

LEARNING ARGUMENTATIVE SUMMARIZATION WITH ITERATIVE REJECTION SAMPLING

Anonymous authors

Paper under double-blind review

ABSTRACT

Summarization is a fundamental task for evaluating language understanding in both humans and machines, and serves as a crucial tool for information processing in our data-rich world. While Large Language Models (LLMs) have shown significant progress in summarization, they still struggle with domain-specific tasks such as zero-shot medical documentation, legal text, and argumentative summarization. To improve argumentative text understanding and summarization, we propose an iterative fine-tuning framework that trains LLMs on high-quality argument-summary pairs generated by the model itself. These pairs are filtered using similarity scores calculated by comparing reconstructed arguments from summaries with the original arguments, using rejection sampling, without external supervision. Our experiments demonstrate that this method improves argument summarization performance, achieving gains up to 11.88% in BERT F1 similarity scores between reconstructed and original arguments, over the vanilla model without such fine-tuning on a dataset of 200 r/ChangeMyView posts.

1 INTRODUCTION

Summarization, a task to perform on natural language on which unrelated or information of lesser importance is filtered out and only the core points are kept, is one of the human skills that LLMs are trained to simulate. A good grasp of summarization can showcase a solid understanding of the material and human language itself. However, research on Large Language Models (LLMs) has shown inconsistent performance across different domains. They have on-par performance when summarizing news, albeit in different styles (Zhang et al., 2024); when adapted to corresponding domains either by in-context learning or fine-tuning, they could perform even better than professionals on summarizing several types of clinical documents (Van Veen et al., 2024). However, when not adapted for the specific medical domains (Tang et al., 2023) or on legal (Deroy et al., 2024) and various long-form domains (Que et al., 2024), LLMs show significant performance degradation with inconsistencies and hallucinations that make them risky and unfit for real-world deployment.

The importance of LLMs to be good summarizers is multi-faceted. For one, the ability to summarize natural language written by humans could signify a solid understanding of the language, a fundamental goal for the language models (Zhao et al., 2025). LLMs, while initially finetuned on mass corpora from the internet on the task of causal language modeling, associate the definition and concept of summarizing with the behavior of extracting the cardinal information out of convoluted pieces of text. However, to make the language models eventually human-adjacent (OpenAI, 2023) in terms of lingual abilities, they should have comparable performance to human experts in all challenging specific domains. Practically, summarizing online media content, including arguments, is instrumental to moderation, analysis, and recommendation (Gurjar et al., 2025; Schluger et al., 2022; Lee et al., 2020; Bhatia et al., 2014; Hasan et al., 2025). For humans, LLMs can also lighten the cognitive load for absorbing the same information, reducing misunderstandings, and cutting down on time consumption by summarizing long or technical documents for people. This includes summarization for research papers (Langston & Ashford, 2024), legal (Deroy et al., 2024), clinical (Van Veen et al., 2024), and log data produced by automated systems, such as databases (Zhou et al., 2023), allowing people to absorb information and form decisions at faster rates than before.

When summarizing the human language, many aspects or characteristics could make the task extra difficult, for humans and automated systems alike. Text filled with complicated and delicate facts,

054 such as clinical and legal documents, obfuscates summarization as they induce factual errors for
055 any human or automated system that does not understand the material well (Afzal et al., 2024).
056 Documents with tighter-knit logic create points that are closer to the cardinal logic and other side
057 information that has looser ties to that logic, requiring the summarizer to have a good grasp on the
058 cardinal logic (Afzal et al., 2024). We could argue that the intricate-factual summarization tasks
059 could be resolved by injecting required information with Retrieval-Augmented Generation (RAG)
060 systems. Therefore, the challenge to better automated summarization is how we prepare these sys-
061 tems, which are often LLMs, so they could capture the cardinal logic of a given piece of natural
062 language, for the evaluation of which we need one instance of language that only captures the logic-
063 heavy nature and less of the factual side. Argumentative text presents particularly acute challenges
064 for LLM summarization that go beyond the factual accuracy issues found in medical or legal do-
065 mains. Unlike news articles or technical documents that follow predictable structures, argumentative
066 discourse requires models to identify and preserve complex logical relationships, implicit premises,
067 counterarguments, and rhetorical strategies constituting cores of persuasive reasoning.

068 Currently, LLMs face significant hurdles when summarizing argumentative text due to the inherent
069 complexity and unique characteristics of such content. Argumentative dialogues often contain con-
070 tradictory utterances with opposing viewpoints and logical relationships, making them substantially
071 more convoluted to summarize compared to standard text (Zhao et al., 2023; Shakil et al., 2024). U-
072 shaped positional bias, a phenomenon caused by the attention mechanisms of current LLMs where
073 the LLM will place more attention to both ends of the text while ignoring the middle (Vaswani et al.,
074 2017; Hsieh et al., 2024; Yi et al., 2025), was also pronounced and degraded the faithfulness in long-
075 form argumentative summarizations (Elaraby & Litman, 2025a). LLMs also struggle with argument
076 prioritization, as human-written summaries prefer to focus on core arguments while using support-
077 ing examples selectively, whereas LLMs try to list all arguments without distinguishing between
078 central and trivial points (Fan et al., 2024). Finally, the performance of such summarization is not
079 consistent across domains. Legal argument generation represents one of the most problematic do-
080 mains, exposing three critical shortcomings: hallucination, inadequate abstention when arguments
081 are untenable, and poor factor utilization where models fail to incorporate relevant factual elements
082 from provided case materials into their generated arguments (Zhang & Ashley, 2025).

082 To address these limitations, we propose an iterative rejection sampling framework for improving
083 argumentative summarization in Large Language Models. Our approach leverages the model’s own
084 generation capabilities to create high-quality training data through a cyclical process:

085 **Summary generation and reconstruction:** The model generates summaries of argumentative texts
086 and expands these summaries back into full arguments.

087 **Rejection sampling:** We select training pairs using semantic similarity scoring to ensure quality,
088 functioning without external supervision signals.

089 **Iterative fine-tuning:** This self-supervised approach prevents the common pitfall of iterative train-
090 ing systems that drift toward verbose, unfocused outputs.

