Revelata at the FinLLM Challenge Task: Improving Financial Text Summarization by Restricted Prompt Engineering and Fine-tuning

Ken Kawamura[1], Zeqian Li[1], Chit-Kwan Lin[1], Bradley McDanel[2]
[1]Revelata, Inc., [2]Franklin and Marshall College
Correspondence: ken@revelata.com

Abstract

Fine-tuning large language models (LLMs) is a promising approach for domain-specific tasks such as financial text summarization. However, the role of prompt design in fine-tuning LLMs, especially on limited training data, remains under-explored. In this paper, we examine the impact of instruction complexity and restricted prompt engineering on fine-tuning instruction-tuned LLMs for financial headline generation. Surprisingly, we find that restricting modifications to a specific portion of the prompt (the "lead-in phrase" for the LLM assistant role) significantly influences the quality of the generated outputs, even outperforming models fine-tuned on more complex instructions. Our results underscore the pivotal role of prompt design in adapting LLMs to specialized domains, and suggest that carefully crafting specific portions of an instruction-tuned LLM's prompt can yield substantial performance gains even with minimal training data.

1 Introduction

Recent advancements in LLMs (Sanh et al., 2021; Brown et al., 2020; et al., 2022, 2023b) are finding wider adoption in finance (Wu et al., 2023; Xie et al., 2024; Yu et al., 2023a), driven in part by shared task challenges such as FinLLM. Here, we discuss our submission to FinLLM Task 2: Financial Text Summarization, in which we investigate how the quality of LLM-generated financial news summaries can be improved by modifying specific parts of conversational prompts when fine-tuning instruction-tuned LLMs.

Surprisingly, when fine-tuning `Meta-Llama-3-8B-Instruct`[1] on a small financial news article dataset (Zhou et al., 2021) with a variety of prompts, the complexity of the prompt instructions given to the model has relatively little impact on summarization performance. In contrast, fine-tuning with the "right" lead-in phrase (i.e., the portion of the prompt immediately before the model begins generating) outperforms models that are fine-tuned on complex instructions, as measured with ROUGE-1 (Lin, 2004).

[1]https://ai.meta.com/blog/meta-llama-3/

2 Related Work

Financial news articles can have critical impacts on the stock market (Tetlock, 2005). Prior research has explored the use of sentiment analysis (Araci, 2019; Sy et al., 2023) on financial news articles to predict stock movement (Zhou et al., 2021; Kalyani et al., 2016; Shah et al., 2018; Mohan et al., 2019). However, sentiment paints news articles with a broad brush, and is less suitable for nuanced analyses. For such analyses, news article summarization plays a critical role.

In general, the goal of news article summarization is to generate a concise text that captures the key points of a longer news article. Prior works have relied on datasets such as the CNN/Daily Mail Corpus (Nallapati et al., 2016) and XSum (Narayan et al., 2018) for evaluations of various summarization methods, which can range from those (Liu and Lapata, 2019) based on BERT (Devlin et al., 2019) to more recent ones based on GPT (Brown et al., 2020) models (Zhang et al., 2023; Goyal et al., 2022). This recent adoption of LLMs (Wu et al., 2023; Yang et al., 2023b; Lee et al., 2024; Yu et al., 2023b) has opened up many possibilities of LLM-based financial news summarization (Xie et al., 2024). In this work, we explore the interaction between prompt design and fine-tuning LLMs for financial news summarization.

3 Task and Dataset

3.1 Task Description

FinLLM Task 2 centers around training an LLM to generate coherent and concise summaries of financial news articles. This task is formulated as

Proceedings of the Joint Workshop of the 8th Financial Technology and Natural Language Processing (FinNLP) and the 1st Agent AI for Scenario Planning (AgentScen), Jeju, South Korea, August 3, 2024

an abstractive summarization problem, where the model is asked to generate a compact summary that captures the essence of the article. In order to guide the model to output such summaries, the participants in the task are allowed to create their own prompt and perform fine-tuning on custom datasets. The organizers detect model cheating when perplexities on training and test data are too close in value, following existing work (Wei et al., 2023) on data leakage.

