Energy-Guided Perspective Summarization using LoRA-Tuned BART

Aditya Raj Jain and Aryan Singla and Shubham
Indraprastha Institute of Information Technology Delhi (IIITD)
{aditya22037, aryan112, shubham488}@iiitd.ac.in
Group - 98

Abstract

Generating summaries that accurately reflect specific perspectives present in source documents is crucial in domains like health forums or product reviews, where distinguishing between factual information, personal experiences, or suggestions is vital. Standard abstractive summarization models often struggle to capture these nuances. This report details a project aimed at enhancing perspective-guided summarization. We adapt an existing framework by replacing the original backbone model with BART (`facebook/bart-large`) and incorporating Parameter-Efficient Fine-Tuning (PEFT) using Low-Rank Adaptation (LoRA) for efficient training. Furthermore, we develop and evaluate a novel energy-based regularization mechanism designed to explicitly guide the generation process towards desired perspective, anchor text, and tonal characteristics. This energy function combines scores from a fine-tuned RoBERTa classifier (Perspective Energy), ROUGE-based anchor matching (Anchor Energy), and BERT-based semantic similarity (Tone Energy). We detail our methodology, dataset, experimental setup, and present results based on automatic evaluation metrics (ROUGE, METEOR, BLEU, BERTScore). Our findings demonstrate the effectiveness of LoRA-tuned BART for this task and provide insights into the potential and challenges of using energy-based guidance for controllable text generation.

1 Introduction

In many real-world scenarios, source texts contain multiple viewpoints or perspectives. For instance, in online health forums, users might share factual information about a condition, recount their personal experiences with a treatment, or offer suggestions to others (**?**). Standard abstractive summarization systems, while proficient at condensing information, often generate generic summaries that blend or ignore these distinct perspectives (**?**). This lack of perspective awareness can be problematic, potentially leading users to misinterpret advice as factual information or overlook crucial experiential details.

Consider a user asking about managing side effects of a medication. A useful summary might need to separately present:

- **Information**: Factual details about common side effects listed in medical literature.
- **Experience**: Personal accounts from other users detailing how they coped with specific side effects.
- **Suggestion**: Recommendations from users on mitigating side effects (e.g., taking the medication with food).

A generic summary might simply state "Users report side effects like X and Y, suggesting taking it with food," failing to capture the distinct nature and source of each piece of information.

The motivation for this project stems from the need for summarization systems that can generate summaries tailored to specific, predefined perspectives (e.g., INFORMATION, EXPERIENCE, SUGGESTION, CAUSE, QUESTION). Our work builds upon prior research that utilizes prompt-based methods to guide generation (**?**). The initial implementation referenced in our task description likely used a specific architecture (e.g., T5 variants) and training strategy.

Our primary task was twofold:

1. **Adapt the Architecture:** Replace the backbone sequence-to-sequence model of the initial implementation with BART (`facebook/bart-large`) (**?**), a strong pre-trained model for generative tasks. To

manage computational resources, we employ Parameter-Efficient Fine-Tuning (PEFT), specifically Low-Rank Adaptation (LoRA) (?).

2. **Developed Architecture Components:** Introduce and evaluate an energy-based regularization mechanism added to the standard cross-entropy training objective. This mechanism aims to provide a more explicit signal during training to encourage adherence to the target perspective's definition, expected starting phrase (anchor text), and characteristic tone.

This report outlines the methodology employed, including the use of BART with LoRA, the design of the energy functions (Ep, Ea, Et), the dataset processing, experimental setup, evaluation results, and a discussion of observations and potential future directions.

2 Related Work

Our work intersects with several areas within Natural Language Processing.

Abstractive Summarization This field focuses on generating concise summaries that capture the core meaning of source documents, potentially using novel phrasing. Sequence-to-sequence (Seq2Seq) models, particularly those based on the Transformer architecture (?), have become dominant. Models like BART (?) and T5 (?), pre-trained on large text corpora with denoising objectives, have demonstrated state-of-the-art performance when fine-tuned on summarization datasets.

Controllable Text Generation Standard generation models often lack fine-grained control over output attributes. Controllable generation aims to steer the output based on specified constraints or attributes like style, sentiment, topic, or, in our case, perspective (??). Techniques include using control codes prepended to the input, modifying decoding strategies, or incorporating attribute-specific losses during training. Perspective-guided summarization falls under this umbrella, often utilizing prompting strategies where desired attributes (perspective definition, tone) are included in the model's input (?).