091 We evaluate our method on one of the Reddit’s r/ChangeMyView dataset ¹, which provides a versa-
092 tile combination of discourse that could help language models to become more generalizable across
093 different argumentative tasks (Dutta et al., 2022). In addition, CMV provides an ideal testbed for ar-
094 gumentative summarization due to its authentic, naturally-occurring argumentative discourse where
095 users present structured reasoning to challenge and defend viewpoints—capturing the complex logi-
096 cal relationships and persuasive strategies that make argument summarization particularly challeng-
097 ing for current LLMs (Syed et al., 2021; Jo et al., 2021; Qin et al., 2025; Elaraby & Litman, 2025b).
098 Through iterative fine-tuning on progressively refined data that prioritizes concise yet comprehen-
099 sive summaries, our method enables the model to better preserve core argumentative elements and
100 logical flow without sacrificing the essential brevity that makes summaries useful.

101
102
103
104
105
106
107

¹<https://huggingface.co/datasets/agentlans/reddit-logic>

2 RELATED WORKS

2.1 NEURAL SUMMARIZATION AND DOMAIN ADAPTATION

The field of neural summarization has been transformed by large-scale Transformer-based sequence-to-sequence models that employ diverse pre-training objectives tailored for text generation tasks (Wang et al., 2020; Raffel et al., 2019; Lewis et al., 2019). Despite these advances, domain adaptation remains a critical challenge for neural summarization systems. Early approaches that combined pre-trained encoders like BERT with decoders trained from scratch suffered from representation mismatches between the two components (Wang et al., 2020; Yang et al., 2019). While unified sequence-to-sequence pre-training approaches like BART, T5, and PEGASUS have addressed this architectural mismatch, significant performance gaps persist when these models are applied to specialized domains without proper adaptation (Wan & Bansal, 2022). Recent work has shown that domain-specific fine-tuning and careful data filtering can substantially improve performance, with filtered datasets leading to better results than training on noisy domain-specific data (Laskar et al., 2023; Duong et al., 2025). The challenge is particularly acute for specialized domains with highly technical terminology, complex argumentative structures, or factual requirements that differ from the general web text typically used in pre-training (Wan & Bansal, 2022).

2.2 ARGUMENTATIVE TEXT UNDERSTANDING

Neural approaches to argumentative text understanding have evolved from traditional rule-based methods to sophisticated transformer-based architectures that can identify and structure argumentative discourse. Recent work has shown that fine-tuning BERT for identifying argumentative units and relationships between them within a text and across texts provides a foundation for understanding complex argumentative structures (Fabbri et al., 2021). More advanced approaches now focus on end-to-end argument graph construction, where structured representations of claims and premises in argumentative text are built by connecting claim and premise argumentative discourse units (Fabbri et al., 2021). These graph-based representations capture the logical flow of arguments more effectively than linear text processing approaches, enabling better preservation of reasoning chains in downstream tasks like summarization. The challenge becomes particularly acute when dealing with debate datasets and persuasive text, where maintaining the integrity of argumentative structure is crucial for generating coherent and logical summaries that preserve the original reasoning patterns.

2.3 EVALUATION OF AUTOMATED SUMMARIES

The evaluation of neural summarization systems has evolved beyond lexical overlap metrics like ROUGE to incorporate semantic similarity measures that better capture summary quality. Current pre-training works in abstractive summarization give more points to the summaries with more words in common with the main text and pay less attention to the semantic similarity between generated sentences and the original document (Salemi et al., 2021). Recent approaches have shown that experimental results using state-of-the-art performance on summarization tasks measured by both ROUGE and BERTScore, with human evaluation demonstrating that using semantic scores significantly improves summarization results (Salemi et al., 2021). Automatic evaluation with BLEU, ROUGE, and METEOR can be complemented by human judges who rate system outputs for relevance and coherence, providing a more comprehensive quality assessment (Hua & Wang, 2020).

Faithfulness assessment has become a critical component of summarization evaluation, addressing the persistent issue that recent models can achieve highly fluent and coherent abstractive summaries, yet the generated summaries often contain factual errors (Wan & Bansal, 2022). Advanced evaluation frameworks now demonstrate that training methods can achieve up to a 14% improvement in faithfulness metrics over existing methods according to automatic evaluation metrics, with evaluations by both GPT-4 and human judges indicating substantially more faithful generations (Duong et al., 2025). Human evaluation protocols have also been refined to assess specific aspects of summary quality, with evaluators scoring factual consistency and coherence on scales that provide more nuanced assessment than binary judgments (Laskar et al., 2023). They enable better understanding of trade-offs between fluency, semantic fidelity, and factual accuracy in summarization systems.

2.4 POLICY OPTIMIZATION METHODS FOR LLM ALIGNMENT

Large Language Models require alignment with human preferences when deployed in real-world settings, leading to the development of various reward-based fine-tuning techniques (Feng et al., 2024). Proximal Policy Optimization (PPO) is a popular approach towards aligning LLMs to human preferences by reward models trained on human preference data (Stiennon et al., 2020). Unlike PPO, Group Relative Policy Optimization (GRPO) operates by evaluating advantages based on the relative performance of multiple sampled outputs for a given prompt, calculating advantages by normalizing rewards relative to group performance statistics such as mean and standard deviation (Kumar et al., 2025; Chen et al., 2025). Reward-ranked Fine-tuning (RAFT) offers a straightforward alternative to complex reinforcement learning approaches by operating through an iterative three-step framework: generating a batch of outputs from the model, scoring outputs using a reward function and filtering high-reward samples, and fine-tuning the model on the filtered data (Dong et al., 2023). Further research reveals that similar, improved rejection-sampling based approaches that remove both the fully correct and incorrect responses also work for GRPO better than the vanilla (Xiong et al., 2025).

3 METHODOLOGY

Many of the challenges outlined in Section 2 are imposed either by the limitations regarding the distribution of topics learned by LLMs, the mechanisms that current LLMs use, or the lack of domain-specific data for adaptation. These challenges, which might have caused the observed hallucination or omission of details, could be mitigated by adapting LLMs to these domains. However, the inability to address and extract the cardinal logic is a more universal flaw that must be addressed. To address the limitation where current LLMs struggle with argumentative summarization due to complex logical relationships and lack of domain-specific data, we propose an iterative rejection sampling framework that addresses these limitations through self-supervised fine-tuning of the model.