3.2 Dataset

We were provided a dataset of 8,000 training samples and 2,000 test samples from the EDTSUM dataset[2] (Zhou et al., 2021; Xie et al., 2024). Each sample consisted of two elements: (1) the text of a financial news article from a source such as Businesswire or PRNewswire; and (2) the article's corresponding title, which served as an approximation of an abstractive summary. Thus, the true task could be better described as "title generation", rather than a more broadly-construed summarization task; this distinction informed our prompt design. Lastly, along with the dataset, the organizers provided a baseline prompt template (Xie et al., 2024) (see Appendix A).

3.2.1 Data Cleaning

Through manual inspection of the training dataset we found that there existed titles that were too short or too long to be qualitatively good titles (Table 1). This led us to examine the distribution of the title lengths (Figure 1), which we found to be long-tailed. We reasoned that outliers were likely to harm model training, and decided to remove samples with titles shorter than four words and titles longer than 69 words (99th percentile). We emphasize we only removed extreme outliers; the remaining samples still reflected the broad spectrum of title complexity in original dataset (e.g., we retained the vast majority of titles that contained subtitles and bullets).

We also found duplicate titles and samples where company names used in the title could not be found in their corresponding article. We filtered out these samples since they would likely negatively impact model training as well. Lastly, we found punctuation missing in many of the titles. While missing punctuation does not impact ROUGE-1, we reasoned that such titles would have a lower probability of being generated by any LLM, since such sequences are likely out-of-distribution with respect to the LLM's pre-training data. To avoid this problem, we used GPT-4 (et al., 2023b) to impute title punctuation (see Appendix B). After cleaning, we had 7,803 training samples remaining.

[2]https://huggingface.co/datasets/TheFinAI/edtsum_train

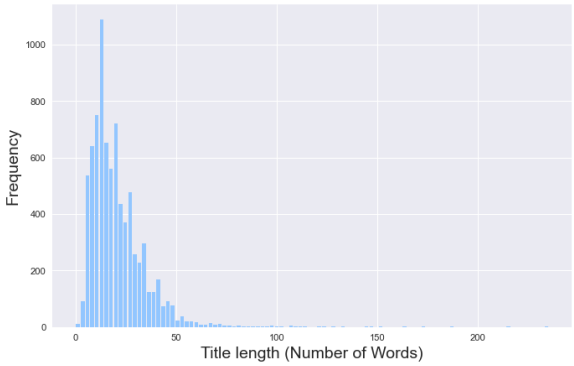

Figure 1: The distribution of news article title length is long-tailed.

4 Approach

We first designed a set of prompts by systematically changing parts of the prompts, and then fine-tuned `Meta-Llama-3-8B-Instruct` on each of these prompts separately. We conjectured that when training data is limited, as in this task, the choice of which part of the prompt to modify could have a large impact on the fine-tuning result.

4.1 Baselines

Our baselines are Gemini Pro (Team et al., 2023) and LLaMA2-70B (et al., 2023a) as reported in FinBen (Xie et al., 2024). They were evaluated in a zero-shot fashion with PIXIU (Xie et al., 2023) using the baseline prompt.

4.2 Prompt Tuning

4.2.1 Parts of a Prompt

LLMs trained on instructions for chat applications switch between the two roles, user and assistant (et al., 2023b,a; Jiang et al., 2023; Roller et al., 2020), allowing a single model to simulate the conversation between two parties and to act as one or the other, depending on the role.

One natural way to engineer a prompt is to refine the instruction given by a user to an assistant *before* the latter's response. In our scenario, this would involve carefully defining the summarization task and giving detailed guidelines for the assistant's response. Figure 2 shows the instruction portion of an example prompt in orange.

Long/Short	Title	Word Count
Short	Annual Financial Report	3
Long	Henry Schein Reports Record First-Quarter 2021 Financial Results from Continuing Operations Total net sales of $2.9 billion up 20.4% versus prior year GAAP diluted EPS from continuing operations of $1.16 versus prior-year GAAP diluted EPS from continuing operations of $0.91 Non-GAAP diluted EPS from continuing operations of $1.24 versus prior-year non-GAAP diluted EPS from continuing operations of $0.94 Reflecting strong first-quarter results, the Company raises guidance for 2021 non-GAAP diluted EPS from continuing operations to be at or above $3.70	80

Table 1: Examples of titles found in the dataset that are either too short or too long to be qualitatively good titles.