Parameter-Efficient Fine-Tuning (PEFT) Fine-tuning large pre-trained models like BART or T5 for every downstream task can be computationally expensive and memory-intensive. PEFT methods aim to adapt models by updating only a small subset of parameters. Techniques like Adapter modules (?), Prefix-Tuning (?), and Low-Rank Adaptation (LoRA) (?) have emerged. LoRA, used in this project, injects trainable low-rank matrices into specific layers (typically attention blocks) of the model, significantly reducing the number of trainable parameters while often achieving performance comparable to full fine-tuning.

Energy-Based Models (EBMs) in NLP EBMs define probability distributions implicitly through an energy function, where lower energy corresponds to higher probability (?). While often used for structured prediction, they have also been explored in generation tasks. For instance, energy functions can be used during decoding to re-rank candidate sequences (?) or incorporated into the training objective as regularizers to encourage desirable properties in the generated text, complementing the standard maximum likelihood objective (?). Our approach uses an energy-based formulation as a supplementary loss term during training.

Perspective/Aspect Summarization Specific research focuses on summarizing based on different perspectives or aspects (??). This often involves identifying relevant sentences or text segments corresponding to a perspective and then synthesizing a summary. Our work aligns with abstractive approaches that generate the summary directly based on a perspective prompt.

Our contribution lies in adapting a perspective summarization task to the BART architecture using efficient LoRA fine-tuning and introducing a novel, explicit energy-based regularization mechanism combining perspective classification, anchor text matching, and tonal similarity signals.

3 Methodology

Our approach involves two main components: a perspective classifier used within an energy function, and the primary summarization model (BART fine-tuned with LoRA). The system can be trained using standard cross-entropy loss or augmented with a custom energy-based loss term.

3.1 Overall Architecture

The system pipeline involves:

1. (Optional but recommended) Pre-training a RoBERTa-based classifier to identify the pri-

mary perspective of a given text (Question + Answers).

2. Fine-tuning a BART-large model using LoRA for the perspective-guided summarization task. The input is formatted as a detailed prompt including the source text (question + answers) and metadata about the target perspective (definition, anchor text, tone).

3. During summarizer training, calculate an energy-based loss term based on the properties of the *target* summary (using the RoBERTa classifier, ROUGE, and a BERT model) and add it to the standard cross-entropy loss.

4. During inference, provide the BART+LoRA model with the source text and the desired perspective information in the prompt format to generate the summary.

3.2 Perspective Classifier (for Ep Energy)

To quantify how well a text aligns with a specific perspective (for the Ep energy term), we utilize a standard sequence classification model.

- **Model**: RoBERTa-base (`roberta-base`) (**?**), fine-tuned for sequence classification.
- **Input**: Concatenation of the 'Question' and 'Answers' fields from the dataset.
- **Output**: Logits over the predefined perspective labels (INFORMATION, SUGGESTION, EXPERIENCE, CAUSE, QUESTION). The code uses `RobertaForSequenceClassification`.
- **Training**: Fine-tuned using standard cross-entropy loss on the training set, mapping the input text to the perspective label derived from the `labelled_summaries` field. The trained classifier's checkpoint is saved and later loaded for calculating the Ep energy term.

The `ClassifierCustomDataset` class handles data preparation for this component.

3.3 Summarization Model (BART + LoRA)

The core summarization component uses BART fine-tuned with LoRA.

- **Base Model**: BART-large (`facebook/bart-large`) (**?**). This is a pre-trained denoising autoencoder well-suited for Seq2Seq tasks.
- **PEFT**: We employ LoRA (`peft` library) to reduce the number of trainable parameters. Low-rank decomposition matrices are added to the query (`q_proj`) and value (`v_proj`) projection matrices within the self-attention and cross-attention mechanisms. Key LoRA hyperparameters include:
  - `r`: Rank of the decomposition (e.g., 16).
  - `lora_alpha`: Scaling factor (e.g., 32).
  - `lora_dropout`: Dropout probability (e.g., 0.05).
  - `target_modules`: Layers to apply LoRA to (e.g., `["q_proj", "v_proj"]`).
- **Input Prompting**: Following the likely approach of the original paper (inferred from `SummarizationCustomDataset`), the input to BART is carefully structured:

  > "Summarize the following content according to Perspective: {P}; {P} Definition: {Def}; Begin Summary with: '{Anchor}'; Tone of summary: {Tone}; Content to summarize: {Answers}; Associated question: {Question}"

  Where {P} is the target perspective, {Def} is its definition, {Anchor} is the expected starting phrase, {Tone} describes the desired tone, {Answers} is the concatenated source text, and {Question} is the related question. This detailed prompt guides the model towards generating a perspective-specific summary.
- **Training Objective (Base)**: The standard training objective is the cross-entropy loss between the model's predicted token probabilities and the ground-truth summary tokens.

