job creation;<br>3) workforce development ; and<br>4) education and training programs. Requires the Secretary, in conjunction with the |         |   |   |   |   |   | inaccuracies.                                                                                                                                                                     | identification number requirements.<br>- **Generated Summary:** The summary generated does not align well with the input, introducing incorrect elements and failing to capture the key points of the legislative text. This indicates potential issues with the model's ability to accurately interpret and summarize legislative content.                                                                                              |
| s<br>h<br>n<br>g<br>2<br>0<br>2<br>5<br>/<br>G<br>P<br>T<br>-<br>V<br>a<br>l<br>k<br>y<br>r<br>i<br>e<br>-<br>L<br>N<br>-<br>1<br>2<br>4<br>m<br>- | A<br>t<br>t<br>o<br>n<br>I<br>y<br><br>The people of the State of California do enact as follows:<br><br><b>SECTION 1.</b><br>Section 17052 is added to the Revenue and Taxation Code, to read:<br>17052.<br>(a) (1) For each taxable year beginning on or after January 1, 2015, there shall be allowed against the "net tax," as defined by Section 17039, an earned income tax credit in an amount equal to an amount determined in accordance with Section 32 of the Internal Revenue Code, relating to earned income, as applicable for federal income tax purposes for the taxable year, except as otherwise provided in this section.<br>(2) (A) The amount of the credit determined under Section 32 of the Internal Revenue Code, relating to earned income, as modified by this section, shall be multiplied by the earned income tax credit adjustment factor for the taxable year.<br>(B) Unless otherwise specified in the annual Budget Act, the earned income tax credit adjustment factor for a taxable year beginning on or after January 1, 2015, shall be 0 percent.<br>(C) The earned income tax credit authorized by this section shall only be operative for taxable years for which resources are authorized in the annual Budget Act for the Franchise Tax Board to oversee and audit returns associated with the credit.<br>(b) (1) In lieu of the table prescribed in Section 32(b)(1) of the Internal Revenue Code, relating to percentages, the credit percentage and the phaseout percentage shall be determined as follows:<br>In the case of an eligible individual with:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | Amends the Revenue Code to allow a tax credit of up to \$3,300 for earned income taxes.<br><br>Allows an individual with a credit percentage of less than 25% of earned income to elect to receive a refundable tax credit under this Act. Defines "qualified earned income"; as any earned income                                                                                     | 2<br>.8 | 2 | 3 | 4 | 2 | 3 | The generated summary partially captures the essence of the tax credit but is marred by the inclusion of unrelated and inaccurate information. It needs to be more focused on the | The model name "shng2025/GPT-Valkyrie_LN-124m_AttnOnly_Billsum" suggests a specialized model for summarizing legislative bills. The "Norm Type" being LN (Layer Normalization) and the "Variant" being AttnOnly (Attention Only) indicate specific architectural choices aimed at improving performance. However, the generated summary does not fully leverage these features, as it introduces inaccuracies and unrelated content. The |

|                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                   |  |  |  |  |  |                                                   |                                                                                                                                                                                                                                           |
|----------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|---------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>-<br/>A<br/>t<br/>t<br/>n<br/>O<br/>n<br/>I<br/>y<br/>-<br/>-<br/>B<br/>i<br/>I<br/>I<br/>s<br/>u<br/>m</p> | <p>The credit percentage is:<br/>The phaseout percentage is:<br/>No qualifying children<br/>7.65%<br/>7.65%<br/>1 qualifying child<br/>34%<br/>34%<br/>2 or more qualifying children<br/>40%<br/>40%</p> <p>(2) (A) In lieu of the table prescribed in Section 32(b)(2)(A) of the Internal Revenue Code, the earned income amount and the phaseout amount shall be determined as follows:</p> <p>In the case of an eligible individual with:</p> <p>The earned income amount is:<br/>The phaseout amount is:<br/>No qualifying children<br/>\$3,290<br/>\$3,290<br/>1 qualifying child<br/>\$4,940<br/>\$4,940<br/>2 or more qualifying children<br/>\$6,935<br/>\$6,935</p> <p>(B) Section 32(b)(2)(B) of the Internal Revenue Code, relating to joint returns, shall not apply.</p> <p>(3) Section 32(b)(3)(A) of the Internal Revenue Code, relating to increased percentage for three or more qualifying children, is modified by substituting "the credit percentage and phaseout percentage is 45 percent" for "the credit percentage is 45 percent."</p> <p>(c) (1) Section 32(c)(1)(A)(ii)(I) of the Internal Revenue Code is modified by substituting "this state" for "the United States."</p> <p>(2) Section 32(c)(2)(A) of the Internal Revenue Code is modified as follows:</p> <p>(A) Section 32(c)(2)(A)(i) of the Internal Revenue Code is modified by deleting "plus" and inserting in lieu thereof the following: "and only if such amounts are subject to withholding pursuant to Division 6 (commencing with Section 13000) of the Unemployment Insurance Code."</p> <p>(B) Section 32(c)(2)(A)(ii) of the Internal Revenue Code shall not apply.</p> <p>(3) Section 32(c)(3)(C) of the Internal Revenue Code, relating to place of abode, is modified by substituting "this state" for "the United States."</p> <p>(d) Section 32(i)(1) of the Internal Revenue Code is modified by substituting "\$3,400" for "\$2,200."</p> <p>(e) In lieu of Section 32(j) of the Internal Revenue Code, relating to inflation adjustments, for taxable years beginning on or after January 1, 2016, the amounts specified in paragraph (2) of subdivision (b) and in subdivision (d) shall be recomputed annually in</p> | <p>earned by an individual who is not a qualifying child (as defined by this Act).</p> <p>Amends title II (Old Age, Survivors and Disability Insurance) (OASDI) of the Social Security Act (SSA) to provide for the payment of OASDI benefits to individuals who have attained age 65 or older (currently, age 65 and older) and who are not eligible for OAS</p> |  |  |  |  |  | specific details provided in the truncated input. | truncated input focuses on amendments to the California Revenue and Taxation Code regarding earned income tax credits, but the summary diverges into irrelevant areas, indicating a need for better alignment with the input text.<br>... |
|----------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|---------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## Appendix K - Question Answering GPT-4o CSV Files

Note: We are displaying only 20 out of the 200 prompts, as including the full CSV file in this document would extend it to over 600 pages. This large volume of data highlights the effectiveness of GPT-4o for evaluation purposes. The complete CSV files are available in our GitHub repository.<sup>68</sup>

| model_name                                   | norm_type | variant  | question                             | context                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | reference_answers                        | generated_answer | Ovverrare | Coorerat | Coomerat | Roverat | Floverat | Conecise | Overall Feedback                                                                                                                                                                                        | Comments on Columns                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|----------------------------------------------|-----------|----------|--------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------|------------------|-----------|----------|----------|---------|----------|----------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD | LN        | AttnOnly | In what country is Normandy located? | The Normans (Norman: Nourmands; French: Normands; Latin: Normanni) were the people who in the 10th and 11th centuries gave their name to Normandy, a region in France. They were descended from Norse ("Norman" comes from "Norseman") raiders and pirates from Denmark, Iceland and Norway who, under their leader Rollo, agreed to swear fealty to King Charles III of West Francia. Through generations of assimilation and mixing with the native Frankish and Roman-Gaulish populations, their descendants would gradually merge with the Carolingian-based cultures of West Francia. The distinct cultural and ethnic identity of the Normans emerged initially in the first half of the 10th century, and it continued to evolve | ['France', 'France', 'France', 'France'] | France           | 5.0       | 5        | 5        | 5       | 5        | 5        | The generated answer is perfect in terms of correctness, completeness, relevance, fluency, and conciseness. It directly answers the question with the appropriate information derived from the context. | <ul style="list-style-type: none"> <li>**Model Name**: The model name "shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD" suggests that it is a specialized model fine-tuned on the SQuAD dataset, which is known for question-answering tasks. The "AttnOnly" variant indicates that it might be using an attention-only mechanism, which can be effective for focusing on relevant parts of the context.</li> <li>**Norm Type**: "LN" likely stands for Layer Normalization, a technique used to stabilize and improve the training of deep neural networks.</li> <li>**Variant**: "AttnOnly" suggests that the model uses an attention-only mechanism, which can help in accurately pinpointing the relevant information in the context.</li> <li>**Question**: The question "In what country is Normandy located?" is straightforward and seeks a specific factual answer.</li> <li>**Context**: The context provides detailed historical and geographical information about Normandy, clearly stating that it is a region in France.</li> <li>**Reference Answers**: The reference answers are all "France," indicating a consensus on the correct answer.</li> <li>**Generated Answer**: The generated answer "France" is correct and matches the reference answers, demonstrating the model's effectiveness in this instance.</li> </ul> |