Instruction
You are a seasoned marketing PR professional brainstorming a captivating headline for a press release at BUSINESS WIRE and PRNewswire

Write a headline with strong SEO potential. Article: *{Body of News Article}*

Just write a title.

Assistant
Title:

{Title}

Figure 2: An example prompt illustrating the different parts of the prompt. The orange text is the instruction a user provides, and the blue text is a lead-in phrase for the assistant's generation. The violet text is the final title that the model generates. Here, we use a simple instruction and "Title: " as a lead-in phrase.

Another way to tune a prompt is to control the assistant's *lead-in* phrase, just before it generates its response. Figure 2 shows an example of a lead-in phrase in blue. Prior works (Kojima et al., 2022; Wei et al., 2022) have shown that zero-shot LLM predictions can be improved by adding *"Let's think step by step."* to the prompt immediately before the response. Along these lines, we manipulated the conversational *lead-in* phrase of the assistant response; e.g., we controlled the start of the assistant response to be *"Title:"* or *"Here is a headline with strong SEO potential:"*. As instruction-tuned models such as `Meta-Llama-3-8B-Instruct` are trained to be conversational, we found that certain lead-in responses are better suited for chat interactions than others, even when the instruction portion remains the same.

4.2.2 Prompt Design

First, we manually created relatively simple prompts (Figure 2). For example, we simply changed the lead-in phrase from *"Answer:"* in the baseline prompt provided by the organizers to *"Title: "* because it better aligns with the task. From this simple prompt, we crafted additional prompts by modifying the instruction and the lead-in phrase parts of the prompt.

Instruction. We designed four different prompts by replacing the simple instruction with more complex instructions, while keeping the lead-in phrase *"Title: "* fixed. These complex prompts had much more detailed instructions than just asking the model to *"Write a headline with strong SEO potential."*. In Table 2, we present the best-performing complex instruction alongside a baseline instruction and a simple instruction. For a comprehensive list of all complex instructions tried, see Appendix C.

Lead-In Phrase. In total, we devised three lead-in phrases: (1) "Title: ", (2) "" (empty string), and (3) "Here is a headline with strong SEO potential: ". When the lead-in is empty, the model is free to start its response in whatever manner it chooses. Phrase 3 originates from our initial prompt exploration efforts; `Meta-Llama-3-8B-Instruct` sometimes started its generations with this phrase. Since the model already produced this lead-in phrase on its own, we conjectured that it could improve the quality of generated titles and kept it.

4.2.3 Model Fine-tuning

We fine-tuned `Meta-Llama-3-8B-Instruct` hosted on huggingface hub[3] using AutoTrain[4], with default settings. To determine the number of epochs to train the model, we first split the training set 9:1, to create a small validation split. We found that six epochs gave the best ROUGE-1 score on the validation split, and subsequently fine-tuned `Meta-Llama-3-8B-Instruct` for six epochs over the entire dataset, resulting in the model we submitted.

[3]https://huggingface.co/meta-llama/Meta-Llama-3-8B-Instruct

[4]https://github.com/huggingface/autotrain-advanced

Type	Instruction
Baseline	You are given a text that consists of multiple sentences. Your task is to perform abstractive summarization on this text. Use your understanding of the content to express the main ideas and crucial details in a shorter, coherent, and natural sounding text.{Body of News Article}
Simple	You are a seasoned marketing PR professional brainstorming a captivating headline for a press release at BUSINESS WIRE and PRNewswire Write a headline with strong SEO potential. Article: {Body of News Article} Just write a title.
Complex	You are a helpful assistant. You are given a challenge. Below is the text of a press release article. The title has been hidden from you. The goal is to figure out the exact title based on the body of the article. You know that articles such as these can have either simple titles or complex titles that include subtitles in a bullet list. However, it is tricky to determine whether an article should have a simple or complex title, so you need to pay careful attention to the content of the article for any hints or clues. Do your best to write the exact title that was hidden from you. {Body of News Article}

Table 2: Instruction Variations. The baseline instruction is provided by the organizers.