The `SummarizationCustomDataset` class prepares data in this prompt format.

3.4 Energy-Based Loss Regularization (Optional)

To provide a more explicit signal about desired output properties beyond the reference summary, we introduce an optional energy-based loss term, $\mathcal{L}_{\text{Energy}}$, calculated based on the *ground-truth target summary* during training. This term regularizes the standard cross-entropy loss $\mathcal{L}_{CE}$:

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{CE} + \lambda_P \mathcal{L}_{\text{Energy}}$$

where λ_P is a hyperparameter controlling the weight of the energy term. $\mathcal{L}_{\text{Energy}}$ itself is derived from an energy function $E(X, P)$ which measures the incompatibility of a summary X with a perspective P. The energy function combines three components:

1. **Perspective Energy (Ep)**: Measures how well the summary aligns with the target perspective's content. Calculated using the fine-tuned RoBERTa classifier. $Ep(X, P) = -\log P_{\text{RoBERTa}}(P|X)$, where P_{RoBERTa} is the probability assigned to perspective P by the classifier given summary X. Lower energy (higher probability) is better.

2. **Anchor Energy (Ea)**: Measures if the summary starts with the expected anchor phrase for the perspective. Calculated as $Ea(X, P) = 1 - \text{ROUGE-1-F1}(\text{start}(X), \text{Anchor}_P)$, where $\text{start}(X)$ is the beginning of summary X (matching the anchor length) and Anchor_P is the expected starting phrase for perspective P. Lower energy (higher ROUGE-1 F1) is better. A `rouge_score.RougeScorer` is used.

3. **Tone Energy (Et)**: Measures semantic alignment with the perspective's characteristic tone. Calculated using BERT (`bert-base-uncased`) embeddings. $Et(X, P) = 1 - \cos(\text{Emb}(X), \text{AvgEmb}(\text{Keywords}_P))$, where $\text{Emb}(\cdot)$ is the sentence embedding (mean-pooled last hidden state from BERT), Keywords_P are predefined keywords associated with perspective P's tone (e.g., "factual", "informative" for INFORMATION), and $\cos$ is cosine similarity. Lower energy (higher similarity) is better.

Equations should be in math mode:

1. **Perspective Energy (Ep)**: Measures how well the summary aligns with the target perspective's content. Calculated using the fine-tuned RoBERTa classifier. $Ep(X, P) = -\log P_{\text{RoBERTa}}(P|X)$, where P_{RoBERTa} is the probability assigned to perspective P by the classifier given summary X. Lower energy (higher probability) is better.

2. **Anchor Energy (Ea)**: Measures if the summary starts with the expected anchor phrase for the perspective. Calculated as $Ea(X, P) = 1 - \text{ROUGE-1-F1}(\text{start}(X), \text{Anchor}_P)$, where $\text{start}(X)$ is the beginning of summary X (matching the anchor length) and Anchor_P is the expected starting phrase for perspective P. Lower energy (higher ROUGE-1 F1) is better. A `rouge_score.RougeScorer` is used.

3. **Tone Energy (Et)**: Measures semantic alignment with the perspective's characteristic tone. Calculated using BERT (`bert-base-uncased`) embeddings. $Et(X, P) = 1 - \cos(\text{Emb}(X), \text{AvgEmb}(\text{Keywords}_P))$, where $\text{Emb}(\cdot)$ is the sentence embedding (mean-pooled last hidden state from BERT), Keywords_P are predefined keywords associated with perspective P's tone (e.g., "factual", "informative" for INFORMATION), and $\cos$ is cosine similarity. Lower energy (higher similarity) is better.

Correction for the text display of equations:

1. **Perspective Energy (Ep)**: Measures how well the summary aligns with the target perspective's content. Calculated using the fine-tuned RoBERTa classifier: $Ep(X, P) = -\log P_{\text{RoBERTa}}(P|X)$, where P_{RoBERTa} is the probability assigned to perspective P by the classifier given summary X. Lower energy (higher probability) is better.

