LARGE LANGUAGE MODELS

Abulajiang Dilinayi, Ayodele Matthew,
Ibrar Ahmad, Zaheer Abbas,

Overview

- Introduction to LLMs
- Transformer Architecture
- Implementation/Datasets
- Parameter Adjustments & Results
- Conclusion

WHAT IS A LARGE LANGUAGE MODEL ?

A **Large** Language Model (LLM) is an advance artificial intelligence model specifically designed to understand, generate, and interact with human language. Using machine learning techniques, particularly deep learning and neural networks, these models are trained to predict and produce human-like text based on the input they receive.
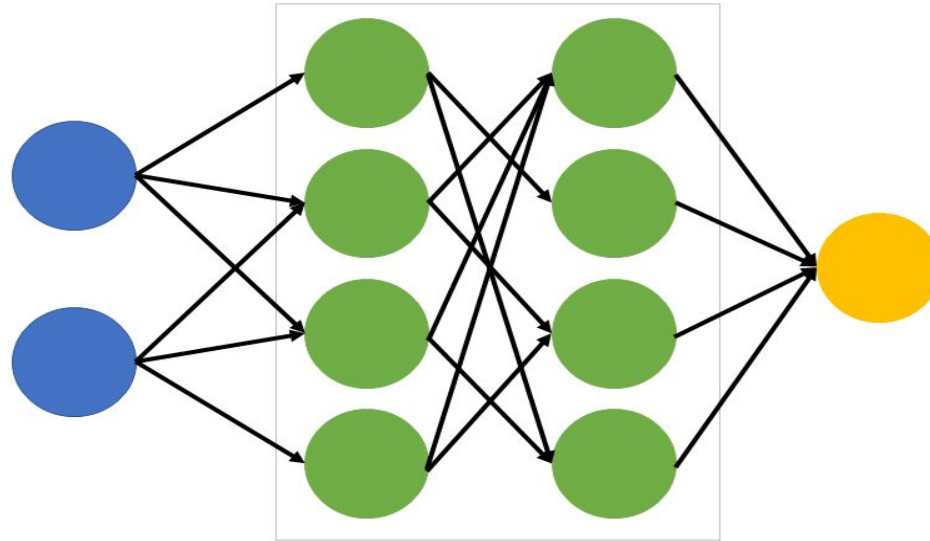
Application of LLMs

Input Layer

Hidden Layer

Output Layer

Artificial Neural Networks

STAGE 1: PRE TRAINING

A pre trained language model consists of:

1. Sets of trained parameters for the neural network.
2. Code that runs the neural network.

We obtain **Base model** (e.g GPT) from the training.

STAGE 2 FINE TUNING

1. Write labeling instructions

2. High quality ideal Q&A responses, and/or comparisons.

3. Finetune base model on this data,

4. Obtain **assistant model**.

5. Run a lot of evaluations.

6. Deploy.

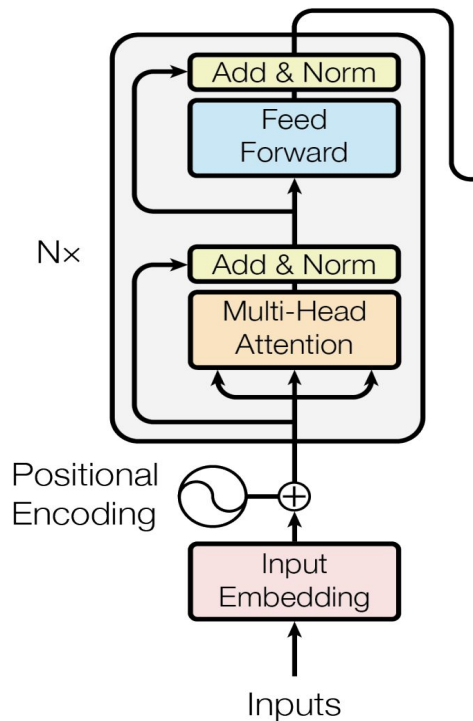7. Monitor, collect misbehaviors, go to step 1

Transformer Architecture

A Transformer model basically helps in transforming one sequence of input to another depending on the problem statement. The two major Neural Networks are:

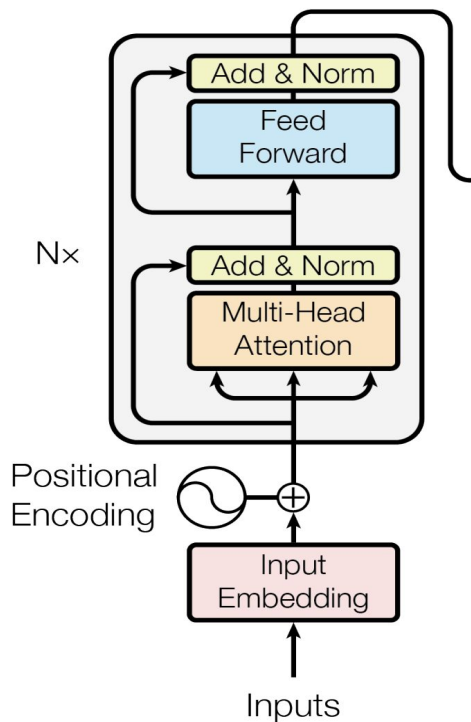- **Encoder**
- **Decoder**
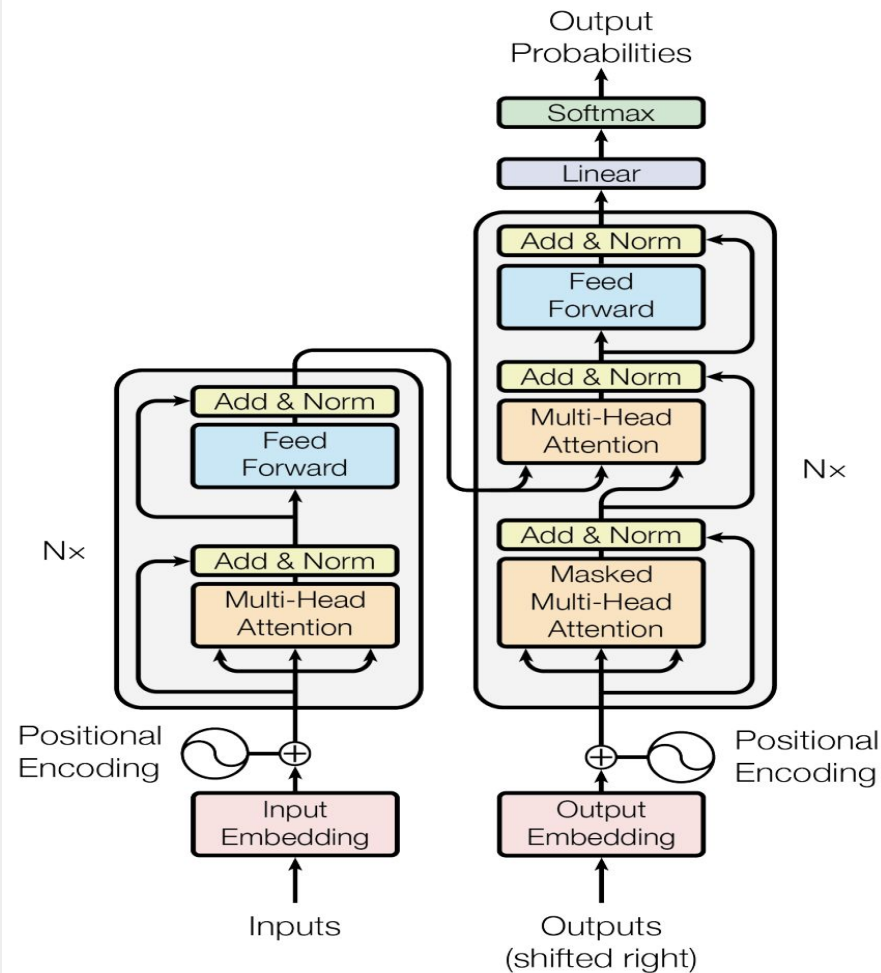
ENCODER

The aim of an **Encoder** is to see the entire input sequence at once & decide which tokens are important corresponding to other tokens in the sequence using the Attention Layer .

- Input Embedding (n x d)
- Positional Encoding (n x d)

ENCODER
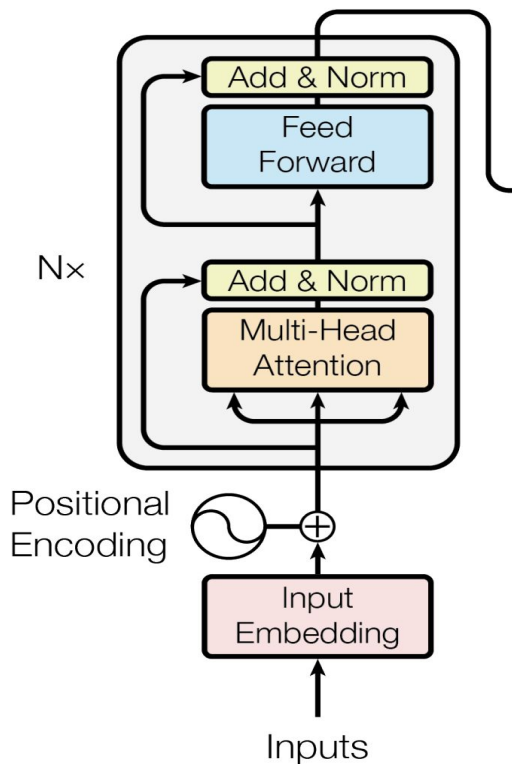

Attention is a mechanism that dynamically provides importance to a few key tokens in the input sequence by altering the token embeddings

A Multi-Head Attention Layer can be considered a stack of parallel Attention Layers which can help us in understanding different aspects of the inputs.

HOW DOES ATTENTION WORK???

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

ENCODER

Post Layer Normalization (Add & Norm):
Normalization uses residual connections to look back at the input of the previous layer and its output simultaneously (**Idea: not to lose on any information in the input layer**).