<sup>68</sup> <https://github.com/Ice-Citron/Extended-Essay-Appendix.git> by (Student, 2024)

|                                              |     |           |                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                       |     |             |                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |  |  |  |  |  |
|----------------------------------------------|-----|-----------|-----------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|
|                                              |     |           |                                               | over the succeeding centuries.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                       |     |             |                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |  |  |  |  |  |
| shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD | L N | Attn Only | When were the Normans in Normandy?            | The Normans (Norman: 'Normands'; French: 'Normands'; Latin: 'Normanni') were the people who in the 10th and 11th centuries gave their name to Normandy, a region in France. They were descended from Norse ("Norman" comes from "Norseman") raiders and pirates from Denmark, Iceland and Norway who, under their leader Rollo, agreed to swear fealty to King Charles III of West Francia. Through generations of assimilation and mixing with the native Frankish and Roman-Gaulish populations, their descendants would gradually merge with the Carolingian-based cultures of West Francia. The distinct cultural and ethnic identity of the Normans emerged initially in the first half of the 10th century, and it continued to evolve over the succeeding centuries. | ['10th and 11th centuries', 'in the 10th and 11th centuries', '10th and 11th centuries', '10th and 11th centuries']                                                                                                                                      | 10th and 11th centuries                                                                                                                                                                                                                                                                                                                                                                                                                               | 5 0 | 5 5 5 5 5 5 | The generated answer is accurate, complete, relevant, fluent, and concise. It perfectly matches the reference answers and addresses the question based on the provided context. | - **Model Name**: The model used is shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD, which appears to be a specialized model for question-answering tasks.<br>- **Norm Type**: LN likely stands for Layer Normalization, a technique used in neural networks to stabilize and accelerate training.<br>- **Variant**: AttnOnly suggests that the model uses attention mechanisms exclusively, which can help in focusing on relevant parts of the context to generate accurate answers.<br>- **Question**: The question is clear and specific, asking for the time period when the Normans were in Normandy.<br>- **Context**: The context provides detailed historical information about the Normans, including their origins and the time period they were in Normandy.<br>- **Reference Answers**: The reference answers are consistent and provide a clear benchmark for evaluating the generated answer.<br>- **Generated Answer**: The generated answer "10th and 11th centuries" is correct and matches the reference answers, demonstrating the model's effectiveness in this task.<br>... |  |  |  |  |  |  |
| shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD | L N | Attn Only | From which countries did the Norse originate? | The Normans (Norman: 'Normands'; French: 'Normands'; Latin: 'Normanni') were the people who in the 10th and 11th centuries gave their name to Normandy, a region in France. They were descended from Norse ("Norman" comes from "Norseman") raiders and pirates from Denmark, Iceland and Norway who, under their leader Rollo, agreed to swear fealty to King Charles III of West Francia. Through                                                                                                                                                                                                                                                                                                                                                                         | ['Denmark, Iceland and Norway', 'Denmark, Iceland and Norway'] | From which countries did the Norse originate? The?The Normans (Norman: 'Normands'; French: 'Normands'; Latin: 'Normanni') were the people who in the 10th and 11th centuries gave their name to Normandy, a region in France. They were descended from Norse ("Norman" comes from "Norseman") raiders and pirates from Denmark, Iceland and Norway who, under their leader Rollo, agreed to swear fealty to King Charles III of West Francia. Through | 4 0 | 5 4 4 4 4 3 | The generated answer is correct and mostly complete but could be more concise and slightly more fluent. It provides the necessary information but                               | - **Model Name**: The model name indicates that this is a specialized version of GPT, likely fine-tuned for question-answering tasks using the SQuAD dataset. The "AttnOnly" variant suggests it uses attention mechanisms, which are crucial for understanding context in QA tasks.<br>- **Norm Type**: "LN" stands for Layer Normalization, a technique used to stabilize and improve the training of deep neural networks.<br>- **Variant**: "AttnOnly" indicates that the model relies solely on attention mechanisms, which are effective for tasks requiring context comprehension.<br>- **Question**: The question is straightforward, asking for the countries of origin of the Norse.<br>- **Context**: The context provides detailed historical information about the Normans and their Norse origins,                                                                                                                                                                                                                                                                               |  |  |  |  |  |  |

|  |  |  |                                                                                                                                                                                                                                                                                                                                                                         |         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |  |  |  |  |  |  |                                                                            |                                                                                                                                                                                                                                                                                                                                                            |
|--|--|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|----------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  |  |  | generations of assimilation and mixing with the native Frankish and Roman-Gaulish populations, their descendants would gradually merge with the Carolingian-based cultures of West Francia. The distinct cultural and ethnic identity of the Normans emerged initially in the first half of the 10th century, and it continued to evolve over the succeeding centuries. | Norway] | ur<br>ma<br>nds<br>;<br>Fre<br>nch<br>:<br>Nor<br>ma<br>nds<br>;<br>Lat<br>in:<br>Nor<br>ma<br>nni<br>)<br>wer<br>e<br>the<br>peo<br>ple<br>wh<br>o<br>in<br>the<br>10t<br>h<br>and<br>11t<br>h<br>cen<br>turi<br>es<br>gav<br>e<br>thei<br>r<br>na<br>me<br>to<br>Nor<br>ma<br>ndy<br>,<br>a<br>re<br>gi<br>on<br>in<br>Fra<br>nce<br>. The<br>y<br>wer<br>e<br>des<br>cen<br>ded<br>fro<br>m<br>Nor<br>se<br>("N<br>or<br>ma<br>n"<br>co<br>me<br>s<br>fro<br>m<br>"N |  |  |  |  |  |  | includes additional context that is not directly relevant to the question. | specifically mentioning Denmark, Iceland, and Norway.<br>- **Reference Answers**: The reference answers are consistent and concise, listing the countries of origin as Denmark, Iceland, and Norway.<br>- **Generated Answer**: The generated answer is correct but includes unnecessary details, making it less concise and slightly less relevant.<br>`` |
|--|--|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|----------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                              |        |                          |                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                      |          |         |   |   |   |   |   |   |                                                                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |  |  |  |  |
|----------------------------------------------|--------|--------------------------|-----------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|----------|---------|---|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|
|                                              |        |                          |                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                      |          |         |   |   |   |   |   |   |                                                                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |  |  |  |  |
| shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD | L<br>N | A<br>ttn<br>O<br>nl<br>y | W<br>ho<br>w<br>as<br>t<br>h<br>e<br>N<br>o<br>r<br>s<br>e<br>l<br>e<br>a<br>d<br>e<br>r<br>? | The Normans (Norman: Nourmands; French: Normands; Latin: Normanni) were the people who in the 10th and 11th centuries gave their name to Normandy, a region in France. They were descended from Norse ("Norman" comes from "Norseman") raiders and pirates from Denmark, Iceland and Norway who, under their leader Rollo, agreed to swear fealty to King Charles III of West Francia. Through generations of assimilation and mixing with the native Frankish and Roman-Gaulish populations, their descendants would gradually merge with the Carolingian-based cultures of West Francia. The distinct cultural and ethnic identity of the Normans emerged initially in the first half of the 10th century, and it continued to evolve over the succeeding centuries. | ['Rollo', 'Rollo', 'Rollo', 'Rollo'] | Rollo    | 5<br>.0 | 5 | 5 | 5 | 5 | 5 | 5 | The generated answer is excellent in all aspects. It is correct, complete, relevant, and concise. The model has performed exceptionally well in providing a precise and accurate response. | - **Model Name**: The model name "shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD" indicates that it is a specialized version of GPT, likely fine-tuned on the SQuAD dataset with an attention-only mechanism and layer normalization (LN).<br>- **Norm Type**: "LN" stands for Layer Normalization, a technique used to stabilize and accelerate the training of deep neural networks.<br>- **Variant**: "AttnOnly" suggests that the model uses an attention-only mechanism, which might focus on improving the attention layers' performance.<br>- **Question**: The question "Who was the Norse leader?" is straightforward and seeks a specific name.<br>- **Context**: The context provides detailed historical information about the Normans and clearly mentions Rollo as the Norse leader.<br>- **Reference Answers**: The reference answers are all "Rollo," indicating a consensus on the correct answer.<br>- **Generated Answer**: The generated answer "Rollo" matches the reference answers perfectly, demonstrating the model's accuracy and reliability.<br>--- |  |  |  |  |
| shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD | L<br>N | A<br>ttn<br>O            | W<br>h<br>a<br>t<br>c<br>e<br>n<br>t<br>u<br>r<br>y<br>d<br>i<br>d                            | The Normans (Norman: Nourmands; French: '10th century, 'the                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | [10th century, 'the                  | 10th cen | 4<br>.8 | 5 | 4 | 5 | 5 | 5 | 5 | The generated                                                                                                                                                                              | - **Model Name**: The model name "shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD" indicates that this is a                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |  |  |  |  |