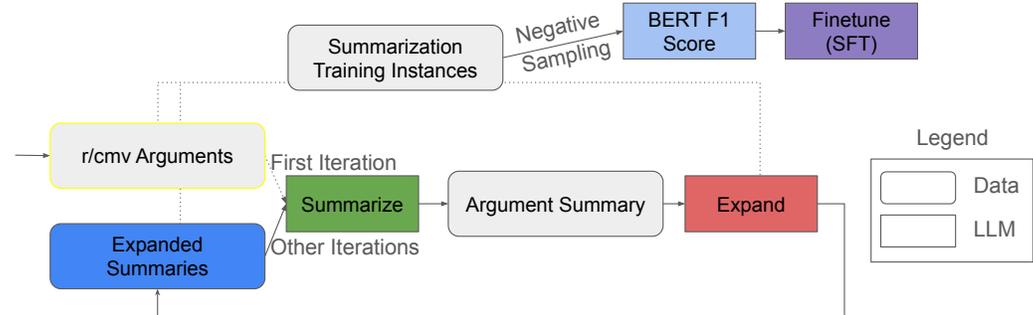


Figure 1: Overview of the iterative rejection sampling framework for argumentative summarization. The process cycles through summary generation, reconstruction, negative sampling, and model fine-tuning across multiple iterations. For each argument or an expanded summary, it was summarized while sampling multiple times, creating diverse summaries that are evaluated and ranked according to the similarity between their expanded summary versions and the original arguments.

3.1 PROBLEM FORMULATION

Let $\mathcal{D} = \{(a_i, s_i)\}_{i=1}^N$ be a dataset of argumentative text-summary pairs, where a_i represents an argumentative text and s_i represents its corresponding summary. Our goal is to learn a summarization function $f_\theta : \mathcal{A} \rightarrow \mathcal{S}$ parameterized by θ , where \mathcal{A} is the space of argumentative texts and \mathcal{S} is the space of summaries. Given the scarcity of high-quality argumentative summarization data, we propose an iterative self-training framework that generates its own training data through reconstruction-based quality assessment. Let $g_\phi : \mathcal{S} \rightarrow \mathcal{A}$ denote a reconstruction function that expands summaries back to argumentative texts. The key insight is that high-quality summaries should enable faithful reconstruction of the original argumentative content.

3.2 ITERATIVE REJECTION SAMPLING FRAMEWORK

Our iterative rejection sampling framework operates through a cyclical process that generates high-quality argumentative text-summary pairs without requiring external supervision, consisting of: summary generation and reconstruction, rejection sampling, and iterative fine-tuning.

3.2.1 SUMMARY GENERATION AND RECONSTRUCTION

At each iteration t , we begin with a set of argumentative texts $\mathcal{A}^{(t)} = \{a_1^{(t)}, a_2^{(t)}, \dots, a_n^{(t)}\}$. For the initial iteration ($t = 0$), this corresponds to the original training corpus. The model $f_{\theta^{(t)}}$ generates k summary candidates for each argumentative text:

$$\mathcal{S}_i^{(t)} = \{s_{i,1}^{(t)}, s_{i,2}^{(t)}, \dots, s_{i,k}^{(t)}\} = \{f_{\theta^{(t)}}(a_i^{(t)})\}_{j=1}^k \quad (1)$$

where each summary $s_{i,j}^{(t)}$ is sampled from the model’s distribution with temperature $T = 1.5$ and minimum probability threshold $p_{min} = 0.1$.

Subsequently, we employ the same model to reconstruct argumentative texts from these summaries using an expansion function $g_{\theta^{(t)}}$:

$$\tilde{a}_{i,j}^{(t)} = g_{\theta^{(t)}}(s_{i,j}^{(t)}) \quad (2)$$

The reconstruction process uses the prompt template: ”Expand the following summaries to one argumentative paragraph with 18 sentences that you might find on reddit’s r/changemyview.”

3.2.2 QUALITY FILTERING VIA RECONSTRUCTION SIMILARITY

To select high-quality training pairs, we implement a rejection sampling mechanism based on semantic similarity between original and reconstructed argumentative texts. We compute BERT-based F1 similarity scores between the original argumentative text $a_i^{(t)}$ and its reconstruction $\tilde{a}_{i,j}^{(t)}$:

$$\text{sim}(a_i^{(t)}, \tilde{a}_{i,j}^{(t)}) = \text{BERT-F1}(a_i^{(t)}, \tilde{a}_{i,j}^{(t)}) \quad (3)$$

We select the top 25% of summary-article pairs based on their reconstruction similarity scores:

$$\mathcal{D}_{selected}^{(t)} = \text{top}_{0.25}\{(a_i^{(t)}, s_{i,j}^{(t)}) : \text{sim}(a_i^{(t)}, \tilde{a}_{i,j}^{(t)})\} \quad (4)$$

This filtering mechanism ensures that selected summaries preserve the core argumentative content and logic of the original texts, enhancing their ability to faithfully reconstruct the source material.

3.2.3 ITERATIVE FINE-TUNING PROCESS

The filtered dataset $\mathcal{D}_{selected}^{(t)}$ is used to fine-tune the model using supervised fine-tuning. The fine-tuning objective minimizes the cross-entropy loss between the model’s predicted summary tokens and the ground truth summaries:

$$\mathcal{L}^{(t)} = - \sum_{(a,s) \in \mathcal{D}_{selected}^{(t)}} \sum_{i=1}^{|s|} \log P_{\theta^{(t)}}(s_i | a, s_{<i}) \quad (5)$$

After fine-tuning, the updated model parameters $\theta^{(t+1)}$ are used for the next iteration. The expanded texts from the current iteration become the input corpus for the subsequent iteration:

$$\mathcal{A}^{(t+1)} = \{\tilde{a}_{i,j}^{(t)} : (a_i^{(t)}, s_{i,j}^{(t)}) \in \mathcal{D}_{selected}^{(t)}\} \quad (6)$$

The motivation of such a process is to utilize the expanded summary as a source for more arguments in order to mitigate overfitting when the model is repeatedly fine-tuned on the same training dataset.