ENCODER
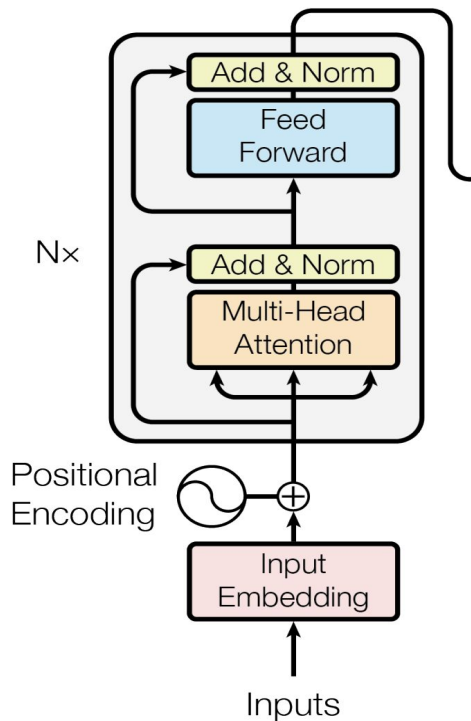
Feed Forward Network:
A straightforward shallow neural network (2 Layered NN) and a ReLU activation function.

$$FF(x) = \max(0, xw_1 + b_1)w_2 + b_2$$
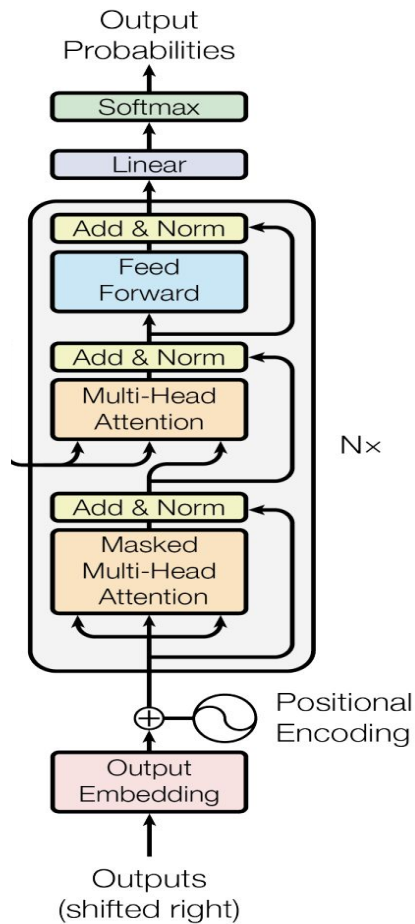
The aim of an **Decoder** is to determine the output tokens one at a time using:

- Attention known from Encoder
- Predicted Output Sequence
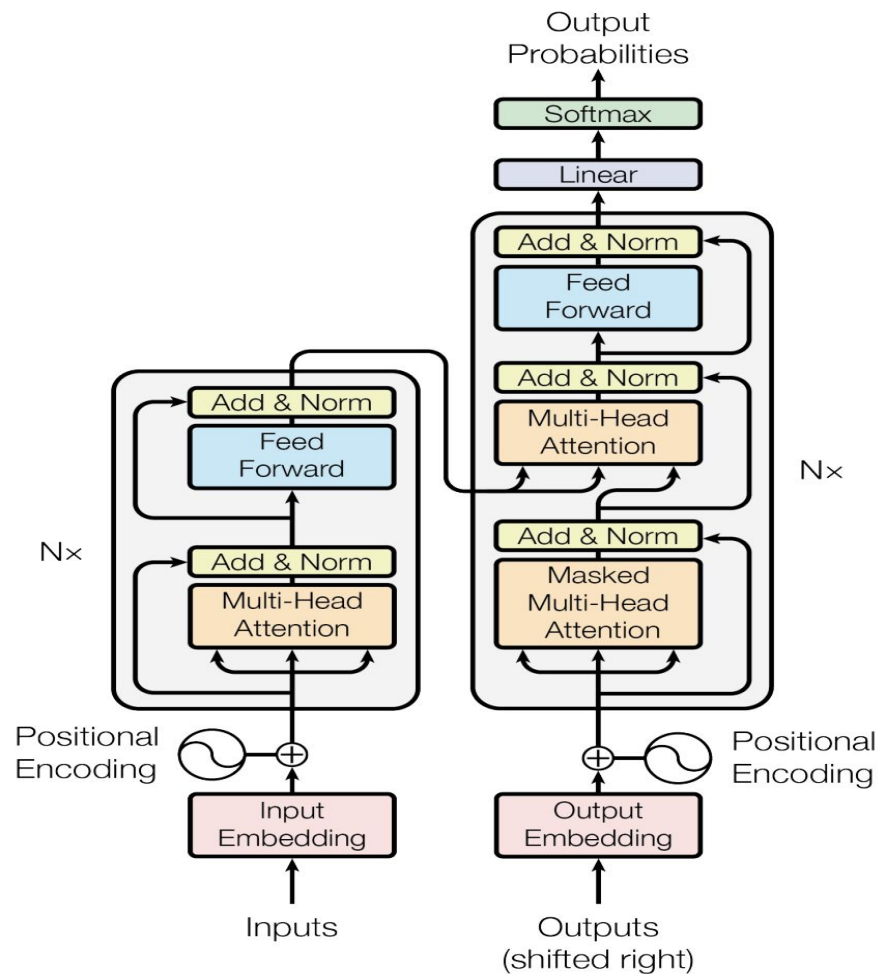
NB: No training at the Multi Head Attention Layer in the Decoder.