|                                                                                |         |                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |          |         |             |   |   |   |   |                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|--------------------------------------------------------------------------------|---------|-------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|---------|-------------|---|---|---|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| kyrie_LN-12<br>4m__<br>AttnO<br>nly__<br>SQuA<br>D                             | nl<br>y | the<br>Norm<br>ans<br>first<br>gain<br>their<br>separ<br>ate<br>identi<br>ty? | Normands; Latin:<br>Normanni) were the<br>people who in the<br>10th and 11th<br>centuries gave their<br>name to Normandy,<br>a region in France.<br>They were<br>descended from<br>Norse ("Norman"<br>comes from<br>"Norseman") raiders<br>and pirates from<br>Denmark, Iceland<br>and Norway who,<br>under their leader<br>Rollo, agreed to<br>swear fealty to King<br>Charles III of West<br>Francia. Through<br>generations of<br>assimilation and<br>mixing with the<br>native Frankish and<br>Roman-Gaulish<br>populations, their<br>descendants would<br>gradually merge with the<br>Carolingian-based<br>cultures of West<br>Francia. The distinct<br>cultural and ethnic<br>identity of the<br>Normans emerged<br>initially in the first<br>half of the 10th<br>century, and it<br>continued to evolve<br>over the succeeding<br>centuries. | first<br>half of<br>the<br>10th<br>centur<br>y',<br>'10th',<br>'10th']                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | tur<br>y |         |             |   |   |   |   | answer<br>is<br>highly<br>accurat<br>e,<br>relevant<br>, and<br>well-con<br>structed<br>. It<br>correctl<br>y identifie<br>s the<br>10th<br>century<br>as the<br>period<br>when<br>the<br>Norman<br>s first<br>gained<br>their<br>separat<br>e identity.<br>While it<br>could<br>be<br>slightly<br>more<br>specific,<br>it still<br>effectiv<br>ely<br>answers<br>the<br>questio<br>n. | specialized model likely fine-tuned on the SQuAD dataset, which is known for question-answering tasks. The model's performance in this instance is strong, suggesting effective training and fine-tuning.<br>- **Norm Type**: "LN" stands for Layer Normalization, a technique used to stabilize and improve the training of deep neural networks. This likely contributes to the model's ability to generate accurate and fluent answers.<br>- **Variant**: "AttnOnly" suggests that the model uses attention mechanisms exclusively, which are known to enhance the model's focus on relevant parts of the context when generating answers. This is evident in the model's precise and relevant response.<br>- **Question**: The question is clear and specific, asking for the century when the Normans first gained their separate identity. This clarity helps in evaluating the model's performance.<br>- **Context**: The context provides detailed historical information about the Normans, including their origins and the period when they developed a distinct identity. This rich context allows the model to generate an accurate and relevant answer.<br>- **Reference Answers**: The reference answers are consistent and provide a clear benchmark for evaluating the generated answer. The generated answer aligns well with these references, indicating its correctness.<br>... |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| shng2<br>025/G<br>PT-Val<br>kyrie_LN-12<br>4m__<br>AttnO<br>nly__<br>SQuA<br>D | L<br>N  | At<br>tn<br>O<br>nl<br>y                                                      | Who<br>gave<br>their<br>name<br>to<br>Norm<br>andy<br>in the<br>1000'<br>s and<br>1100'<br>s                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | The Normans<br>(Norman:<br>Normands; French:<br>Normands; Latin:<br>Normanni) were the<br>people who in the<br>10th and 11th<br>centuries gave their<br>name to Normandy,<br>a region in France.<br>They were<br>descended from<br>Norse ("Norman"<br>comes from<br>"Norseman") raiders<br>and pirates from<br>Denmark, Iceland<br>and Norway who,<br>under their leader<br>Rollo, agreed to<br>swear fealty to King<br>Charles III of West<br>Francia. Through<br>generations of<br>assimilation and<br>mixing with the<br>native Frankish and<br>Roman-Gaulish<br>populations, their | □        | Wh<br>o | 1<br>.<br>0 | 1 | 1 | 1 | 1 | 1                                                                                                                                                                                                                                                                                                                                                                                      | The<br>generat<br>ed<br>answer<br>is<br>entirely<br>inadequ<br>ate. It<br>does<br>not<br>provide<br>any<br>useful<br>informati<br>on,<br>fails to<br>address<br>the<br>questio<br>n, and<br>is not<br>well-for<br>med.<br>The<br>model<br>needs<br>signific<br>ant<br>improve                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | - **Model Name**: The model name indicates that it is a variant of GPT-Valkyrie with attention-only mechanisms and layer normalization, fine-tuned on the SQuAD dataset. This suggests it should be capable of handling question-answering tasks, but the performance in this instance is poor.<br>- **Norm Type**: LN stands for Layer Normalization, a technique used to stabilize and improve the training of deep neural networks. Despite this, the model's output in this case is not satisfactory.<br>- **Variant**: AttnOnly indicates that the model uses attention mechanisms exclusively. Attention mechanisms are designed to help the model focus on relevant parts of the input, but this did not translate into a correct or relevant answer here.<br>- **Question**: The question is clear and asks for the identity of the people who gave their name to Normandy in the 10th and 11th centuries.<br>- **Context**: The context provides detailed information about the Normans, their origins, and their leader Rollo. This information should have been used to generate a correct and complete answer. |