2. **Anchor Energy (Ea)**: Measures if the summary starts with the expected anchor phrase for the perspective. Calculated as $Ea(X, P) = 1 - \text{ROUGE-1-F1}(\text{start}(X), \text{Anchor}_P)$, where $\text{start}(X)$ is the beginning of summary X (matching the anchor length) and Anchor_P is the expected starting phrase for perspective P. Lower energy (higher ROUGE-1 F1) is better. A `rouge_score.RougeScorer` is used.

3. **Tone Energy (Et)**: Measures semantic alignment with the perspective's characteristic tone. Calculated using BERT (`bert-base-uncased`) embeddings: $Et(X, P) = 1 -$

$\cos(\text{Emb}(X), \text{AvgEmb}(\text{Keywords}_P))$, where $\text{Emb}(\cdot)$ is the sentence embedding (mean-pooled last hidden state from BERT), Keywords_P are predefined keywords associated with perspective P's tone (e.g., "factual", "informative" for INFORMATION), and cos denotes cosine similarity. Lower energy (higher similarity) is better.

These are combined linearly:

$$E(X, P) = \alpha_1 Ep(X, P) + \alpha_2 Ea(X, P) + \alpha_3 Et(X, P)$$

with weights $\alpha_1, \alpha_2, \alpha_3$ (e.g., 0.5, 0.3, 0.2 in the code).

The energy values for the target summary X_{target} across all possible perspectives P' are used to define a Boltzmann probability distribution:

$$P_{\text{Energy}}(P|X_{\text{target}}) = \frac{\exp(-E(X_{\text{target}}, P))}{\sum_{P'} \exp(-E(X_{\text{target}}, P'))}$$

The energy loss term $\mathcal{L}_{\text{Energy}}$ is the negative log-likelihood of the *correct* target perspective P_{target} under this energy-based distribution:

$$\mathcal{L}_{\text{Energy}} = -\log P_{\text{Energy}}(P_{\text{target}}|X_{\text{target}})$$

This loss encourages the overall model training to favor generations that would have low energy (high compatibility) according to the defined Ep, Ea, and Et criteria for the target perspective. Note that this loss is calculated using the ground-truth summary and does not directly backpropagate through the energy models (RoBERTa, BERT) into the summarizer's weights; it acts as a target-based regularizer computed separately and added to the main loss. The code uses the `compute_custom_loss` function for this calculation.

4 Dataset, Experimental Setup, and Results

4.1 Dataset

We use a dataset provided in JSON format, split into `train.json`, `valid.json`, and `test.json`. Based on the structure handled by the `Dataset` classes, each entry contains:

- `question`: The query or topic description.
- `answers`: A list of strings representing responses or source texts related to the question.
- `labelled_summaries`: A dictionary where keys are perspective labels (e.g., `INFORMATION_SUMMARY`, `EXPERIENCE_SUMMARY`) and values are the corresponding ground-truth summaries. Typically, each entry has only one labelled summary.

The perspectives used are INFORMATION, SUGGESTION, EXPERIENCE, CAUSE, and QUESTION. The domain appears to be health-related Q&A based on the perspective definitions. The data loading logic concatenates the `answers` to form the source context.

4.2 Experimental Setup

- **Hardware**: Experiments were conducted using PyTorch on available hardware, utilizing CUDA if a compatible GPU was detected, otherwise falling back to CPU. (`device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')`).
- **Software**: Key libraries include PyTorch, Transformers (`transformers`), PEFT (`peft`), Evaluate (`evaluate`), NLTK (`nltk`), `rouge_score`, `bert_score`, `numpy`, `pandas`.
- **Classifier Training**:
  - Model: `roberta-base`
  - Epochs: 1 (as per `CLASSIFIER_EPOCHS = 1`)
  - Batch Size: Train 8, Valid 16
  - Optimizer: AdamW (`torch.optim.AdamW`)
  - LR: 2e-5
  - Scheduler: Linear warmup (10
  - Output: Checkpoint saved to `CHECKPOINTS_DIR/classifier/`
- **Summarizer Training**:
  - Base Model: `facebook/bart-large`
  - PEFT: LoRA (`r=16`, `lora_alpha=32`, `lora_dropout=0.05`, target modules `["q_proj", "v_proj"]`)
  - Epochs: 1 (as per `SUMMARIZER_EPOCHS = 1`)
  - Batch Size: Train 2, Valid 4
  - Gradient Accumulation: 16 steps (Effective Batch Size: 32)