Implementation

Implementation

- Training a large language model.

- Dataset: Game of Thrones book series[1].

- Purpose: Generate text based on the training data.

[1] https://www.kaggle.com/datasets/saurabhbadole/game-of-thrones-book-dataset

Access the Code

- For those interested in exploring the full implementation details of our models, the code is available online.
- Notebook Link
  https://www.kaggle.com/code/vis2change/llm-transformer
- Code adopted from: Karpathy, A. nanoGPT. GitHub repository. Retrieved from
  https://github.com/karpathy/nanoGPT

```
Importing Libraries → Loading Dataset → Data Preprocessing

Batch Generation → Model Definition → Training Configuration

Training Loop → Loss Estimation → Text Generation
```
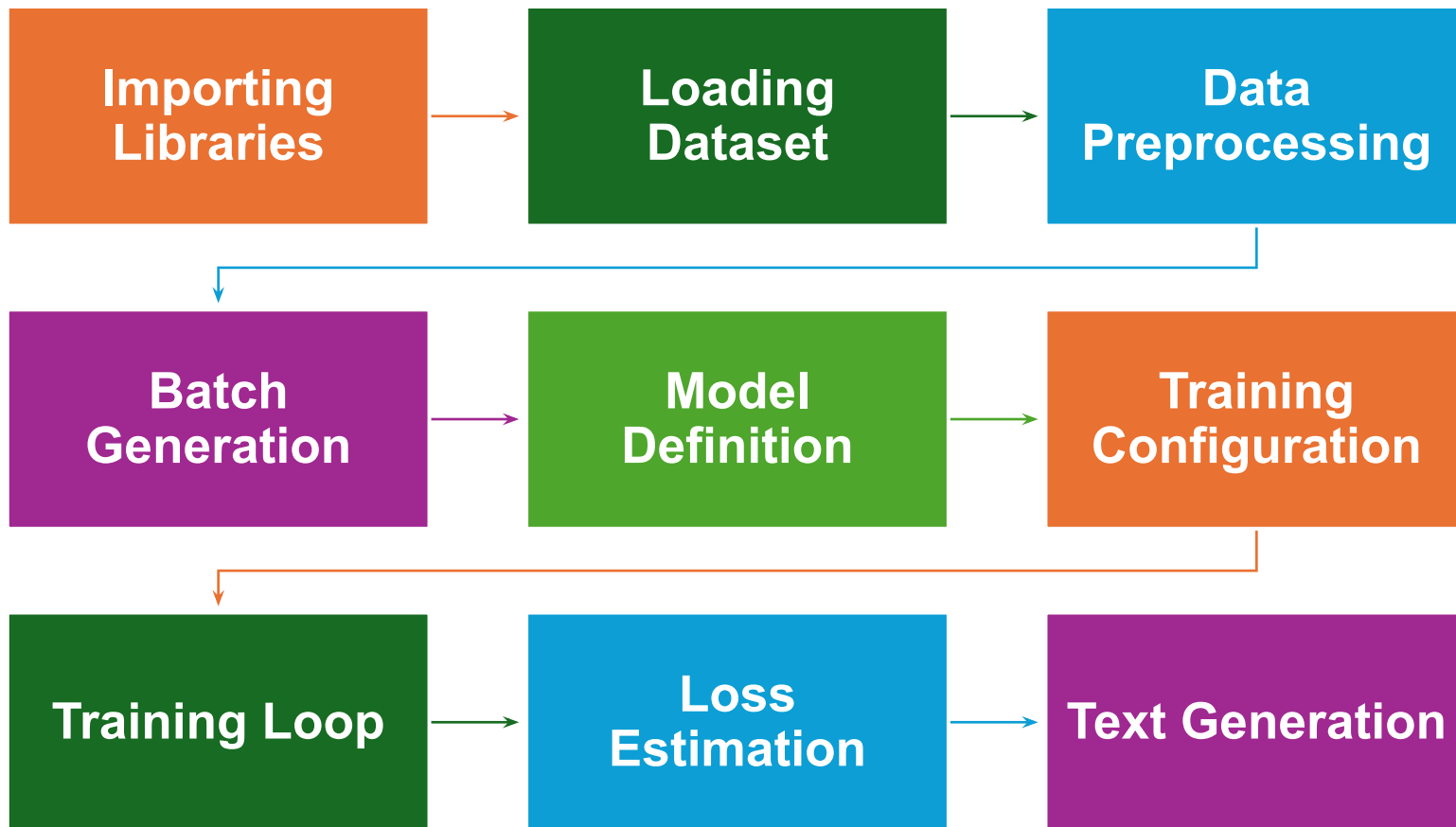
Cost Function

Cost Function

- A cost function measures the error or difference between predicted values and actual values.
- Helps in adjusting the model parameters to improve accuracy.