3.2.4 ALGORITHM OVERVIEW

Algorithm 1 presents the complete iterative rejection sampling framework:

Algorithm 1 Iterative Rejection Sampling for Argumentative Summarization

Require: Initial dataset $\mathcal{A}^{(0)}$, model $f_{\theta^{(0)}}$, iterations T , sample size k

Ensure: Fine-tuned model $f_{\theta^{(T)}}$

```

1: for  $t = 0$  to  $T - 1$  do
2:   // Generation Phase
3:   for each  $a_i^{(t)} \in \mathcal{A}^{(t)}$  do
4:     Generate  $k$  summaries:  $S_i^{(t)} = \{f_{\theta^{(t)}}(a_i^{(t)})\}_{j=1}^k$ 
5:     Reconstruct:  $\tilde{A}_i^{(t)} = \{g_{\theta^{(t)}}(s_{i,j}^{(t)})\}_{j=1}^k$ 
6:   end for
7:   // Filtering Phase
8:   Compute similarity scores:  $\text{sim}(a_i^{(t)}, \tilde{a}_{i,j}^{(t)})$  for all pairs
9:   Select top 25%:  $\mathcal{D}_{selected}^{(t)} = \text{top}_{0.25}(\text{sim})$ 
10:  // Training Phase
11:  Fine-tune  $f_{\theta^{(t)}}$  on  $\mathcal{D}_{selected}^{(t)}$  to obtain  $f_{\theta^{(t+1)}}$ 
12:  Update corpus:  $\mathcal{A}^{(t+1)} = \{\tilde{a} : (a, s) \in \mathcal{D}_{selected}^{(t)}\}$ 
13: end for
14: return  $f_{\theta^{(T)}}$ 

```

4 EVALUATION AND RESULTS

4.1 EXPERIMENT SETTINGS

We employ a learning rate of $\eta = 1 \times 10^{-5}$ with the AdamW optimizer using 8-bit quantization for memory efficiency. Training uses a batch size of 4 per device with gradient accumulation steps of 2, effectively yielding a batch size of 8. Each iteration consists of 2 training epochs with 5 warmup steps and weight decay $\lambda = 0.01$ in the fine-tuning phase. We have a total of 4 iterations for each LLM. The loading and fine-tuning of LLMs use a fixed random seed of 3407 for reproducibility. We implement our framework using LoRA (Low-Rank Adaptation) for efficient fine-tuning, with parameters $r = 16$, $\alpha = 16$, targeting projection matrices in attention and feed-forward layers. The maximum sequence length is set to 2048 tokens with a maximum generation length of 1024 tokens.

For computational efficiency, we process data in batches and implement memory management techniques including gradient checkpointing and 4-bit quantization. The framework runs for 8 iterations on each dataset, with early stopping if no suitable training examples are found in any iteration. The BERT scoring model used is DeBERTa-large-MNLI from the DeBERTa family (He et al., 2021).

4.2 RESULTS AND ANALYSIS

As shown in Table 1, during the iterative process, the score had improvements from 0.84% up to 11.88%, even though the score might also trend down as a sign of overfitting outside of the Gemma experiments. The Gemma experiments exhibit a different pattern of overfitting from the beginning, which likely stems from overfitting and requires finding better hyper-parameters, like learning rates.

The larger, commercially available LLMs perform better on this task than the models we have fine-tuned. However, the BERT F1 score is still low, indicating the difficulty of this task. In addition, Qwen/Qwen3-4B-Instruct-2507 performs the best across iterations, with an average F1 score of 0.121 (not shown in Table 1) but reaching a plateau, as it might have been aligned to such tasks and domains before fine-tuning, while Llama-3.2-3B has the largest proportional performance gain.

Therefore, we will further analyze the trajectory of Llama-3.2-3B and especially as compared to 1B. Since the increase of length of the summary is not penalized, LLMs might increase the token length of the summarization between iterations to achieve a better score measured by this framework. However, as shown in Figure 2, Llama-3.2-3B and Llama-3.2-1B perform differently on the