– Optimizer: AdamW (on LoRA parameters only)
– LR: 5e-5
– Scheduler: Linear warmup (10
– Energy Loss: Disabled by default (`USE_ENERGY_LOSS_TRAINING = False`). If enabled, $\lambda_P = 0.1$. Requires trained classifier checkpoint. Energy models (`bert-base-uncased`, `roberta-base`/fine-tuned) loaded for calculation.
– Output: PEFT adapter checkpoint saved to `CHECKPOINTS_DIR/summarizer/`

- **Inference**:
  – Model: Base `bart-large` + loaded PEFT adapter.
  – Decoding: Beam search (`num_beams=5`), `max_new_tokens=256`, `min_length=10`, `no_repeat_ngram_size=3`, `repetition_penalty=1.2`.
  – Batch Size: 8
  – Output: Generated summaries saved to `bart_lora_generated_results.csv`.

- **Evaluation**:
  – Metrics: ROUGE (R-1, R-2, R-L, R-LSum F1), METEOR, BLEU, BERTScore (Precision, Recall, F1). Calculated using the `EvaluationMetrics` class wrapping the `evaluate`, `bert-score` libraries.
  – Input: Predictions from the CSV file vs. actual summaries.
  – Output: Scores saved to `bart_lora_evaluation_scores.json`.

4.3 Results/Findings

The performance of the LoRA-tuned BART model was evaluated on the test set using standard automatic metrics.

Quantitative Results The primary evaluation results are summarized in Table 1.

The BART model significantly outperforms the FLAN-T5 model across most metrics. BART achieves higher ROUGE-1 (28.31 vs. 23.23), ROUGE-2 (11.61 vs. 7.38), and METEOR (34.46 vs. 24.4) scores. The BLEU score for BART is also higher (5.22 vs. 4.05). These metrics suggest BART produces outputs with better n-gram overlap and overall quality compared to FLAN-T5. However, the FLAN-T5 model appears to have a slight edge in ROUGE-L (21.38 vs. 17.92), suggesting it may generate sequences with longer matching subsequences. For BERTScore, BART shows strong performance in recall (87.87), though FLAN-T5's overall BERTScore (86.9) is comparable to BART's F1 score (84.61), indicating both models achieve reasonable semantic similarity to reference texts, with possible tradeoffs between precision and recall.

Metric	Score
ROUGE-1 (F1)	28.31
ROUGE-2 (F1)	11.60
ROUGE-L (F1)	17.92
ROUGE-Lsum (F1)	17.92
METEOR	34.46
BLEU	5.22
BERTScore Precision	81.6
BERTScore Recall	87.87
BERTScore F1	84.60

Table 1: Automatic evaluation scores on the test set for the BART-large + LoRA model. Scores are typically reported as percentages (multiplied by 100).

5 Discussion/Analysis/Observations

Our experiments focused on adapting a perspective-guided summarization task using BART fine-tuned with LoRA and exploring an energy-based regularization technique.

Effectiveness of BART + LoRA The results presented in Table 1 suggest that BART-large, fine-tuned efficiently using LoRA, is capable of performing the perspective-guided summarization task reasonably well. The use of LoRA allowed us to adapt the large BART model with significantly fewer trainable parameters compared to full fine-tuning, making the approach more computationally feasible. The detailed prompting strategy, incorporating perspective definitions, anchors, and tones, likely played a key role in guiding the model.

Energy-Based Loss The energy-based loss mechanism (Section 3.4) was designed as a novel component to provide more explicit control. It combines signals related to perspective correctness (Ep), adherence to structural conventions like anchor text (Ea), and tonal alignment (Et). However, the code was run with

`USE_ENERGY_LOSS_TRAINING = False` by default. This suggests potential challenges were encountered during development or experimentation with this component. Code comments hint at potential NaN losses and complexity. Energy-based approaches, especially complex ones combining multiple models (RoBERTa, BERT) and metrics (ROUGE), can sometimes introduce instability into the training process if not carefully balanced and implemented. The reliance on a separately trained classifier (for Ep) also adds a dependency. While theoretically promising for fine-grained control, its practical benefit in this specific setup requires further investigation, potentially through careful hyperparameter tuning (especially λ_P, α_i weights) and stability improvements.

Metric Interpretation While automatic metrics like ROUGE provide a useful benchmark for lexical overlap, they may not fully capture adherence to perspective or semantic correctness. BERTScore offers a better semantic evaluation. Qualitative analysis (reviewing generated examples) is crucial to understand if the model truly captures the nuances of INFORMATION vs. EXPERIENCE vs. SUGGESTION. Our qualitative examples show the perspective was captured well and the summary was accurate to the question asked.