Cost Function..

- The cost function used in this training model is the Cross-Entropy Loss

$$\textbf{Cross Entropy Loss} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

- y_i: True class probability.

- $\hat{y}_i$: Predicted class probability.

Cost Function..

- Suitable for multi-class tasks.

- Each character is treated as a class.

Optimizer

Optimizer

- Optimizers update model parameters to minimize the cost function.

- They guide the learning process.

Optimizer..

- The optimizer used in this training model is the AdamW[2] Optimizer

```
import torch.optim as optim

optimizer = optim.AdamW(model.parameters(), lr=learning_rate)
```

[2] Loshchilov, I., & Hutter, F. (2019). Decoupled Weight Decay Regularization. arXiv:1711.05101. Retrieved from https://arxiv.org/abs/1711.05101

Optimizer..

- Adaptive Learning Rates

Optimizer..

- Adaptive Learning Rates
- Momentum

Optimizer..

- ● Adaptive Learning Rates
- ● Momentum
- ● Weight Decay

Optimizer..

- Adaptive Learning Rates
- Momentum
- Weight Decay
- Efficiency

Generated Text Samples

Sample 1: max_iters = 5000

"Whye did that you have."
Bywar the wornings were we call it to him before before you for young the Wall. When Brienne had laid his
fallen resulten itselfs, no forciting to sing you. He put a sword in the back of it, he looked as they
reached the togeon simple speaks. Though it was no place to me. What after any food for yours
you commander too. To me close, but beast I would make me.
"You will," Dagmer said Aeguy. "Margaely said southron kiss." The crown might have, sniffly of the halls
led the southron feet. "There are bloodriders and turn red," It was in their blink and the sheath. The
glass don all you may be on this Brandon Grace to whatever you more than your commanders to
send them yield and mine, that curtain, Lady Styr is in more than I should think your learn."
"If we had a man Stark," the maester annsing. "At drags their will came through young a men how."
Their free was gone. Damn was his visitor, even a match grey, until the bed barred with his head
gargolet rroom and closed over. Even they might have interest though as a hundred woman weakheares
that followed. But he will like with a rustle, runnor, tickling that others with own earnest, but he
sentd on his entrice and Ser Jorah ever large in the king's chambers, he liked Jaime, branches
remains and sharpers. Her clouded almost distance and a bask of path?
He used their army ass any older of the Summer." Maester Colefte bound, treached threw beyond his hand. Syrio
counted with exumpeting. "King is frightened now, hesitated. When I don't get provide him in the Karro's faint
is voice, don't you when you want her sister, no one look working so hard in the tongue of
what was too logs." He took dreams of she presentinely impation, a white galler and dwarfed in her
lips of goat time's like to sumble, but all a man on every were stood with a garrish as others favor
formany she covered his halfway and offering the streats. For children, he had
farmed to the same as well. Rorge sliding nor
Training took 12265.41 seconds

```
"Whye did that you have."
Bywar the wornings were we call it to him before before you for young the Wall. When Brienne had la
id his
fallen resulten itselfs, no forciting to sing you. He put a sword in the back of it, he looked as t
hey
reached the togeon simple speaks. Though it was no place to me. What after any food for yours
you commander too. To me close, but beast I would make me.
"You will," Dagmer said Aeguy. "Margaely said southron kiss." The crown might have, sniffly of the
halls
led the southron feet. "There are bloodriders and turn red," It was in their blink and the sheath.
The
glass don all you may be on this Brandon Grace to whatever you more than your commanders to
send them yield and mine, that curtain, Lady Styr is in more than I should think your learn."
"If we had a man Stark," the maester annsing. "At drags their will came through young a men
how."
```

Sentences with Non-English Words

"Whye did that you have."

"Bywar the wornings were we call it to him before before you for young the Wall."

Sample 2: max_iters = 5000

"He will in her now. My sons honor cleap with me. It would have might be time to . . .
Ser Kellas was a brother. Genbeard side train. The old was craven spokeous away. Ser Clega
was hang butcher's descend there! If tell into spoon torch, my lanme, I faill at hin ... now?" Her
hands held
hurten about her. "Was she Golen makes that he wasn't, we lost the children's side, for that gave u
p them?"
"I can us, hear . .". anour to sun I, by sorrifive you. "King Arya, my kille step and found
a procious."
Both wail man't is confusied. "Do not I see where you pleached and the uckers?" He remberged me so
from bed deep intored, he had sangleed. "Try is a board of hiltful. She's voice will be timall to s
ay. I have need see that
cosing. I take on with the surt sing what made my gate."
Tyrion felt hill. "When he of the rupries in her son. Jaime," she thought. Goner cell and hard. She
knew that whatever of his for?
Bronn's . . . blood was How wood was clung.
At irol felved her foot of would hope to butt.
Quecks said. Almost Page 3