324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377

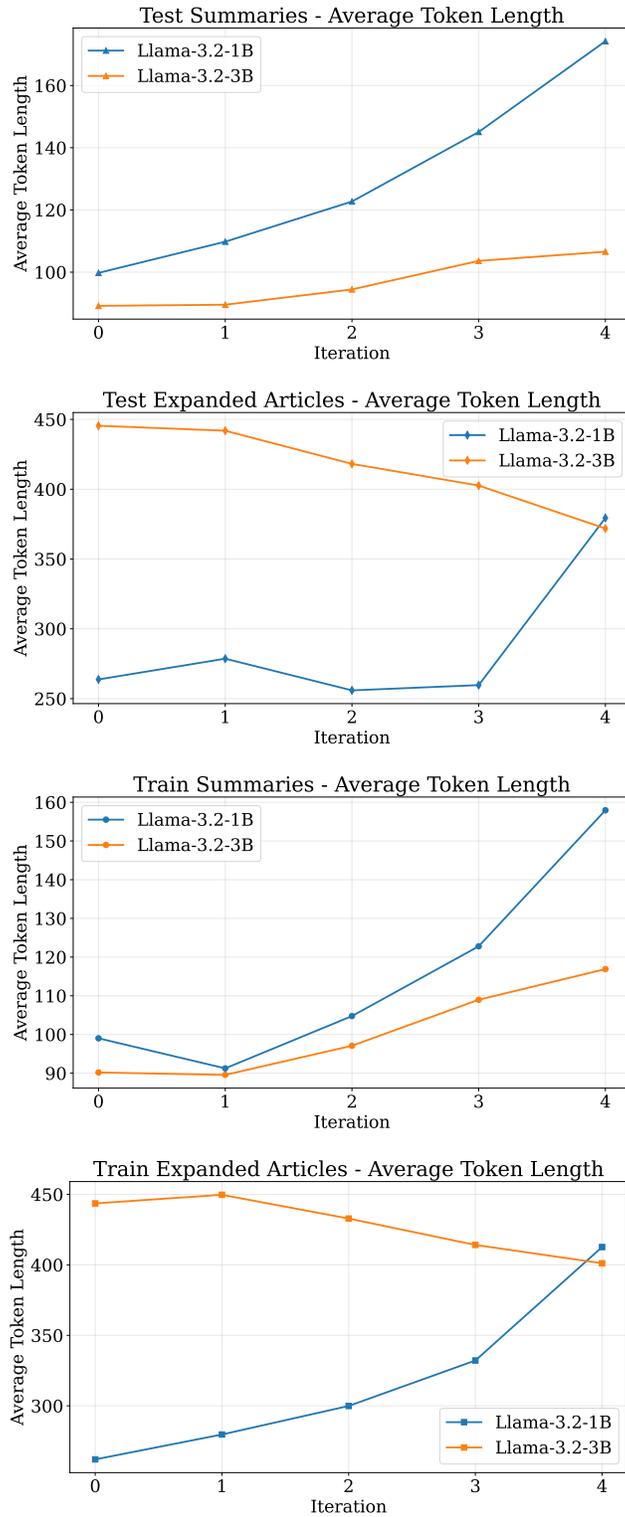


Figure 2: Visualizations on how Llama-3.2-3B and Llama-3.2-1B behaves differently at average summarization and expansion token numbers. While Llama-3.2-3B outperforms Llama-3.2-1B in terms of performance gains, it has a more moderate increase on summarization compared to Llama-3.2-1B, and actually exhibits a decrease on the expanded summary, indicating that the knowledge on summarization of the arguments might have transferred onto the expansion task.

Table 1: Average BERT-F1 scores across the test set. The Gemma experiments are cut-off as they have overfitted and would require changes to the hyper-parameters, such as the learning rate.

Model	Iterations				
	0	1	2	3	4
meta-llama/Llama-3.2-1B-Instruct	0.064	0.073	0.068	0.057	0.041
meta-llama/Llama-3.2-3B-Instruct	0.101	0.102	0.105	0.112	0.113
Google/Gemma-3-1B-it	0.090	0.089	0.089		
Google/Gemma-3-4B-it	0.103	0.098	0.097		
Qwen/Qwen3-4B-Instruct-2507	0.119	0.118	0.120	0.119	0.120
Deepseek/chat (DeepSeek-V3.1)	0.123				
OpenAI/gpt-4o-mini	0.105				
Claude/sonnet-4-20250514	0.120				

summarization task. For summaries (top row), both the training and test sets show a steady increase in length as iterations progress. The growth is particularly pronounced for the 1B model, which increases from around 100 tokens to nearly 160, while the 3B model increases more moderately and stabilizes around 90–115 tokens. This suggests that smaller models are more prone to “length inflation” in iterative summarization, whereas larger models remain more stable. For expanded articles (bottom row), the two models diverge. The 1B model consistently produces longer outputs with each round, increasing from roughly 260 to more than 400 tokens. In contrast, the 3B model starts with longer generations (~ 450 tokens) but gradually contracts to around 390 tokens as iterations proceed. This might indicate that in the case of Llama-3.2-3B, there might be a more significant and effective learning of the summarization of arguments that is transferred to the task of expansion, and the model also adapted better the summarization task than Llama-3.2-1B.

5 CONCLUSION

In this work, we introduced an iterative rejection sampling framework that addresses the significant challenges Large Language Models (LLMs) face when summarizing argumentative text. Our approach leverages a self-supervised cyclical process that generates high-quality argument-summary pairs without requiring external human annotation, addressing the fundamental problem of domain adaptation for argumentative discourse.

Our key contributions include: (1) a novel reconstruction-based quality assessment mechanism that uses semantic similarity between original and reconstructed arguments to filter training data, (2) an iterative fine-tuning framework that progressively improves model performance through self-generated examples, and (3) comprehensive evaluation demonstrating improvements up to 11.88% in BERT F1 similarity scores on the r/ChangeMyView dataset.

The experimental results reveal important insights into model behavior during iterative training. While our framework consistently improves performance for larger models like Llama-3.2-3B, we observed that smaller models (Llama-3.2-1B) are more susceptible to length inflation, producing increasingly verbose outputs that may compromise the core objective of concise summarization. This suggests that model capacity plays a crucial role in maintaining the balance between comprehensiveness and brevity in argumentative summarization.

Our analysis of token length dynamics shows that the Llama-3.2-3B model not only outperforms smaller variants but also demonstrates more stable behavior, even showing contraction in expanded text length over iterations. This indicates that larger models may be better at transferring learned summarization capabilities to related tasks like argument expansion, suggesting a more robust understanding of argumentative structure.

The framework’s self-supervised nature makes it particularly valuable for domains where high-quality training data is scarce. By eliminating the need for human-annotated argument-summary pairs, our approach enables scalable adaptation to specialized argumentative domains while maintaining quality through reconstruction-based filtering.

Limitations and Future Work. Despite these promising results, several limitations warrant attention. The framework’s tendency toward length inflation in smaller models suggests the need

for explicit length constraints or penalties in the training objective. Additionally, our evaluation is currently limited to a single dataset (r/ChangeMyView), and broader evaluation across diverse argumentative domains would strengthen the generalizability claims.

Future research directions include investigating alternative quality metrics beyond reconstruction similarity, exploring the framework’s applicability to other specialized domains requiring logical reasoning, and developing methods to mitigate length inflation while preserving content quality. The integration of explicit logical structure preservation mechanisms could further enhance the framework’s ability to maintain argumentative coherence.

Our work demonstrates that self-supervised iterative training can meaningfully improve argumentative summarization capabilities in LLMs, providing a scalable path toward better domain adaptation for complex reasoning tasks. Our framework is characterized by its use of a pair of tasks that counter each other out, summarization and expansion, to obtain a self-supervised signal for iterative fine-tuning. The same concept should be applicable to a broader range of such task pairs, such as creating and solving programming problems.