|                                                                                    |        |                          |                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |   |                                                                          |                            |                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |  |  |                                                                      |                                                                                                                                                                                                                                                                          |
|------------------------------------------------------------------------------------|--------|--------------------------|-----------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|--------------------------------------------------------------------------|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|----------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                    |        |                          |                                                                 | descendants would gradually merge with the Carolingian-based cultures of West Francia. The distinct cultural and ethnic identity of the Normans emerged initially in the first half of the 10th century, and it continued to evolve over the succeeding centuries.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |   |                                                                          |                            |                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |  |  | ment in understanding and generating relevant and complete answers . | - **Reference Answers**: There are no reference answers provided, but the context itself is sufficient to derive the correct answer.<br>- **Generated Answer**: The generated answer "Who" is not useful and does not address the question in any meaningful way.<br>``` |
| shng2<br>025/G<br>PT-Val<br>kyrie_<br>LN-12<br>4m__<br>AttnO<br>nly__<br>SQuA<br>D | L<br>N | At<br>tn<br>O<br>nl<br>y | What<br>is<br>Franc<br>e a<br>regio<br>n of?                    | The Normans (Norman: Nourmands; French: Normands; Latin: Normanni) were the people who in the 10th and 11th centuries gave their name to Normandy, a region in France. They were descended from Norse ("Norman" comes from "Norseman") raiders and pirates from Denmark, Iceland and Norway who, under their leader Rollo, agreed to swear fealty to King Charles III of West Francia. Through generations of assimilation and mixing with the native Frankish and Roman-Gaulish populations, their descendants would gradually merge with the Carolingian-based cultures of West Francia. The distinct cultural and ethnic identity of the Normans emerged initially in the first half of the 10th century, and it continued to evolve over the succeeding centuries. | □ | Wh<br>at<br>.0                                                           | 1<br>1<br>1<br>1<br>1<br>1 | The generat<br>ed<br>answer<br>is<br>entirely<br>inadequ<br>ate. It<br>does<br>not<br>address<br>the<br>questio<br>n, lacks<br>relevan<br>ce, and<br>is not<br>well-for<br>med.<br>The<br>model<br>needs<br>signific<br>ant<br>improve<br>ment in<br>underst<br>anding<br>and<br>generati<br>ng<br>context<br>ually<br>appror<br>iate<br>answers<br>. | - **Model Name**: The model name indicates that it is a variant of GPT-Valkyrie with attention-only mechanisms, fine-tuned on the SQuAD dataset. This suggests it should be capable of handling question-answering tasks, but the performance in this instance is poor.<br>- **Norm Type**: LN (Layer Normalization) is a common technique used in neural networks to stabilize and accelerate training. However, the normalization technique does not seem to have helped in generating a correct answer here.<br>- **Variant**: AttnOnly indicates that the model uses attention mechanisms exclusively. Attention mechanisms are designed to focus on relevant parts of the input, but in this case, the model failed to focus on the relevant context.<br>- **Question**: The question "What is France a region of?" is straightforward and expects an answer indicating that France is a country, not a region.<br>- **Context**: The context provides information about Normandy being a region in France, which should help the model understand that France is not a region but a country.<br>- **Reference Answers**: The absence of reference answers means there is no benchmark for comparison, but the generated answer is clearly inadequate.<br>- **Generated Answer**: The answer "What" is nonsensical in this context and does not provide any useful information.<br><br>Overall, the model's performance in this instance is unsatisfactory, and it requires significant improvements to handle such questions effectively.<br>``` |  |  |                                                                      |                                                                                                                                                                                                                                                                          |
| shng2<br>025/G<br>PT-Val<br>kyrie_<br>LN-12<br>4m__<br>AttnO<br>nly__<br>SQuA<br>D | L<br>N | At<br>tn<br>O<br>nl<br>y | Who<br>did<br>King<br>Char<br>les III<br>swear<br>fealty<br>to? | The Normans (Norman: Nourmands; French: Normands; Latin: Normanni) were the people who in the 10th and 11th centuries gave their name to Normandy, a region in France.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | □ | Wh<br>o<br>did<br>Kin<br>g<br>Ch<br>arle<br>s III<br>sw<br>ear<br>2<br>2 | 2<br>1<br>2<br>2<br>3<br>3 | The generat<br>ed<br>answer<br>fails to<br>correctl<br>y<br>address<br>the<br>questio<br>n                                                                                                                                                                                                                                                            | - **Model Name**: The model name indicates that it is a specialized version of GPT designed for question-answering tasks, specifically trained on the SQuAD dataset.<br>- **Norm Type**: LN stands for Layer Normalization, a technique used to stabilize and improve the training of deep neural networks.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |  |  |                                                                      |                                                                                                                                                                                                                                                                          |

|  |  |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |  |                                                                                                                                                                                                                  |  |  |  |  |  |  |                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |  |  |  |  |  |  |  |  |  |
|--|--|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|--|--|--|
|  |  |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |  |                                                                                                                                                                                                                  |  |  |  |  |  |  |                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |  |  |  |  |  |  |  |  |  |
|  |  |  | They were descended from Norse ("Norman" comes from "Norseman") raiders and pirates from Denmark, Iceland and Norway who, under their leader Rollo, agreed to swear fealty to King Charles III of West Francia. Through generations of assimilation and mixing with the native Frankish and Roman-Gaulish populations, their descendants would gradually merge with the Carolingian-based cultures of West Francia. The distinct cultural and ethnic identity of the Normans emerged initially in the first half of the 10th century, and it continued to evolve over the succeeding centuries. |  | fealty to? The Normans (Norman: No ur mands; French: Normands; Latin: Normanni) were the people who in the 10th and 11th centuries gave their name to Normandy, a region in France. They were descended from Nor |  |  |  |  |  |  | n, leading to a low score in correctness and relevance. While the answer is somewhat fluent and concise, it does not provide the necessary information to fully answer the question. | - <b>Variant</b> : AttnOnly suggests that the model uses an attention-only mechanism, which focuses on the most relevant parts of the input context to generate the answer.<br>- <b>Question</b> : The question is straightforward and asks for a specific piece of information.<br>- <b>Context</b> : The context provides detailed historical information about the Normans and their relationship with King Charles III of West Francia.<br>- <b>Reference Answers</b> : No reference answers are provided, which makes it challenging to compare the generated answer against a gold standard.<br>- <b>Generated Answer</b> : The generated answer misinterprets the question, leading to an incorrect and incomplete response.<br>... |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|

|                                        |        |                          |                                                                |                                                                                                                                  |   |                                                                                                                                                                                     |                                                       |         |   |   |   |   |   |                                            |                                                                                                                                                                                                                                                                              |
|----------------------------------------|--------|--------------------------|----------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------|---------|---|---|---|---|---|--------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| shng2025/GPT-Valkyrie_LN-124m_AttnOnly | L<br>N | A<br>ttn<br>O<br>nl<br>y | W<br>hen<br>did<br>the<br>Fr<br>ank<br>ish<br>iden<br>ti<br>ty | The Normans (Norman: Nourmands; French: Normands; Latin: Normanni) were the people who in the 10th and 11th centuries gave their | □ | se ("Nor man" comes from "Nor mans") raiders and pirates from Denmark, Iceland and Norway who, under their leader Rollo, agreed to swear fealty to King Charles III of West Francia | firs<br>t<br><br>half<br>of<br>the<br>10t<br>h<br>cen | 2<br>.4 | 1 | 1 | 1 | 4 | 5 | The generated answer is concise and fluent | - **Model Name**: The model used is shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD, which is designed for question-answering tasks.<br>- **Norm Type**: LN (Layer Normalization) is used, which is a common technique in neural networks to stabilize and accelerate training. |
|----------------------------------------|--------|--------------------------|----------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------|---------|---|---|---|---|---|--------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                              |        |              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                             |                   |   |   |   |   |   |   |                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|----------------------------------------------|--------|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|-------------------|---|---|---|---|---|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SQuAD                                        |        | emerge?      | name to Normandy, a region in France. They were descended from Norse ("Norman" comes from "Norseman") raiders and pirates from Denmark, Iceland and Norway who, under their leader Rollo, agreed to swear fealty to King Charles III of West Francia. Through generations of assimilation and mixing with the native Frankish and Roman-Gaulish populations, their descendants would gradually merge with the Carolingian-based cultures of West Francia. The distinct cultural and ethnic identity of the Normans emerged initially in the first half of the 10th century, and it continued to evolve over the succeeding centuries. | tur<br>y                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                             |                   |   |   |   |   |   |   | but fails in correctness, completeness, and relevance. It provides an incorrect timeframe for the emergence of the Frankish identity by mistakenly referring to the Norman identity instead. | - **Variant**: AttnOnly indicates that the model uses attention mechanisms, which are crucial for understanding context in question-answering tasks.<br>- **Question**: The question asks about the emergence of the Frankish identity, which is not addressed in the provided context.<br>- **Context**: The context discusses the emergence of the Norman identity, not the Frankish identity, leading to the model's incorrect answer.<br>- **Reference Answers**: There are no reference answers provided, which makes it challenging to compare the generated answer against a correct one.<br>- **Generated Answer**: The generated answer "first half of the 10th century" is incorrect as it refers to the Norman identity, not the Frankish identity.<br>...                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD | L<br>N | Attn<br>Only | Who was the duke in the battle of Hastings?                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | The Norman dynasty had a major political, cultural and military impact on medieval Europe and even the Near East. The Normans were famed for their martial spirit and eventually for their Christian piety, becoming exponents of the Catholic orthodoxy into which they assimilated. They adopted the Gallo-Romance language of the Frankish land they settled, their dialect becoming known as Norman, Normaund or Norman French, an important literary language. The Duchy of Normandy, which they formed by treaty with the French crown, was a great fief of medieval France, and under Richard I of Normandy was | ['William the Conqueror', 'William the Conqueror', 'William the Conqueror'] | Will<br>iam<br>.0 | 5 | 5 | 5 | 5 | 5 | 5 | The generated answer is accurate, complete, relevant, fluent, and concise. It meets all the criteria for a high-quality response.                                                            | - **Model Name**: The model name "shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD" suggests that it is a specialized model fine-tuned on the SQuAD dataset, which is known for question-answering tasks. The "AttnOnly" variant indicates that it might be using an attention-only mechanism, which can be effective for focusing on relevant parts of the context.<br>- **Norm Type**: "LN" likely stands for Layer Normalization, a technique used to stabilize and improve the training of deep neural networks.<br>- **Variant**: "AttnOnly" suggests that the model uses an attention-only mechanism, which can help in accurately pinpointing the relevant information in the context.<br>- **Question**: The question is clear and specific, asking for the name of the duke involved in the Battle of Hastings.<br>- **Context**: The context is detailed and provides sufficient information to answer the question accurately. It mentions the Norman conquest of England and specifically names William the Conqueror as the leader.<br>- **Reference Answers**: The reference answers are consistent and all point to "William the Conqueror," which aligns with the generated answer.<br>- **Generated Answer**: The generated answer "William the Conqueror" is correct and matches the reference answers, |