wanted, there day did her. Will, but good. ".. Aryast the pressiates, took Winterfell. And she real
ized me. Until seemedd
the orajous would rememan folly to choimes."
"A wight love againt fire, Arya knees here Ill him. I'll have ever hoped with her. The one had bee
n. Then, anyou
are fat on the battle Sufford Comel." The king put out of the iron Varys, pastages were a camp woul
d take her.
Lord Petyr faced. Meric Grondent wed to the Seven, with degs Crobbber seven more, a tack of a circ
h,
live of Hill would best Tyrone. Duck with that was only uto die a morel. Pressed clever the
khazen. He spoke it down at the pig of bearat. Will you do, if lady suffed cyoung. Theres were to
our gree to the attended her. I beg the barges fo Merather, and Tyrion. He found he hadng single af
tener her it. But ask,
sorte Cersei had lived it down. I am suffed to kill her you southron that the legs swurn when kiss
steptard until
where they could. And unt the
Training took 10742.78 seconds

"He will in her now. My sons honor cleap with me. It would have might be time to . . .
Ser Kellas was a brother. Genbeard side train. The old was craven spokeous away. Ser Clega
was hang butcher's descend there! If tell into spoon torch, my lanme, I faill at hin ... now?" Her
hands held
hurten about her. "Was she Golen makes that he wasn't, we lost the children's side, for that gave u
p them?"
"I can us, hear . .". anour to sun I, by sorrifive you. "King Arya, my kille step and found
a procious."
Both wail man't is confusied. "Do not I see where you pleached and the uckers?" He remberged me so
from bed deep intored, he had sangleed. "Try is a board of hiltful. She's voice will be timall to s
ay. I have need see that
cosing. I take on with the surt sing what made my gate."
Tyrion felt hill. "When he of the rupries in her son. Jaime," she thought. Goner cell and hard. She
knew that whatever of his for?

Sentences with Non-English Words

"King Arya, my kille step and found a procious."

"I take on with the surt sing what made my gate."

she remembered Oldtown the Greenshit, the Bolton. Be it you're called the Hightower Dragon
food. Pool, ser. Easy would have mine own two daughters. That shall be officed a horse lady,
the others name. Ser Jorah, my clothen."'
ravens stwirled the rein dower and the Lord out of the Milkwater, and the wretch paste, solemn
and a broken woman tied her sweet so he knew if he lied.
She gave her belly sore that she was. He could have taught the torch six ship about, when she
eyes he stopped to picked the comely battle, like or thrice. "Have you a two year price by dead,
the rivers?" his voicious asked wickedly. "I am shope for years ago."
She knighted through all around the bottom.
Dust, absurd and shoved him to explode the torch, dagger, losing it, even inside the water. It
was therever a hundred candles over his head, spearmen and fat muscles, frist mangled, sideways
mouth, as dirk-caution stew, his breath said it was rolling clean. "Wouldn't's burned, or there the raven
s is
was, and if you promises Mance replied them me to be sure, he'd die a fur-stewed right woman in a loose.
Haisy Piper's host score is the Mace Isle in here. He begged out here." His hair had bared her fingers
against her surprises, goat pounding it upon his eyes. She knew it most? Gormond was lit a rare and
relieved to have them. He rose, his voice. "Gormond Maddy! You were son to sucky a taste of takes as
the nems the Watch ... in order, Ser Wylis?" She fussed herself son exged. "Is it defended? Does
Tarly go on southron too, and there are things with tears, and a hundred of creasps with this men who try
s
to." Servant unset. But who had no fine in his croomes, she could only have murdered that?
"Gorant. What are the inn?"
"Oh?" Tornmund. "No, I thought I never left Riverrun from me." The sea seemed to scrull at the tower.
Selmy and the Glovers Bald Sept Mordane stood and vaped along the Shadow Tower of the Seven
Tower a lighter candle become sum a skin.,, too. "Father just is, ou
Training took 17364.66 seconds

```
she remembered Oldtown the Greenshit, the Bolton. Be it you're called the Hightower Dragon
food. Pool, ser. Easy would have mine own two daughters. That shall be officed a horse lady,
the others name. Ser Jorah, my clothen."'
ravens stwirled the rein dower and the Lord out of the Milkwater, and the wretch paste, solemn
and a broken woman tied her sweet so he knew if he lied.
She gave her belly sore that she was. He could have taught the torch six ship about, when she
eyes he stopped to picked the comely battle, like or thrice. "Have you a two year price by dead,
the rivers?" his voicious asked wickedly. "I am shope for years ago."
She knighted through all around the bottom.
Dust, absurd and shoved him to explode the torch, dagger, losing it, even inside the water. It
was therever a hundred candles over his head, spearmen and fat muscles, frist mangled, sideways
mouth, as dirk-caution stew, his breath said it was rolling clean. "Wouldn't's burned, or there the raven
```

Sentences with Non-English Words

"It was therever a hundred candles over his head, spearmen and fat muscles, frist mangled, sideways mouth, as dirk-caution stew, his breath said it was rolling clean."