REFERENCES

- Anum Afzal, Ribin Chalumattu, Florian Matthes, and Laura Mascarell. AdaptEval: Evaluating large language models on domain adaptation for text summarization. In Sachin Kumar, Vidhisha Balachandran, Chan Young Park, Weijia Shi, Shirley Anugrah Hayati, Yulia Tsvetkov, Noah Smith, Hannaneh Hajishirzi, Dongyeop Kang, and David Jurgens (eds.), *Proceedings of the 1st Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual (CustomNLP4U)*, pp. 76–85, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.customnlp4u-1.8. URL <https://aclanthology.org/2024.customnlp4u-1.8/>.
- Sumit Bhatia, Prakhar Biyani, and Prasenjit Mitra. Summarizing online forum discussions – can dialog acts of individual messages help? pp. 2127–2131, 01 2014. doi: 10.3115/v1/D14-1226.
- Xiwen Chen, Wenhui Zhu, Peijie Qiu, Xuanzhao Dong, Hao Wang, Haiyu Wu, Huayu Li, Aristeidis Sotiras, Yalin Wang, and Abolfazl Razi. Dra-grpo: Exploring diversity-aware reward adjustment for rl-zero-like training of large language models. *ArXiv*, abs/2505.09655, 2025. URL <https://api.semanticscholar.org/CorpusID:278636216>.
- Aniket Deroy, Kripabandhu Ghosh, and Saptarshi Ghosh. Applicability of large language models and generative models for legal case judgement summarization. *Artificial Intelligence and Law*, July 2024. ISSN 1572-8382. doi: 10.1007/s10506-024-09411-z. URL <https://doi.org/10.1007/s10506-024-09411-z>.
- Hanze Dong, Wei Xiong, Deepanshu Goyal, Yihan Zhang, Winnie Chow, Rui Pan, Shizhe Diao, Jipeng Zhang, KaShun SHUM, and Tong Zhang. RAFT: Reward ranked finetuning for generative foundation model alignment. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL <https://openreview.net/forum?id=m7p507zblY>.
- Song Duong, Florian Le Bronnec, Alexandre Allauzen, Vincent Guigue, Alberto Lumbreras, Laure Soulier, and Patrick Gallinari. Scope: A self-supervised framework for improving faithfulness in conditional text generation. *ArXiv*, abs/2502.13674, 2025. URL <https://api.semanticscholar.org/CorpusID:276450028>.
- Subhabrata Dutta, Jeevesh Juneja, Dipankar Das, and Tanmoy Chakraborty. Can unsupervised knowledge transfer from social discussions help argument mining? In *Annual Meeting of the Association for Computational Linguistics*, 2022. URL <https://api.semanticscholar.org/CorpusID:247627719>.
- Mohamed Elaraby and Diane Litman. Arc: Argument representation and coverage analysis for zero-shot long document summarization with instruction following llms, 2025a. URL <https://arxiv.org/abs/2505.23654>.
- Mohamed S. Elaraby and Diane J. Litman. Arc: Argument representation and coverage analysis for zero-shot long document summarization with instruction following llms. *ArXiv*, abs/2505.23654, 2025b. URL <https://api.semanticscholar.org/CorpusID:278996263>.

- 486 Alexander R. Fabbri, Faiaz Rahman, Imad Rizvi, Borui Wang, Haoran Li, Yashar Mehdad, and
487 Dragomir R. Radev. Convosumm: Conversation summarization benchmark and improved ab-
488 stractive summarization with argument mining. *ArXiv*, abs/2106.00829, 2021. URL <https://api.semanticscholar.org/CorpusID:235294200>.
489
- 490 Yuchen Fan, Xin Zhong, Yazhe Wan, Chengsi Wang, Haonan Cheng, Gaoche Wu, Ning Ding, and
491 Bowen Zhou. Eva-score: Evaluating abstractive long-form summarization on informativeness
492 through extraction and validation, 2024. URL <https://arxiv.org/abs/2407.04969>.
493
- 494 Qizhang Feng, Siva Rajesh Kasa, Hyokun Yun, Choon Hui Teo, and S. Bodapati. Exposing
495 privacy gaps: Membership inference attack on preference data for llm alignment. In *Inter-
496 national Conference on Artificial Intelligence and Statistics*, 2024. URL [https://api.
497 semanticscholar.org/CorpusID:271064786](https://api.semanticscholar.org/CorpusID:271064786).
498
- 499 Omkar Gurjar, Agam Goyal, and Eshwar Chandrasekharan. Argcmv: An argument summarization
500 benchmark for the llm-era, 2025. URL <https://arxiv.org/abs/2508.19580>.
501
- 502 Emrul Hasan, Mizanur Rahman, Chen Ding, Jimmy Xiangji Huang, and Shaina Raza. Review-
503 based recommender systems: A survey of approaches, challenges and future perspectives. *ACM
504 Comput. Surv.*, 58(1), September 2025. ISSN 0360-0300. doi: 10.1145/3742421. URL <https://doi.org/10.1145/3742421>.
505
- 506 Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. Deberta: Decoding-enhanced bert
507 with disentangled attention, 2021. URL <https://arxiv.org/abs/2006.03654>.
508
- 509 Cheng-Yu Hsieh, Yung-Sung Chuang, Chun-Liang Li, Zifeng Wang, Long Le, Abhishek Kumar,
510 James Glass, Alexander Ratner, Chen-Yu Lee, Ranjay Krishna, and Tomas Pfister. Found
511 in the middle: Calibrating positional attention bias improves long context utilization. In Lun-
512 Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Findings of the Association for Computa-
513 tional Linguistics: ACL 2024*, pp. 14982–14995, Bangkok, Thailand, August 2024. Association
514 for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.890. URL <https://aclanthology.org/2024.findings-acl.890/>.
515
- 516 Xinyu Hua and Lu Wang. Pair: Planning and iterative refinement in pre-trained transformers for long
517 text generation. *ArXiv*, abs/2010.02301, 2020. URL [https://api.semanticscholar.
518 org/CorpusID:222140805](https://api.semanticscholar.org/CorpusID:222140805).
519
- 520 Yohan Jo, Haneul Yoo, Jinyeong Bak, Alice H. Oh, Chris Reed, and Eduard H. Hovy. Knowledge-
521 enhanced evidence retrieval for counterargument generation. *ArXiv*, abs/2109.09057, 2021. URL
522 <https://api.semanticscholar.org/CorpusID:237571419>.
- 523 Komal Kumar, Tajamul Ashraf, Omkar Thawakar, Rao Muhammad Anwer, Hisham Cholakkal,
524 Mubarak Shah, Ming-Hsuan Yang, Phillip H.S. Torr, Salman H. Khan, and Fahad Shahbaz Khan.
525 Llm post-training: A deep dive into reasoning large language models. *ArXiv*, abs/2502.21321,
526 2025. URL <https://api.semanticscholar.org/CorpusID:276725054>.
527
- 528 Oliver Langston and Brian Ashford. Automated summarization of multiple document abstracts and
529 contents using large language models. *Authorea Preprints*, 2024.
530
- 531 Md Tahmid Rahman Laskar, Mizanur Rahman, Israt Jahan, Enamul Hoque, and Jimmy X. Huang.
532 Can large language models fix data annotation errors? an empirical study using debatepe-
533 dia for query-focused text summarization. In *Conference on Empirical Methods in Natural
534 Language Processing*, 2023. URL [https://api.semanticscholar.org/CorpusID:
535 266175937](https://api.semanticscholar.org/CorpusID:266175937).
- 536 Sung-Chul Lee, Jaeyoon Song, Eun-Young Ko, Seongho Park, Jihee Kim, and Juho Kim. Solu-
537 tionchat: Real-time moderator support for chat-based structured discussion. In *Proceedings of
538 the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pp. 1–12, New
539 York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450367080. doi:
10.1145/3313831.3376609. URL <https://doi.org/10.1145/3313831.3376609>.