|  |  |  |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |  |  |  |  |  |  |  |  |  |                                                                  |
|--|--|--|--|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|--|--|--|------------------------------------------------------------------|
|  |  |  |  | forged into a cohesive and formidable principality in feudal tenure. The Normans are noted both for their culture, such as their unique Romanesque architecture and musical traditions, and for their significant military accomplishments and innovations. Norman adventurers founded the Kingdom of Sicily under Roger II after conquering southern Italy on the Saracens and Byzantines, and an expedition on behalf of their duke, William the Conqueror, led to the Norman conquest of England at the Battle of Hastings in 1066. Norman cultural and military influence spread from these new European centres to the Crusader states of the Near East, where their prince Bohemond I founded the Principality of Antioch in the Levant, to Scotland and Wales in Great Britain, to Ireland, and to the coasts of north Africa and the Canary Islands. |  |  |  |  |  |  |  |  |  | demonstrating the model's effectiveness in this instance.<br>... |
|--|--|--|--|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|--|--|--|------------------------------------------------------------------|

|                                                                                    |        |                          |                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                           |         |   |   |   |   |   |                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|------------------------------------------------------------------------------------|--------|--------------------------|------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------|---------|---|---|---|---|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| shng2<br>025/G<br>PT-Val<br>kyrie_<br>LN-12<br>4m__<br>AttnO<br>nly__<br>SQuA<br>D | L<br>N | At<br>tn<br>O<br>nl<br>y | Who<br>ruled<br>the<br>duc<br>h<br>y<br>of<br>Norm<br>andy | The Norman dynasty had a major political, cultural and military impact on medieval Europe and even the Near East. The Normans were famed for their martial spirit and eventually for their Christian piety, becoming exponents of the Catholic orthodoxy into which they assimilated. They adopted the Gallo-Romance language of the Frankish land they settled, their dialect becoming known as Norman, Normaund | ['Rich<br>ard I',<br>'Richa<br>rd I',<br>'Richa<br>rd I'] | 1<br>.0 | 1 | 1 | 1 | 1 | 1 | The generated answer "nan" is entirely inadequate. It does not provide any correct, complete, or relevant information, and it lacks fluency. The model failed to | - **Model Name**: The model name indicates that it is a variant of GPT-Valkyrie with attention-only mechanisms, fine-tuned on the SQuAD dataset. This suggests it should be capable of answering questions based on provided context, but it failed in this instance.<br>- **Norm Type**: LN (Layer Normalization) is a common technique used in neural networks to stabilize and accelerate training. However, the normalization technique did not seem to help in generating a correct answer here.<br>- **Variant**: AttnOnly indicates that the model uses attention mechanisms exclusively, which are typically effective for understanding context. Despite this, the model did not perform well.<br>- **Question**: The question is clear and straightforward, asking for the ruler of the duchy of Normandy.<br>- **Context**: The context provides sufficient information to answer the |
|------------------------------------------------------------------------------------|--------|--------------------------|------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------|---------|---|---|---|---|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|  |  |  |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |  |  |  |  |  |  |  |  |                                                 |                                                                                                                                                                                                                                                                                                                                                          |
|--|--|--|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|--|--|-------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  |  |  |  | or Norman French, an important literary language. The Duchy of Normandy, which they formed by treaty with the French crown, was a great fief of medieval France, and under Richard I of Normandy was forged into a cohesive and formidable principality in feudal tenure. The Normans are noted both for their culture, such as their unique Romanesque architecture and musical traditions, and for their significant military accomplishments and innovations. Norman adventurers founded the Kingdom of Sicily under Roger II after conquering southern Italy on the Saracens and Byzantines, and an expedition on behalf of their duke, William the Conqueror, led to the Norman conquest of England at the Battle of Hastings in 1066. Norman cultural and military influence spread from these new European centres to the Crusader states of the Near East, where their prince Bohemond I founded the Principality of Antioch in the Levant, to Scotland and Wales in Great Britain, to Ireland, and to the coasts of north Africa and the Canary Islands. |  |  |  |  |  |  |  |  | generate a meaningful response to the question. | question correctly, mentioning Richard I as the ruler of the duchy of Normandy.<br>- **Reference Answers**: The reference answers are consistent and correct, all indicating "Richard I."<br>- **Generated Answer**: The generated answer "nan" is a placeholder or error value, indicating a failure in the model's response generation process.<br>... |
|--|--|--|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|--|--|-------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                    |        |                          |                                                    |                                                                                                                                                                                                                                |                                                                         |                   |         |   |   |   |   |   |                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|------------------------------------------------------------------------------------|--------|--------------------------|----------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|-------------------|---------|---|---|---|---|---|-----------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| shng2<br>025/G<br>PT-Val<br>kyrie_<br>LN-12<br>4m__<br>AttnO<br>nly__<br>SQuA<br>D | L<br>N | A<br>ttn<br>O<br>nl<br>y | What<br>religi<br>on<br>were<br>the<br>Norm<br>ans | The Norman dynasty had a major political, cultural and military impact on medieval Europe and even the Near East. The Normans were famed for their martial spirit and eventually for their Christian piety, becoming exponents | ['Cath<br>olic',<br>'Catho<br>lic<br>orthod<br>oxy',<br>'Catho<br>lic'] | Chr<br>isti<br>an | 4<br>.2 | 4 | 3 | 4 | 5 | 5 | The<br>generat<br>ed<br>answer<br>is<br>generall<br>y<br>correct<br>and<br>relevant<br>but<br>lacks | - **Model Name**: The model name indicates that this is a specialized version of GPT, likely fine-tuned for question-answering tasks using the SQuAD dataset. This suggests the model should be proficient in generating contextually accurate answers.<br>- **Norm Type**: LN stands for Layer Normalization, a technique used to stabilize and improve the training of deep neural networks. This is a standard practice in modern NLP models. |
|------------------------------------------------------------------------------------|--------|--------------------------|----------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|-------------------|---------|---|---|---|---|---|-----------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|  |  |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |  |  |  |  |  |  |  |  |  |  |  |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|--|--|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|--|--|--|--|--|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  |  |  | <p>of the Catholic orthodoxy into which they assimilated. They adopted the Gallo-Romance language of the Frankish land they settled, their dialect becoming known as Norman, Normaund or Norman French, an important literary language. The Duchy of Normandy, which they formed by treaty with the French crown, was a great fief of medieval France, and under Richard I of Normandy was forged into a cohesive and formidable principality in feudal tenure. The Normans are noted both for their culture, such as their unique Romanesque architecture and musical traditions, and for their significant military accomplishments and innovations. Norman adventurers founded the Kingdom of Sicily under Roger II after conquering southern Italy on the Saracens and Byzantines, and an expedition on behalf of their duke, William the Conqueror, led to the Norman conquest of England at the Battle of Hastings in 1066. Norman cultural and military influence spread from these new European centres to the Crusader states of the Near East, where their prince Bohemond I founded the Principality of Antioch in the Levant, to Scotland and Wales in Great Britain, to Ireland, and to the coasts of north Africa and the Canary Islands.</p> |  |  |  |  |  |  |  |  |  |  |  |  | <p>the specificity of the reference answers. It is well-written and concise, but it could be improved by specifying that the Normans were Catholics.</p> <ul style="list-style-type: none"><li>- <b>Variant**:</b> AttnOnly indicates that the model uses attention mechanisms, which are crucial for understanding and generating contextually relevant text. This variant likely focuses on leveraging attention to improve answer accuracy.</li><li>- <b>Question**:</b> The question is straightforward and asks for the religion of the Normans, which is clearly stated in the context.</li><li>- <b>Context**:</b> The context provides detailed information about the Normans, including their adoption of Catholic orthodoxy. This specificity is important for generating a precise answer.</li><li>- <b>Reference Answers**:</b> The reference answers are specific ("Catholic," "Catholic orthodoxy") and provide a clear benchmark for evaluating the generated answer.</li><li>- <b>Generated Answer**:</b> The generated answer "Christian" is correct but less specific than the reference answers, indicating room for improvement in capturing finer details from the context.</li><li>...</li></ul> |
|--|--|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|--|--|--|--|--|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                  |        |               |                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |   |                                                                   |             |   |   |   |   |   |                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|----------------------------------------------------------------------------------|--------|---------------|------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|-------------------------------------------------------------------|-------------|---|---|---|---|---|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| shng2<br>025/G<br>PT-Val<br>kyrie_<br>LN-12<br>4m_<br>AttnO<br>nly_<br>SQuA<br>D | L<br>N | At<br>tn<br>O | What<br>type<br>of<br>major<br>impac<br>t did<br>the<br>Norm<br>an<br>dynas<br>ty<br>have<br>on<br>mode<br>rn<br>Europ<br>e? | The Norman dynasty had a major political, cultural and military impact on medieval Europe and even the Near East. The Normans were famed for their martial spirit and eventually for their Christian piety, becoming exponents of the Catholic orthodoxy into which they assimilated. They adopted the Gallo-Romance language of the Frankish land they settled, their dialect becoming known as Norman, Normaund or Norman French, an important literary language. The Duchy of Normandy, which they formed by treaty with the French crown, was a great fief of medieval France, and under Richard I of Normandy was forged into a cohesive and formidable principality in feudal tenure. The Normans are noted both for their culture, such as their unique Romanesque architecture and musical traditions, and for their significant military accomplishments and innovations. Norman adventurers founded the Kingdom of Sicily under Roger II after conquering southern Italy on the Saracens and Byzantines, and an expedition on behalf of their duke, William the Conqueror, led to the Norman conquest of England at the Battle of Hastings in 1066. Norman cultural and military influence spread from these new European centres to the Crusader states of the Near East, where | □ | poli<br>tica<br>l,<br>cult<br>ura<br>l<br>and<br>mili<br>tar<br>y | 4<br>.<br>4 | 5 | 3 | 5 | 4 | 5 | The generated answer is correct and highly relevant but lacks completeness in terms of detail. It is concise and fluent, making it a good but not perfect response. | - **Model Name**: The model used is shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD, which appears to be a specialized model for question-answering tasks.<br>- **Norm Type**: LN likely stands for Layer Normalization, a technique used in neural networks to stabilize and accelerate training.<br>- **Variant**: AttnOnly suggests that the model uses attention mechanisms exclusively, which can help in focusing on relevant parts of the context.<br>- **Question**: The question asks about the major impacts of the Norman dynasty on modern Europe, which is well-answered by the generated response.<br>- **Context**: The context provides detailed information about the political, cultural, and military impacts of the Norman dynasty, which the generated answer correctly identifies but does not elaborate on.<br>- **Reference Answers**: No reference answers are provided, making it challenging to compare the generated answer against a gold standard.<br>- **Generated Answer**: The generated answer is "political, cultural and military," which is correct but could be more detailed to fully address the question.<br>... |
|----------------------------------------------------------------------------------|--------|---------------|------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|-------------------------------------------------------------------|-------------|---|---|---|---|---|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                              |        |                      |                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |     |        |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |  |
|----------------------------------------------|--------|----------------------|--------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|--------|---|---|---|---|---|---|---|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|                                              |        |                      |                                            | their prince Bohemond I founded the Principality of Antioch in the Levant, to Scotland and Wales in Great Britain, to Ireland, and to the coasts of north Africa and the Canary Islands.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |     |        |   |   |   |   |   |   |   |                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |  |
| shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD | L<br>N | Attn<br>O<br>nl<br>y | Who was famous for their Christian spirit? | The Norman dynasty had a major political, cultural and military impact on medieval Europe and even the Near East. The Normans were famed for their martial spirit and eventually for their Christian piety, becoming exponents of the Catholic orthodoxy into which they assimilated. They adopted the Gallo-Romance language of the Frankish land they settled, their dialect becoming known as Norman, Normaund or Norman French, an important literary language. The Duchy of Normandy, which they formed by treaty with the French crown, was a great fief of medieval France, and under Richard I of Normandy was forged into a cohesive and formidable principality in feudal tenure. The Normans are noted both for their culture, such as their unique Romanesque architecture and musical traditions, and for their significant military accomplishments and innovations. Norman adventurers founded the Kingdom of Sicily under Roger II after conquering southern Italy on the Saracens and Byzantines, and an expedition on behalf of their duke, William the Conqueror, led to | Who | 1<br>0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | The generated answer "Who" is entirely inadequate. It fails to provide a correct, complete, relevant, fluent, and concise response to the question. The model needs significant improvement in understanding and generating appropriate answers based on the given context. | - **Model Name**: The model name "shng2025/GPT-Valkyrie_LN-124m_AttnOnly_SQuAD" suggests that it is a specialized model trained on the SQuAD dataset with attention-only mechanisms and layer normalization (LN). Despite these features, the model's performance in this instance is poor.<br>- **Norm Type**: LN (Layer Normalization) is a technique used to stabilize and improve the training of deep neural networks. However, in this case, it did not contribute to generating a correct answer.<br>- **Variant**: AttnOnly indicates that the model uses attention mechanisms exclusively. Attention mechanisms are designed to focus on relevant parts of the input, but the model failed to apply this effectively here.<br>- **Question**: The question "Who was famed for their Christian spirit?" is clear and straightforward, asking for a specific group or individual known for their Christian piety.<br>- **Context**: The context provides detailed information about the Normans, including their martial spirit and Christian piety. It clearly states that the Normans were famed for their Christian piety.<br>- **Reference Answers**: There are no reference answers provided, but the context itself is sufficient to determine the correct answer.<br>- **Generated Answer**: The generated answer "Who" is not a valid response to the question and indicates a failure in the model's comprehension and generation capabilities.<br>`` |  |