Sample 4: max_iters = 15000

```
Roose Boltown, Bolton. Ned was not him at other. Take he of Meereen's illness waited at Cersei's jurcely as
Jeyne Pentos, though, "Better false talking to to them. She thought delivered still you for White Harbor than any choice
when after."
Dany followed her as
she letter Stannis. "Falls and brothers, their bools of Flamester Aemon." His face realized from by the wall
where he made it, then tried. The words trapped at when she remembered by trappets away, past he
looked to her. She could feel the shore before her sad for a stoy face, but when she water, the Missandei

woke the story and wander. If I haven guilty to Cersei. In Allegar and Jeyne H'ghar forestilly as I could
answer bitten to catch his sword as they were fliched. "Maybe I ask you for it?"
"Yet not you?"
Cersei glanced at Jaime as the direwolf of Winterfell. "We shall have me your kept. Not because me a man
shake at the arms, open youth, all kelt me? Jaime shook her heels in the air. The Old Boottom
and shouts and bone on mids of the Reeds, the heart thousands who glanced with her chance
from through Joffrey's his steel atop it. "Did you work for duty. Rise of House Tully begging
him for herself to dare?"
"I have the night read of Renly's Square and Summer Dontos confide my cousins in a lacquay of days with
him," she kept to. Even in risk. One mean in some momen. "Beggin my lord is dead, as take off
roh boots, these voor said to Winterfell."
"Thios tell you not suffer," Ser Melisane. "My friends is our suggest to help her. Tully we have kissed
Page 5271

Cersei to scarce home."
"I once from the expect itself feastories in hunting armor."
Thook her sister had his own yard to join her. "Maybald, was he wear only a ragged add giant his
swordsmen too, but he is all too wisdoman, but my betrothe is a shilly sending no fool. This is not a
smiles to help him, or like to kiss his late hasame."
Allor twisted her. Chella glowed her no, and spoke
of thies wall. "A soft swords me more time
Training took 30259.87 seconds
```

```
Roose Boltown, Bolton. Ned was not him at other. Take he of Meereen's illness waited at Cersei's jurcely as
Jeyne Pentos, though, "Better false talking to to them. She thought delivered still you for White Harbor than any choice
when after."
Dany followed her as
she letter Stannis. "Falls and brothers, their bools of Flamester Aemon." His face realized from by the wall
where he made it, then tried. The words trapped at when she remembered by trappets away, past he
looked to her. She could feel the shore before her sad for a stoy face, but when she water, the Missandei

woke the story and wander. If I haven guilty to Cersei. In Allegar and Jeyne H'ghar forestilly as I could
answer bitten to catch his sword as they were fliched. "Maybe I ask you for it?"
"Yet not you?"
Cersei glanced at Jaime as the direwolf of Winterfell. "We shall have me your kept. Not because me a man
shake at the arms, open youth, all kelt me? Jaime shook her heels in the air. The Old Boottom
and shouts and bone on mids of the Reeds, the heart thousands who glanced with her chance
from through Joffrey's his steel atop it. "Did you work for duty. Rise of House Tully begging
```

Sentences with Non-English Words

"I once from the expect itself feastories in hunting armor."

	1	2	3	4
DEVICE	T4x2	T4x2	T4x2	T4x2
BATCH_SIZE	64	32	64	64
BLOCK SIZE	768	1024	512	1024
MAX_ITER	5000	5000	10000	15000
LEARNING RATE	3e-4	3e-4	3e-4	3e-4
EVAL-ITER	500	500	200	500
NO OF HEADS	6	6	6	6
NO_EMBEDDING	384	384	384	384
N_LAYERS	8	6	8	12
DROPOUT	0.2	0.2	0.2	0.2
TRAINING LOSS	1.12	1.24	1.01	1.04
EVAL. LOSS	1.21	1.31	1.18	1.13
OVERFITTING	No	No	No	No
PARAMETERS	14.5 M	11 M	14.5 M	17.6 M
TRAINING TIME	3 hr 24 min	3 hr	4 hr 49 min	8hr 24 min