- 540 Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdel rahman Mohamed,
541 Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence
542 pre-training for natural language generation, translation, and comprehension. In *Annual*
543 *Meeting of the Association for Computational Linguistics*, 2019. URL [https://api.](https://api.semanticscholar.org/CorpusID:204960716)
544 [semanticscholar.org/CorpusID:204960716](https://api.semanticscholar.org/CorpusID:204960716).
- 545 OpenAI. Planning for agi and beyond. [https://openai.com/index/
546 planning-for-agi-and-beyond/](https://openai.com/index/planning-for-agi-and-beyond/), February 2023. Accessed: 2025-09-12.
- 547
548 Wenjuan Qin, Weiran Wang, Yuming Yang, and Tao Gui. Machine-assisted writing evaluation: Ex-
549 ploring pre-trained language models in analyzing argumentative moves. *ArXiv*, abs/2503.19279,
550 2025. URL <https://api.semanticscholar.org/CorpusID:277313716>.
- 551
552 Haoran Que, Feiyu Duan, Liqun He, Yutao Mou, Wangchunshu Zhou, Jiaheng Liu, Wenge Rong,
553 Zekun Moore Wang, Jian Yang, Ge Zhang, Junran Peng, Zhaoxiang Zhang, Songyang Zhang, and
554 Kai Chen. Hellobench: Evaluating long text generation capabilities of large language models,
555 2024. URL <https://arxiv.org/abs/2409.16191>.
- 556
557 Colin Raffel, Noam M. Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena,
558 Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified
559 text-to-text transformer. *J. Mach. Learn. Res.*, 21:140:1–140:67, 2019. URL [https://api.](https://api.semanticscholar.org/CorpusID:204838007)
560 [semanticscholar.org/CorpusID:204838007](https://api.semanticscholar.org/CorpusID:204838007).
- 561
562 Alireza Salemi, Emad Kebriaei, Ghazal Neisi Minaei, and Azadeh Shakery. Arman: Pre-training
563 with semantically selecting and reordering of sentences for persian abstractive summarization.
564 In *Conference on Empirical Methods in Natural Language Processing*, 2021. URL [https://api.](https://api.semanticscholar.org/CorpusID:237453315)
565 [semanticscholar.org/CorpusID:237453315](https://api.semanticscholar.org/CorpusID:237453315).
- 566
567 Charlotte Schluger, Jonathan P. Chang, Cristian Danescu-Niculescu-Mizil, and Karen Levy. Proac-
568 tive moderation of online discussions: Existing practices and the potential for algorithmic support.
569 *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2), November 2022. doi: 10.1145/3555095. URL
<https://doi.org/10.1145/3555095>.
- 570
571 Hassan Shakil, Ahmad Farooq, and Jugal Kalita. Abstractive text summarization: State of the art,
572 challenges, and improvements. *Neurocomput.*, 603(C), October 2024. ISSN 0925-2312. doi:
573 [10.1016/j.neucom.2024.128255](https://doi.org/10.1016/j.neucom.2024.128255). URL [https://doi.org/10.1016/j.neucom.2024.](https://doi.org/10.1016/j.neucom.2024.128255)
574 [128255](https://doi.org/10.1016/j.neucom.2024.128255).
- 575
576 Nisan Stiennon, Long Ouyang, Jeff Wu, Daniel M. Ziegler, Ryan J. Lowe, Chelsea Voss, Alec
577 Radford, Dario Amodei, and Paul Christiano. Learning to summarize from human feed-
578 back. *ArXiv*, abs/2009.01325, 2020. URL [https://api.semanticscholar.org/
CorpusID:221665105](https://api.semanticscholar.org/CorpusID:221665105).
- 579
580 Shahbaz Syed, Khalid Al-Khatib, Milad Alshomary, Henning Wachsmuth, and Martin Potthast.
581 Generating informative conclusions for argumentative texts. In *Findings*, 2021. URL [https://api.](https://api.semanticscholar.org/CorpusID:235294159)
582 [semanticscholar.org/CorpusID:235294159](https://api.semanticscholar.org/CorpusID:235294159).
- 583
584 Liyan Tang, Zhaoyi Sun, Betina Idnay, Jordan G. Nestor, Ali Soroush, Pierre A. Elias, Ziyang Xu,
585 Ying Ding, Greg Durrett, Justin F. Rousseau, Chunhua Weng, and Yifan Peng. Evaluating large
586 language models on medical evidence summarization. *npj Digital Medicine*, 6(1):158, August
587 2023. ISSN 2398-6352. doi: 10.1038/s41746-023-00896-7. URL [https://doi.org/10.](https://doi.org/10.1038/s41746-023-00896-7)
[1038/s41746-023-00896-7](https://doi.org/10.1038/s41746-023-00896-7).
- 588
589 Dave Van Veen, Cara Van Uden, Louis Blankemeier, Jean-Benoit Delbrouck, Asad Aali, Chris-
590 tian Bluethgen, Anuj Pareek, Malgorzata Polacin, Eduardo Pontes Reis, Anna Seehofnerová,
591 Nidhi Rohatgi, Poonam Hosamani, William Collins, Neera Ahuja, Curtis P. Langlotz, Jason
592 Hom, Sergios Gatidis, John Pauly, and Akshay S. Chaudhari. Adapted large language mod-
593 els can outperform medical experts in clinical text summarization. *Nature Medicine*, 30(4):
1134–1142, April 2024. ISSN 1546-170X. doi: 10.1038/s41591-024-02855-5. URL <https://doi.org/10.1038/s41591-024-02855-5>.