|                                                                                    |        |                          |                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |     |         |   |   |   |   |   |   |                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |
|------------------------------------------------------------------------------------|--------|--------------------------|-------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|---------|---|---|---|---|---|---|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|                                                                                    |        |                          |                                                             | the Norman conquest of England at the Battle of Hastings in 1066. Norman cultural and military influence spread from these new European centres to the Crusader states of the Near East, where their prince Bohemond I founded the Principality of Antioch in the Levant, to Scotland and Wales in Great Britain, to Ireland, and to the coasts of north Africa and the Canary Islands.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |     |         |   |   |   |   |   |   |                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |
| shng2<br>025/G<br>PT-Val<br>kyrie_<br>LN-12<br>4m__<br>AttnO<br>nly__<br>SQuA<br>D | L<br>N | At<br>tn<br>O<br>nl<br>y | Who<br>assim<br>ilated<br>the<br>Roma<br>n<br>langu<br>age? | The Norman dynasty had a major political, cultural and military impact on medieval Europe and even the Near East. The Normans were famed for their martial spirit and eventually for their Christian piety, becoming exponents of the Catholic orthodoxy into which they assimilated. They adopted the Gallo-Romance language of the Frankish land they settled, their dialect becoming known as Norman, Normaund or Norman French, an important literary language. The Duchy of Normandy, which they formed by treaty with the French crown, was a great fief of medieval France, and under Richard I of Normandy was forged into a cohesive and formidable principality in feudal tenure. The Normans are noted both for their culture, such as their unique Romanesque architecture and musical traditions, and for their significant military accomplishments and innovations. | Who | 1<br>.0 | 1 | 1 | 1 | 1 | 1 | 1 | The generated answer is entirely inadequate. It does not address the question, lacks relevance, and is not well-formed. The model needs significant improvement in understanding and generating appropriate responses based on the context. | - **Model Name**: The model name indicates that it is a specialized version of GPT, likely fine-tuned for question-answering tasks using the SQuAD dataset. However, the performance in this instance is poor.<br>- **Norm Type**: LN stands for Layer Normalization, a technique used to stabilize and improve the training of deep neural networks.<br>- **Variant**: AttnOnly suggests that the model uses attention mechanisms exclusively, which should theoretically help in focusing on relevant parts of the context. However, this did not translate into a correct answer here.<br>- **Question**: The question is clear and asks for the entity that assimilated the Roman language.<br>- **Context**: The context provides detailed information about the Normans, including their adoption of the Gallo-Romance language, which is relevant to the question.<br>- **Reference Answers**: There are no reference answers provided, which makes it difficult to compare the generated answer against a standard.<br>- **Generated Answer**: The generated answer "Who" is not a valid response and indicates a failure in the model's ability to comprehend and generate a meaningful answer.<br>`` |  |