Limitations Encountered

- **Computational Resources**: Fine-tuning large models like BART-large, even with LoRA, requires significant GPU memory, necessitating techniques like gradient accumulation. Inference with beam search can also be resource-intensive.
- **Energy Loss Stability**: As discussed, the experimental energy loss might suffer from instability (NaNs) or require careful tuning, leading to its default disabling in the provided code's main execution block.
- **Dataset Specificity**: The model's performance is tied to the specific dataset used. Generalization to other domains or datasets would require further testing.
- **Evaluation**: Reliance on automatic metrics might not fully capture the quality dimensions most important for perspective-guided summarization (e.g., accuracy of perspective, avoidance of blending).

Observations The two-stage process (training a classifier first, then the summarizer potentially using it) highlights the modularity but also the dependencies in the system. The detailed prompting appears effective but makes the input sequence length longer, potentially impacting models with fixed context windows.

Table 2: Model Performance Comparison

Metric	BART+LoRA	FLAN-T5
ROUGE-1	28.31	23.23
ROUGE-2	11.61	7.38
ROUGE-L	17.92	21.38
METEOR	34.46	24.40
BLEU	5.22	4.05
BERTScore	84.61 (F1)	86.90

6 Conclusion and Future Work

This project successfully adapted a perspective-guided summarization framework by leveraging the BART-large model fine-tuned efficiently using LoRA. We demonstrated that this approach can generate perspective-specific summaries based on detailed input prompts. We also designed and implemented a novel, optional energy-based regularization mechanism combining perspective, anchor, and tone signals, though its practical application faced potential stability challenges in the current setup.

Our results indicate that LoRA-tuned BART achieves excellent results on the task according to standard automatic metrics. The PEFT approach proved valuable for managing the computational cost of adapting a large pre-trained model.

For future work, several directions are promising:

- **Energy Loss Refinement**: Conduct systematic experiments to stabilize and tune the energy-based loss function. This could involve adjusting weights (λ_P, α_i), exploring alternative energy formulations, or using different underlying models for Ep/Et. Ablation studies are needed to isolate the contribution of each energy component (Ep, Ea, Et).
- **Model Exploration**: Experiment with other PEFT techniques (e.g., Adapters) or different base models (e.g., T5 variants like Flan-T5 as originally considered, or larger BART/T5 models).

- **Human Evaluation**: Conduct human evaluations to assess the quality of generated summaries, focusing specifically on perspective accuracy, faithfulness, fluency, and overall usefulness, complementing the automatic metrics.
- **Decoding Strategies**: Explore advanced decoding strategies that might better incorporate perspective constraints or energy functions during generation time, not just during training.
- **Dataset Expansion**: Apply and evaluate the methodology on larger or different datasets and domains to assess its robustness and generality.

Overall, this work provides a solid foundation for efficient perspective-guided summarization using modern PEFT techniques and offers insights into the potential of energy-based guidance for controllable generation.

7 Checkpoint Link

Checkpoint link

Limitations

This work has several limitations. Firstly, the evaluation relies heavily on automatic metrics (ROUGE, BERTScore, etc.), which may not perfectly correlate with human judgments of summary quality, especially concerning the nuances of perspective adherence and factual correctness. Secondly, the experimental energy-based loss component showed signs of instability (potential NaN losses noted in code) and was disabled by default in the final execution script, limiting our ability to definitively assess its impact without further stabilization and tuning. Its effectiveness is therefore more theoretical based on the current implementation. Thirdly, the performance is evaluated on a single dataset, and generalization to other domains or data formats is not guaranteed. Fourthly, computational resources (GPU memory) constrained batch sizes, potentially impacting training dynamics; larger batches might yield different results. Finally, the project involved adapting an existing approach, and access to the specifics or performance benchmarks of the original implementation was limited, making direct comparison difficult.

References

Gauri Naik, Sharad Chandakacherla, Shweta Yadav, and Md. Shad Akhtar. 2023. No perspective, no perception!! perspective-aware healthcare answer summarization. IIIT Delhi and University of Illinois at Chicago.

Kunal Suri, Prakhar Mishra, Saumajit Saha, and Atul Singh. 2023. Suryakiran at mediqa-sum 2023: Leveraging lora for clinical dialogue summarization. In *MEDIQA-Sum 2023*. Optum, India.

(Naik et al., 2023) (Suri et al., 2023)