Model	Zero-Shot/Fine-tune	Instruction	Lead-in	ROUGE-1
Baselines				
`Gemini Pro`	Zero-shot	Baseline	"Answer: "	0.39
`LLaMA2-70B`	Zero-shot	Baseline	"Answer: "	0.25
Ours				
`Meta-Llama-3-8B-Instruct`	Zero-shot	Simple	"Title: "	0.402
	Fine-tune	Simple	"Title: "	0.446
	Fine-tune	Complex	"Title: "	0.441
	Fine-tune	Simple	""	0.412
	Fine-tune	Simple	"Here is a headline with strong SEO potential:"	**0.500**

Table 3: Test ROUGE-1 Score on EDTSUM. We only show the best performing result for the prompts with complex instructions.

5 Results

As shown in Table 3, zero-shot title prediction by `Meta-Llama-3-8B-Instruct` with a simple instruction (ROUGE-1: 0.402) already outperforms both `Gemni Pro` (ROUGE-1: 0.39) and `LLaMA2-70B` (ROUGE-1: 0.25). Fine-tuning further improves the ROUGE-1 score from 0.402 to 0.446, using a simple instruction and "Title: " as a lead-in phrase. This result underscores the significance of fine-tuning for adapting foundation models to a specific downstream task.

Surprisingly, we observed that varying the instruction has marginal effect on ROUGE-1 score when the model is fine-tuned. In fact, even a best performing complex instruction with detailed guidelines and a careful task definition (ROUGE-1: 0.441) performed worse than a simple instruction (ROUGE-1: 0.446) by 0.005.

In contrast, varying the lead-in phrase has a substantive impact on performance. Among the fine-tuned models, the model that performed worst (ROUGE-1: 0.412) had an empty lead-in phrase. Meanwhile, by simply replacing "Title: " with "Here is a headline with strong SEO potential: ", and keeping the simple instruction, we achieved our best result (ROUGE-1: 0.500). This suggests that when fine-tuning a model trained for chat applications, tailoring how the assistant starts its conversational response (i.e., the lead-in phrase) is substantially more important than giving complex instructions, if we want the model to achieve high ROUGE-1 performance.

6 Conclusions

Our study highlights the crucial role of prompt engineering in fine-tuning LLMs. Specifically, we find that refining the lead-in phrase of the assistant response significantly improves performance when fine-tuning instruction-tuned models such as `Meta-Llama-3-8B-Instruct`.

However, manually crafting these prompts can be resource-intensive in practical deployments. In future works, we plan to explore automated approaches to optimize lead-in phrases using frameworks such as Optimization by Prompting (OPRO) (Yang et al., 2023a).

References

Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models. *ArXiv*, abs/1908.10063.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeff Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *ArXiv*, abs/2005.14165.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Hugo Touvron et al. 2023a. Llama 2: Open foundation and fine-tuned chat models. *ArXiv*, abs/2307.09288.

Josh Achiam et al. 2023b. GPT-4 Technical Report.

Teven Le Scao et al. 2022. BLOOM: A 176B-Parameter Open-Access Multilingual Language Model. *ArXiv*, abs/2211.05100.

Tanya Goyal, Junyi Jessy Li, and Greg Durrett. 2022. News summarization and evaluation in the era of gpt-3. *ArXiv*, abs/2209.12356.

Albert Qiaochu Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, L'elio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *ArXiv*, abs/2310.06825.

Joshi Kalyani, Prof. H. N. Bharathi, and Prof. Rao Jyothi. 2016. Stock trend prediction using news sentiment analysis. *ArXiv*, abs/1607.01958.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *ArXiv*, abs/2205.11916.

Jean Lee, Nicholas Stevens, Soyeon Caren Han, and Minseok Song. 2024. A survey of large language models in finance (finllms). *ArXiv*, abs/2402.02315.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. *ArXiv*, abs/1908.08345.