|                                                                                    |        |                          |                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |   |         |   |   |   |   |   |   |                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |  |
|------------------------------------------------------------------------------------|--------|--------------------------|-----------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|---------|---|---|---|---|---|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
|                                                                                    |        |                          |                                                           | Norman adventurers founded the Kingdom of Sicily under Roger II after conquering southern Italy on the Saracens and Byzantines, and an expedition on behalf of their duke, William the Conqueror, led to the Norman conquest of England at the Battle of Hastings in 1066. Norman cultural and military influence spread from these new European centres to the Crusader states of the Near East, where their prince Bohemond I founded the Principality of Antioch in the Levant, to Scotland and Wales in Great Britain, to Ireland, and to the coasts of north Africa and the Canary Islands.                                                                                    |   |         |   |   |   |   |   |   |                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |  |
| shng2<br>025/G<br>PT-Val<br>kyrie_<br>LN-12<br>4m__<br>AttnO<br>nly__<br>SQuA<br>D | L<br>N | At<br>tn<br>O<br>nl<br>y | Who<br>ruled<br>the<br>cou<br>try of<br>Norm<br>andy<br>? | The Norman dynasty had a major political, cultural and military impact on medieval Europe and even the Near East. The Normans were famed for their martial spirit and eventually for their Christian piety, becoming exponents of the Catholic orthodoxy into which they assimilated. They adopted the Gallo-Romance language of the Frankish land they settled, their dialect becoming known as Norman, Normaund or Norman French, an important literary language. The Duchy of Normandy, which they formed by treaty with the French crown, was a great fief of medieval France, and under Richard I of Normandy was forged into a cohesive and formidable principality in feudal | 0 | 1<br>.0 | 1 | 1 | 1 | 1 | 1 | 1 | The generated answer "nan" is entirely inadequate. It does not provide any information relevant to the question, lacks completeness, and is not a valid or fluent response. The model failed to generate a meaningful answer based on the context provided. | - **Model Name**: The model name indicates that it is a variant of GPT-Valkyrie with attention-only mechanisms and trained on the SQuAD dataset. This suggests it should be capable of answering questions based on context, but it failed in this instance.<br>- **Norm Type**: LN (Layer Normalization) is a common technique used in neural networks to stabilize and accelerate training. However, the normalization technique did not seem to help in generating a correct answer here.<br>- **Variant**: AttnOnly indicates that the model uses attention mechanisms exclusively. Attention mechanisms are designed to focus on relevant parts of the input, but the model did not utilize this effectively in this case.<br>- **Question**: The question is clear and specific, asking for the ruler of Normandy. The context provides sufficient information to answer this question.<br>- **Context**: The context is rich with historical details about the Norman dynasty, including specific rulers like Richard I and William the Conqueror. The model should have been able to extract this information to generate a correct answer.<br>- **Reference Answers**: No reference answers are provided, but the context itself contains the necessary information to answer the question accurately.<br>`` |  |

|                                                                                    |        |                          |                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |   |                                                                                                                                                                        |        |   |   |   |   |   |                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |  |
|------------------------------------------------------------------------------------|--------|--------------------------|-----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|
|                                                                                    |        |                          |                                                                                   | tenure. The Normans are noted both for their culture, such as their unique Romanesque architecture and musical traditions, and for their significant military accomplishments and innovations. Norman adventurers founded the Kingdom of Sicily under Roger II after conquering southern Italy on the Saracens and Byzantines, and an expedition on behalf of their duke, William the Conqueror, led to the Norman conquest of England at the Battle of Hastings in 1066. Norman cultural and military influence spread from these new European centres to the Crusader states of the Near East, where their prince Bohemond I founded the Principality of Antioch in the Levant, to Scotland and Wales in Great Britain, to Ireland, and to the coasts of north Africa and the Canary Islands. |   |                                                                                                                                                                        |        |   |   |   |   |   |                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |  |
| shng2<br>025/G<br>PT-Val<br>kyrie_<br>LN-12<br>4m__<br>AttnO<br>nly__<br>SQuA<br>D | L<br>N | At<br>tn<br>O<br>nl<br>y | What<br>princi<br>pality<br>did<br>Willia<br>m the<br>conq<br>uerer<br>found<br>? | The Norman dynasty had a major political, cultural and military impact on medieval Europe and even the Near East. The Normans were famed for their martial spirit and eventually for their Christian piety, becoming exponents of the Catholic orthodoxy into which they assimilated. They adopted the Gallo-Romance language of the Frankish land they settled, their dialect becoming known as Norman, Normaund or Norman French, an important literary language. The Duchy of Normandy,                                                                                                                                                                                                                                                                                                      | □ | feudal<br>ten<br>ture<br>. The<br>Nor<br>ma<br>ns<br>are<br>not<br>ed<br>bot<br>h<br>for<br>thei<br>r<br>cult<br>ure<br>,<br>suc<br>h<br>as<br>thei<br>r<br>uni<br>que | 1<br>6 | 1 | 1 | 1 | 3 | 2 | The generated answer fails to correctly and completely address the question. It is not relevant to the specific query about the principality founded by William the Conqueror. | - **Model Name**: The model name indicates that this is a specialized version of GPT designed for question-answering tasks, specifically trained on the SQuAD dataset. However, the performance in this instance is subpar.<br>- **Norm Type**: LN (Layer Normalization) is a technique used to stabilize and improve the training of deep neural networks. Despite this, the model's output in this case is not satisfactory.<br>- **Variant**: AttnOnly suggests that the model uses attention mechanisms exclusively, which should theoretically help in focusing on relevant parts of the context. However, the model failed to focus on the relevant information in this instance.<br>- **Question**: The question is clear and specific, asking about the principality founded by William the Conqueror.<br>- **Context**: The context provides sufficient information to answer the question correctly. It mentions the Duchy of Normandy, which is the correct answer. |  |  |

|  |  |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                |                                                                                                               |                                                                                                                                                                                                                                                                                                               |
|--|--|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  |  |  | <p>which they formed by treaty with the French crown, was a great fief of medieval France, and under Richard I of Normandy was forged into a cohesive and formidable principality in feudal tenure. The Normans are noted both for their culture, such as their unique Romanesque architecture and musical traditions, and for their significant military accomplishments and innovations. Norman adventurers founded the Kingdom of Sicily under Roger II after conquering southern Italy on the Saracens and Byzantines, and an expedition on behalf of their duke, William the Conqueror, led to the Norman conquest of England at the Battle of Hastings in 1066. Norman cultural and military influence spread from these new European centres to the Crusader states of the Near East, where their prince Bohemond I founded the Principality of Antioch in the Levant, to Scotland and Wales in Great Britain, to Ireland, and to the coasts of north Africa and the Canary Islands.</p> | <p>Romanesque architecture and musical traditions, and for their significant military accomplishments and innovations. Norman adventurers founded the Principality of Antioch in the Levant, to Scotland and Wales in Great Britain, to Ireland, and to the coasts of north Africa and the Canary Islands.</p> | <p>While the text is somewhat fluent, it is incomplete and lacks conciseness in relation to the question.</p> | <p>- **Reference Answers**: No reference answers are provided, but the context clearly indicates the correct answer.<br/> - **Generated Answer**: The generated answer is incorrect and incomplete, failing to address the question about the principality founded by William the Conqueror.<br/>     ...</p> |
|--|--|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                  |        |              |                                                     |                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                         |                                                                                         |         |   |   |   |   |   |                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |  |  |  |  |  |  |  |  |
|----------------------------------------------------------------------------------|--------|--------------|-----------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|---------|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|--|--|
|                                                                                  |        |              |                                                     |                                                                                                                                                                                                                                                                                                                                                             | on the Saracens and Byzantines, and an expedition on behalf of their duke, William the Conqueror, led to the Norman conquest of England |                                                                                         |         |   |   |   |   |   |                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |  |  |  |  |  |  |  |  |
| shng2<br>025/G<br>PT-Val<br>kyrie_<br>LN-12<br>4m_<br>AttnO<br>nly_<br>SQuA<br>D | L<br>N | Attn<br>Only | What is the origin al meani ng of the word Norm an? | The English name "Normans" comes from the French words Normans/Norman, plural of Norman, modern French normand, which is itself borrowed from Old Low Franconian Nortmann "Northman" or directly from Old Norse Norðmaðr, Latinized variously as Nortmannus, Normannus, or Nordmannus (recorded in Medieval Latin, 9th century) to mean "Norseman, Viking". | ['Viking', 'Norse man, Viking', 'Norse man, Viking']                                                                                    | What is the ori gin al me ani ng of the wor d Nor ma n? The En glis h na me "Nor ma ns" | 3<br>.4 | 4 | 3 | 4 | 3 | 3 | The generat ed answer is mostly correct and relevant but lacks completeness and conciseness. It provides a detailed etymolo gical backgro und but fails to explicitl y mention | - **Model Name**: The model name indicates that this is a specialized version of GPT, likely fine-tuned for question-answering tasks on the SQuAD dataset. The "AttnOnly" variant suggests it uses attention mechanisms, which are crucial for understanding context.<br>- **Norm Type**: "LN" likely stands for Layer Normalization, a technique used to stabilize and accelerate the training of deep neural networks.<br>- **Variant**: "AttnOnly" indicates that the model uses attention mechanisms, which help in focusing on relevant parts of the input context to generate more accurate answers.<br>- **Question**: The question is clear and straightforward, asking for the original meaning of the word "Norman."<br>- **Context**: The context provides a detailed etymological background of the word "Norman," tracing its origins through various languages and historical periods.<br>- **Reference Answers**: The reference answers are concise and directly address the question, providing the expected terms "Viking" and "Norseman, Viking." |  |  |  |  |  |  |  |  |