- 594 Ashish Vaswani, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez,
595 Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Neural Information Processing*
596 *Systems*, 2017. URL <https://api.semanticscholar.org/CorpusID:13756489>.
597
- 598 David Wan and Mohit Bansal. Factpegasus: Factuality-aware pre-training and fine-tuning for ab-
599 stractive summarization. In *North American Chapter of the Association for Computational Lin-*
600 *guistics*, 2022. URL <https://api.semanticscholar.org/CorpusID:248810934>.
- 601 Zhengjue Wang, Zhibin Duan, Hao Zhang, C. Wang, Long Tian, Bo Chen, and Mingyuan
602 Zhou. Friendly topic assistant for transformer based abstractive summarization. In *Confer-*
603 *ence on Empirical Methods in Natural Language Processing*, 2020. URL [https://api.](https://api.semanticscholar.org/CorpusID:226262313)
604 [semanticscholar.org/CorpusID:226262313](https://api.semanticscholar.org/CorpusID:226262313).
- 605 Wei Xiong, Jiarui Yao, Yuhui Xu, Bo Pang, Lei Wang, Doyen Sahoo, Junnan Li, Nan Jiang,
606 Tong Zhang, Caiming Xiong, and Hanze Dong. A minimalist approach to llm reasoning:
607 from rejection sampling to reinforce. *ArXiv*, abs/2504.11343, 2025. URL [https://api.](https://api.semanticscholar.org/CorpusID:277786703)
608 [semanticscholar.org/CorpusID:277786703](https://api.semanticscholar.org/CorpusID:277786703).
- 609
610 Zhilin Yang, Zihang Dai, Yiming Yang, Jaime G. Carbonell, Ruslan Salakhutdinov, and Quoc V.
611 Le. Xlnet: Generalized autoregressive pretraining for language understanding. *ArXiv*,
612 abs/1906.08237, 2019. URL [https://api.semanticscholar.org/CorpusID:](https://api.semanticscholar.org/CorpusID:195069387)
613 [195069387](https://api.semanticscholar.org/CorpusID:195069387).
- 614 Zihao Yi, Delong Zeng, Zhenqing Ling, Haohao Luo, Zhe Xu, Wei Liu, Jian Luan, Wanxia
615 Cao, and Ying Shen. Attention basin: Why contextual position matters in large language
616 models. *ArXiv*, abs/2508.05128, 2025. URL [https://api.semanticscholar.org/](https://api.semanticscholar.org/CorpusID:280546653)
617 [CorpusID:280546653](https://api.semanticscholar.org/CorpusID:280546653).
- 618
619 Li Zhang and Kevin D. Ashley. Mitigating manipulation and enhancing persuasion: A reflective
620 multi-agent approach for legal argument generation. *ArXiv*, abs/2506.02992, 2025. URL [https://](https://api.semanticscholar.org/CorpusID:279118723)
621 api.semanticscholar.org/CorpusID:279118723.
- 622 Tianyi Zhang, Faisal Ladhak, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori B.
623 Hashimoto. Benchmarking large language models for news summarization. *Transactions of the*
624 *Association for Computational Linguistics*, 12:39–57, 2024. doi: 10.1162/tacl.a.00632. URL
625 <https://aclanthology.org/2024.tacl-1.3/>.
- 626 Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min,
627 Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen,
628 Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and
629 Ji-Rong Wen. A survey of large language models, 2025. URL [https://arxiv.org/abs/](https://arxiv.org/abs/2303.18223)
630 [2303.18223](https://arxiv.org/abs/2303.18223).
- 631
632 Xiutian Zhao, Ke Wang, and Wei Peng. ORCHID: A Chinese debate corpus for target-independent
633 stance detection and argumentative dialogue summarization. In Houda Bouamor, Juan Pino, and
634 Kalika Bali (eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Lan-*
635 *guage Processing*, pp. 9358–9375, Singapore, December 2023. Association for Computational
636 Linguistics. doi: 10.18653/v1/2023.emnlp-main.582. URL [https://aclanthology.org/](https://aclanthology.org/2023.emnlp-main.582/)
637 [2023.emnlp-main.582/](https://aclanthology.org/2023.emnlp-main.582/).
- 638
639 Xuanhe Zhou, Guoliang Li, Zhaoyan Sun, Zhiyuan Liu, Weize Chen, Jianming Wu, Jiesi Liu, Ruo-
640 hang Feng, and Guoyang Zeng. D-bot: Database diagnosis system using large language models,
2023. URL <https://arxiv.org/abs/2312.01454>.

642 6 APPENDIX

644 6.1 ETHICS STATEMENT

645
646 This work investigates iterative summarization and expansion of argumentative text using Large
647 Language Models (LLMs). We primarily rely on the publicly available Reddit ChangeMyView
dataset following the license (CC-BY-4.0). This dataset has been widely adopted in prior research

648 on argumentation and reasoning. To minimize risks to privacy, we use only de-identified text that
649 has already been released for research purposes, and no attempt is made to deanonymize or trace
650 content back to individual users.

651 The models used in our experiments include both open-source LLMs (via Unsloth) and hosted APIs
652 (Anthropic Claude, OpenAI GPT, DeepSeek). We follow the provider’s terms of service and no
653 sensitive or private data are submitted to the APIs. Generated outputs are used strictly for research
654 evaluation and not for deployment in real-world decision-making systems.

655 Potential ethical concerns include the amplification of biases present in the training data and the
656 possibility of producing harmful or misleading content. To mitigate these risks, we limit our experi-
657 ments to summarization and expansion tasks within a controlled research setting and only evaluation
658 metrics such as BERTScore for quality control. The results are not intended for direct downstream
659 applications where misinformation or bias could cause harm.

660 Our study complies with the the principles of research integrity: we document model configurations,
661 hyperparameters, and evaluation protocols to ensure reproducibility. In this study, no human sub-
662 jects, private information, or sensitive demographic characteristics are involved and therefore IRB
663 approval was not required.

664 665 6.2 LARGE LANGUAGE MODELS USAGE 666

667 The background information, the ideation process, and the writing of this paper were done with the
668 aid of LLMs capable of searching for information online. The details of the algorithm have been
669 modified and optimized with information from LLMs, and the writing was also improved using
670 LLMs.

671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701