Saloni Mohan, Sahitya Mullapudi, Sudheer Sammeta, Parag Vijayvergia, and D. Anastasiu. 2019. Stock price prediction using news sentiment analysis. *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 205–208.

Ramesh Nallapati, Bowen Zhou, Cícero Nogueira dos Santos, Çaglar Gülçehre, and Bing Xiang. 2016. Abstractive text summarization using sequence-to-sequence rnns and beyond. In *Conference on Computational Natural Language Learning*.

Shashi Narayan, Shay B. Cohen, and Mirella Lapata. 2018. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. *ArXiv*, abs/1808.08745.

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2020. Recipes for building an open-domain chatbot. In *Conference of the European Chapter of the Association for Computational Linguistics*.

Victor Sanh, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, Manan Dey, M Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal V. Nayak, Debajyoti Datta, Jonathan D. Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng-Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Févry, Jason Alan Fries, Ryan Teehan, Stella Biderman, Leo Gao, Tali Bers, Thomas Wolf, and Alexander M. Rush. 2021. Multitask prompted training enables zero-shot task generalization. *ArXiv*, abs/2110.08207.

Dev Shah, Haruna Isah, and Farhana H. Zulkernine. 2018. Predicting the effects of news sentiments on the stock market. *2018 IEEE International Conference on Big Data (Big Data)*, pages 4705–4708.

Eugene Sy, Tzu-Cheng Peng, Shih-Hsuan Huang, and Yung-Chun Chang Heng-Yu Lin. 2023. Fine-grained argument understanding with bert ensemble techniques: A deep dive into financial sentiment analysis. In *Taiwan Conference on Computational Linguistics and Speech Processing*.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Paul C. Tetlock. 2005. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Huai hsin Chi, F. Xia, Quoc Le, and Denny Zhou. 2022. Chain of thought prompting elicits reasoning in large language models. *ArXiv*, abs/2201.11903.

Tianwen Wei, Liang Zhao, Lichang Zhang, Bo Zhu, Lijie Wang, Haihua Yang, Biye Li, Cheng Cheng, Weiwei Lü, Rui Hu, Chenxia Li, Liu Yang, Xilin Luo, Xue Gang Wu, Lunan Liu, Wenjun Cheng, Peng Cheng, Jianhao Zhang, Xiaoyu Zhang, Lei Lin, Xiaokun Wang, Yutuan Ma, Chuanhai Dong, Yanqi Sun, Yifu Chen, Yongyi Peng, Xiaojuan Liang, Shuicheng Yan, Han Fang, and Yahui Zhou. 2023. Skywork: A more open bilingual foundation model. *ArXiv*, abs/2310.19341.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. Bloomberggpt: A large language model for finance. *ArXiv*, abs/2303.17564.

Qianqian Xie, Weiguang Han, Zhengyu Chen, Ruoyu Xiang, Xiao Zhang, Yueru He, Mengxi Xiao, Dong Li, Yongfu Dai, Duanyu Feng, Yijing Xu, Haoqiang Kang, Ziyan Kuang, Chenhan Yuan, Kailai Yang, Zheheng Luo, Tianlin Zhang, Zhiwei Liu, Guojun Xiong, Zhiyang Deng, Yuechen Jiang, Zhiyuan Yao, Haohang Li, Yangyang Yu, Gang Hu, Jiajia Huang, Xiao-Yang Liu, Alejandro Lopez-Lira, Benyou Wang, Yanzhao Lai, Hao Wang, Min Peng, Sophia Ananiadou, and Jimin Huang. 2024. The finben: An holistic financial benchmark for large language models. *Preprint*, arXiv:2402.12659.

Qianqian Xie, Weiguang Han, Xiao Zhang, Yanzhao Lai, Min Peng, Alejandro Lopez-Lira, and Jimin Huang. 2023. Pixiu: A large language model, instruction data and evaluation benchmark for finance. *Preprint*, arXiv:2306.05443.

Chengrun Yang, Xuezhi Wang, Yifeng Lu, Hanxiao Liu, Quoc V. Le, Denny Zhou, and Xinyun Chen. 2023a. Large language models as optimizers. *ArXiv*, abs/2309.03409.