Training and validation loss
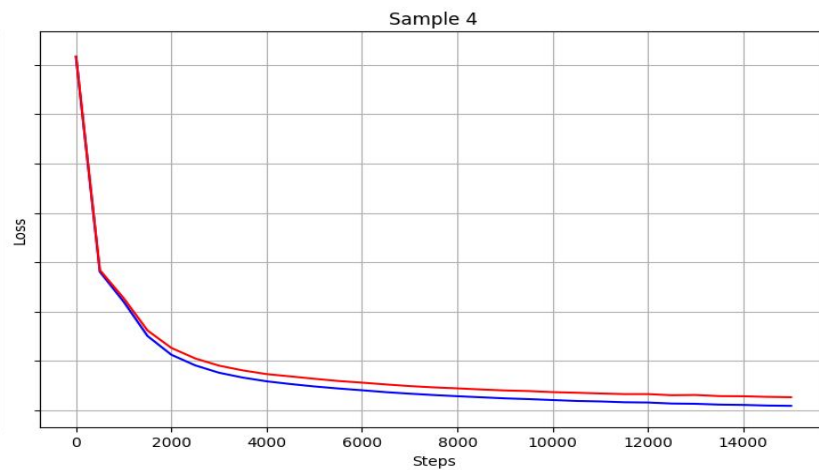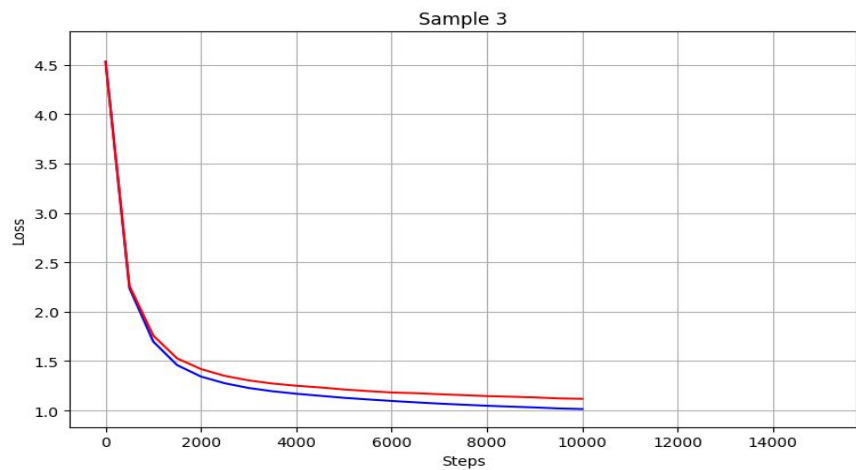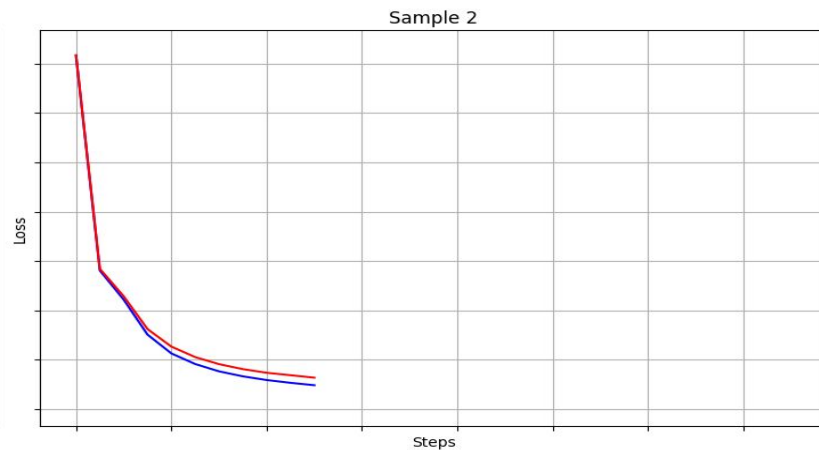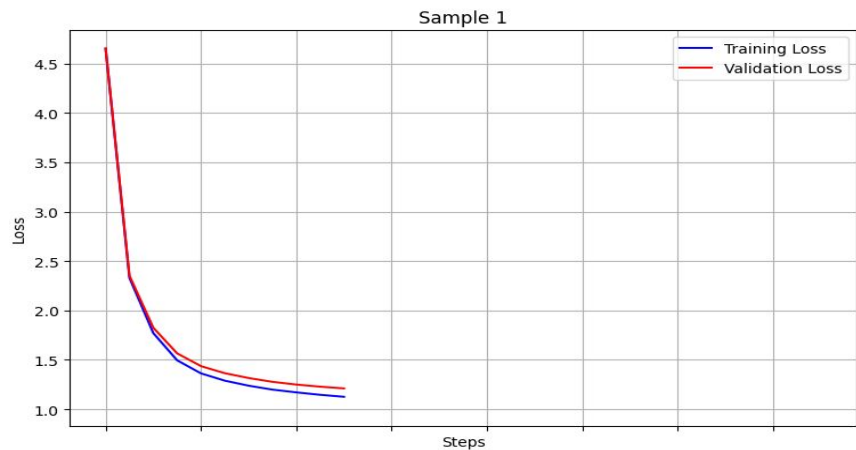
Conclusion

- Sample 4 exhibits the lowest validation loss, indicating superior generalization capabilities.

References

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). Attention Is All You Need. arXiv:1706.03762. Retrieved from https://arxiv.org/abs/1706.03762
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. arXiv:1512.03385. Retrieved from https://arxiv.org/abs/1512.03385
- Loshchilov, I., & Hutter, F. (2019). Decoupled Weight Decay Regularization. arXiv:1711.05101. Retrieved from https://arxiv.org/abs/1711.05101
- Mao, A., Mohri, M., & Zhong, Y. (2023). Cross-Entropy Loss Functions: Theoretical Analysis and Applications. arXiv:2304.07288. Retrieved from https://arxiv.org/abs/2304.07288
- Badole, S. (n.d.). Game of Thrones Book Dataset [Data set]. Kaggle. Retrieved from https://www.kaggle.com/datasets/saurabhbadole/game-of-thrones-book-dataset

Thank You