comes from the French words Normans/ Normanz, plural of Norman, modern French normand, which is itself borrowed from Old Low Franco-norman "Nort hm an" or directly from Old Norse Norðm aðr, Latinized various

"Viking" and could be more directly focused on the original meaning. The fluency is acceptable but could be improved with better punctuation and sentence structure.

- \*\*Generated Answer\*\*: The generated answer is detailed but somewhat verbose and lacks explicit mention of "Viking," which is part of the expected answer. It also has minor issues with fluency and conciseness.  
``

|                                                                                    |        |                          |                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                  |                                                                    |          |         |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |  |  |  |
|------------------------------------------------------------------------------------|--------|--------------------------|------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|----------|---------|---|---|---|---|---|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|
|                                                                                    |        |                          |                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                  |                                                                    |          |         |   |   |   |   |   |   |                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |  |  |  |
| shng2<br>025/G<br>PT-Val<br>kyrie_<br>LN-12<br>4m__<br>AttnO<br>nly__<br>SQuA<br>D | L<br>N | At<br>tn<br>O<br>nl<br>y | When<br>was<br>the<br>Latin<br>versi<br>on of<br>the<br>word<br>Norm<br>an<br>first<br>recor<br>ded? | The English name<br>"Normans" comes<br>from the French<br>words<br>Normans/Normanç,<br>plural of Normant,<br>modern French<br>normand, which is<br>itself borrowed from<br>Old Low Franconian<br>Nortmann<br>"Northman" or<br>directly from Old<br>Norse Norðmaðr,<br>Latinized variously<br>as Nortmannus,<br>Normannus, or<br>Nordmannus<br>(recorded in<br>Medieval Latin, 9th<br>century) to mean<br>"Norseman, Viking". | ly<br>as<br>Nor<br>tma<br>nn<br>us,<br>Nor<br>ma<br>nn<br>us,<br>or<br>Nor<br>dm<br>ann<br>us<br>(re<br>cor<br>ded<br>in<br>Me<br>die<br>val<br>Lat<br>in,<br>9th<br>cen<br>tur<br>y)<br>to<br>me<br>an<br>"N<br>ors<br>em<br>an | ['9th<br>centur<br>y', '9th<br>centur<br>y', '9th<br>centur<br>y'] | Wh<br>en | 1<br>.6 | 1 | 1 | 1 | 1 | 2 | 3 | The<br>generat<br>ed<br>answer<br>"When"<br>is<br>inccore<br>ct,<br>incompl<br>ete, and<br>irrelev<br>ant to the<br>questio<br>n. It<br>does<br>not<br>provide<br>the<br>necessa<br>ry<br>informa<br>tion and<br>fails to<br>meet the<br>criteria<br>for a<br>good<br>answer.<br>The<br>model<br>needs<br>signific<br>ant<br>improve | - <b>Model Name**:</b> The model name<br>indicates that it is a specialized version of<br>GPT, likely fine-tuned for the SQuAD<br>dataset. However, the performance on this<br>specific question is poor.<br>- <b>Norm Type**:</b> LN (Layer Normalization)<br>is a common technique used in neural<br>networks to stabilize and accelerate<br>training. It is unclear if this contributed to<br>the poor performance.<br>- <b>Variant**:</b> AttnOnly suggests that the<br>model uses attention mechanisms<br>exclusively, which should theoretically<br>help in focusing on relevant parts of the<br>context. However, this did not translate<br>into a correct answer in this case.<br>- <b>Question**:</b> The question is clear and<br>specific, asking for the time period when<br>the Latin version of the word "Norman"<br>was first recorded.<br>- <b>Context**:</b> The context provides<br>detailed information about the origins and<br>Latinization of the word "Norman,"<br>including the specific time period (9th<br>century).<br>- <b>Reference Answers**:</b> The reference<br>answers are consistent and correct, all<br>indicating "9th century."<br>- <b>Generated Answer**:</b> The generated<br>answer "When" is entirely inadequate and<br>does not address the question.<br>`` |  |  |  |

|                                                                                    |        |                          |                                                                                                      |                                                                                                                                                                                                                                                                                                                                                               |   |                       |                                 |                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |  |  |  |  |  |  |                                                                          |  |
|------------------------------------------------------------------------------------|--------|--------------------------|------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|-----------------------|---------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|--------------------------------------------------------------------------|--|
|                                                                                    |        |                          |                                                                                                      |                                                                                                                                                                                                                                                                                                                                                               |   |                       |                                 |                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |  |  |  |  |  |  | ment in understanding and generating contextually appropriate responses. |  |
| shng2<br>025/G<br>PT-Val<br>kyrie_<br>LN-12<br>4m__<br>AttnO<br>nly__<br>SQuA<br>D | L<br>N | At<br>tn<br>O<br>nl<br>y | What<br>name<br>come<br>s<br>from<br>the<br>Engli<br>sh<br>word<br>s<br>Norm<br>ans/N<br>orma<br>nz? | The English name "Normans" comes from the French words Normans/Normanç, plural of Normant, modern French normand, which is itself borrowed from Old Low Franconian Nortmann "Northman" or directly from Old Norse Norðmaðr, Latinized variously as Nortmannus, Normannus, or Nordmannus (recorded in Medieval Latin, 9th century) to mean "Norseman, Viking". | □ | Nor<br>ma<br>ns<br>.0 | 5<br>5<br>5<br>5<br>5<br>5<br>5 | The generated answer is accurate, complete, relevant, fluent, and concise. It perfectly matches the requirements of the question based on the provided context. | - **Model Name**: The model name indicates that it is a variant of GPT-Valkyrie with attention-only mechanisms, fine-tuned on the SQuAD dataset. This suggests it is designed for question-answering tasks.<br>- **Norm Type**: LN stands for Layer Normalization, a technique used to stabilize and improve the training of deep neural networks.<br>- **Variant**: AttnOnly indicates that the model uses attention mechanisms exclusively, which can enhance its ability to focus on relevant parts of the context.<br>- **Question**: The question is clear and specific, asking for the name derived from the English words Normans/Normanç.<br>- **Context**: The context provides detailed etymological information about the name "Normans," tracing its origins through various languages and historical forms.<br>- **Reference Answers**: The absence of reference answers means the evaluation relies solely on the context and the generated answer.<br>- **Generated Answer**: The generated answer "Normans" is correct and meets all the evaluation criteria effectively.<br>... |  |  |  |  |  |  |                                                                          |  |