Hongyang Yang, Xiao-Yang Liu, and Chris Wang. 2023b. Fingpt: Open-source financial large language models. *ArXiv*, abs/2306.06031.

Yangyang Yu, Haohang Li, Zhi Chen, Yuechen Jiang, Yang Li, Denghui Zhang, Rong Liu, Jordan W. Suchow, and Khaldoun Khashanah. 2023a. Finmem: A performance-enhanced llm trading agent with layered memory and character design. In *AAAI Spring Symposia*.

Yangyang Yu, Haohang Li, Zhi Chen, Yuechen Jiang, Yang Li, Denghui Zhang, Rong Liu, Jordan W. Suchow, and Khaldoun Khashanah. 2023b. Finmem: A performance-enhanced llm trading agent with layered memory and character design. *Preprint*, arXiv:2311.13743.

Tianyi Zhang, Faisal Ladhak, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori Hashimoto. 2023. Benchmarking large language models for news summarization. *Transactions of the Association for Computational Linguistics*, 12:39–57.

Zhihan Zhou, Liqian Ma, and Han Liu. 2021. Trade the event: Corporate events detection for news-based event-driven trading. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2114–2124, Online. Association for Computational Linguistics.

A Baseline Prompt Provided By Organizers

```
You are given a text that consists of multiple sentences. Your task is to perform abstractive summarization on this text. Use your understanding of the content to express the main ideas and crucial details in a shorter, coherent, and natural sounding text.{Body of News Article}
Answer:
```

B GPT-4 Instruction Used to Add Punctuation

```
You are a helpful proofreader. The text below has no period punctuation. Please add it back. Respond with only the updated text.\n\nText:
```

C Prompts with Various Instructions

Instruction
You are a helpful assistant.

You have written a press release for your employer. The text of it follows these instructions. You need to now write a suitable title for the press release. You know that some press releases in the past have had a single title, while others have had a main title accompanied by subtitles. Taking that into account, you should write a title that is appropriate for this article. In any case, do your best to write a title that will make the reader feel interested in reading the article itself, and to ensure that your title has high SEO potential.

Here is the article:

{Body of News Article}

Assistant
Title:

{Title}

Instruction
You are a financial research analyst.

You ran a web scraper script that scrapes press release articles from company and newswire websites. However, there was a bug in the script that accidentally left out all the titles. You know some articles have a single title, while others have a main title followed by subtitles. Knowing this, do your best to write an appropriate title for the scraped article below.

{Body of News Article}

Assistant
Title:

{Title}

Instruction
You are a large language model.

You are given a prompt to generate the title of a financial news article, e.g., a press release. Even though you aren't allowed to know the actual title of the article, the title you generate must have a high ROUGE-1 score with respect to the actual title of the article. Since it's a ROUGE-1 score, you want to maximize the number of words that overlap with the actual title, regardless of the order in which they appear in the title.

Here is the article:

{Body of News Article}

Assistant
Title:

{Title}

Instruction
You are a helpful assistant.

You are given a challenge. Below is the text of a press release article. The title has been hidden from you. The goal is to figure out the exact title based on the body of the article. You know that articles such as these can have either simple titles or complex titles that include subtitles in a bullet list. However, it is tricky to determine whether an article should have a simple or complex title, so you need to pay careful attention to the content of the article for any hints or clues. Do your best to write the exact title that was hidden from you.

{Body of News Article}

Assistant
Title:

{Title}

D Prompts with Various Lead-in Phrases

Instruction
You are a seasoned marketing PR professional brainstorming a captivating headline for a press release at BUSINESS WIRE and PRNewswire

Write a headline with strong SEO potential. Article: *{Body of News Article}*

Just write a title.

Assistant

{Title}

Instruction
You are a seasoned marketing PR professional brainstorming a captivating headline for a press release at BUSINESS WIRE and PRNewswire

Write a headline with strong SEO potential. Article: *{Body of News Article}*

Just write a title.

Assistant
Here is a headline with strong SEO potential:

{Title}