University of Glasgow at the FinLLM Challenge Task: Adapting Llama for Financial News Abstractive Summarization

Lubingzhi Guo , Javier Sanz-Cruzado , Richard McCreadie

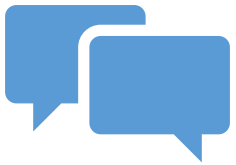
FinNLP-AgentScen 2024

Task 2: Financial Text Summarization

- Objective
  - ➢Use Large Language Models to generate coherent and concise summaries of the given financial news content

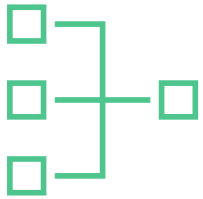
- Research Focus: **Domain Adaptation**
  - ➢LLMs are usually pre-trained on general domain corpora
  - ➢We need effective strategies for adapting LLMs to the financial domain

LLM Adaptation Methods



Chain of Thought Prompting

Design specific prompts to guide the LLM towards generating the desired output and leverage the existing knowledge and reasoning capabilities of the LLM



Adapter Layers

Train lightweight modules inserted into the LLM architecture and adapt the LLM to specific tasks by updating a smaller set of weights

Reinforcement Learning

Further optimize the LLM by maximizing the reward signal derived from feedback to align responses more closely with specific objectives

Chain of Thought Prompting

Approach 1: Chain of Thought

- SumCoT[1]

Article: LONDON--(BUSINESS WIRE)--Technavio has been monitoring the all-season tire market in Europe and it is poised to grow by USD 3.42 billion during 2020-2024, progressing at a CAGR of almost 9% during the forecast period.(…)

Guiding Questions for Prompting

We test two variants

SumCoT

What are the important entities in this document?
What are the important dates in this document?
What events are happening in this document?
What is the result of these events?

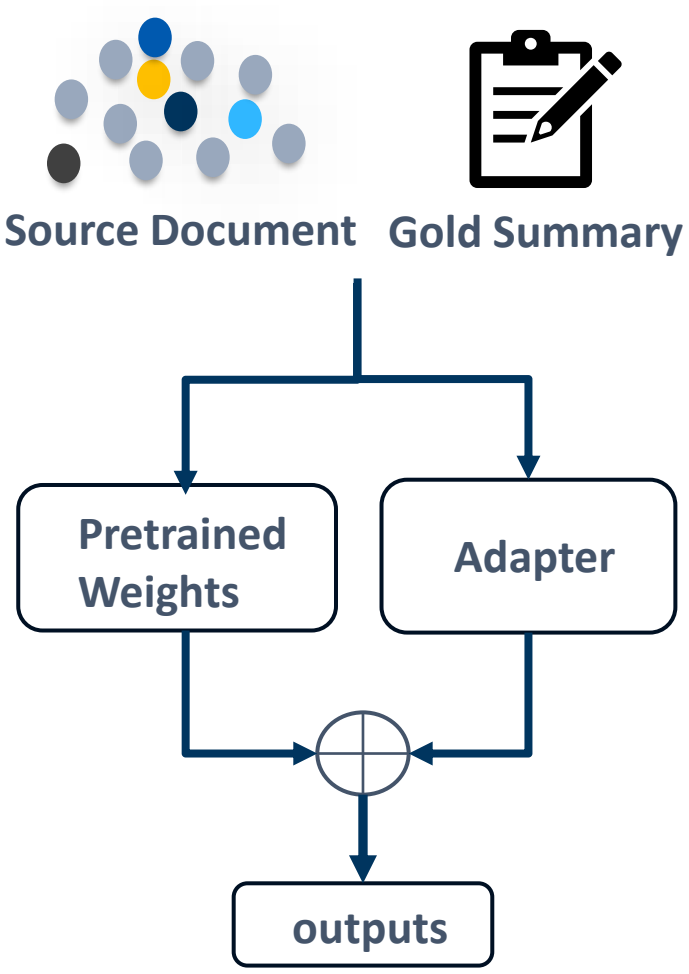
5WCoT

Who is involved? …
What happened? …
Where did it occur? …
When did it take place? …
Why did it happen?...

[1] Element-aware summarization with large language models: Expert-aligned evaluation and chain-ofthought method. In 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 8640–8665, Toronto, Canada. Association for Computational Linguistics.

Tuning the LLM

Approach 2: LLM Adapters



Training Instruction Template

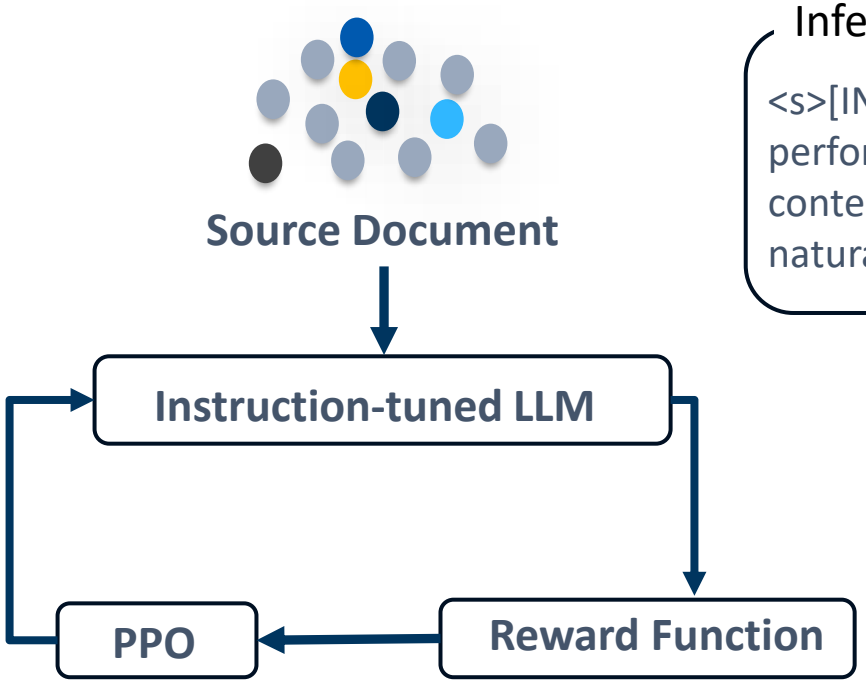
<s>[INST]You are given a text that consists of multiple sentences. Your task is to perform abstractive summarization on this text. Use your understanding of the content to express the main ideas and crucial details in a shorter, coherent, and natural sounding text. Text: {text} Answer:[/INST] {answer}</s>

Configuration

- LLM: 4bit quantized Llama3-8b model
- Adapter: QLoRA finetuning with a rank of 16 across all applicable modules

Reinforcement Learning

Approach 3: Reinforcement Learning



Inference Instruction Template

<s>[INST]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. Text: {text} Answer:[/INST]

Configuration

- LLM: best fine-tuned checkpoint
- Reward Function: averaged score of 4 final performance metrics with length penalty

$$Reward = L\left(\frac{ROUGE_1 + ROUGE_2 + ROUGE_L + BERTScore}{4}\right)$$

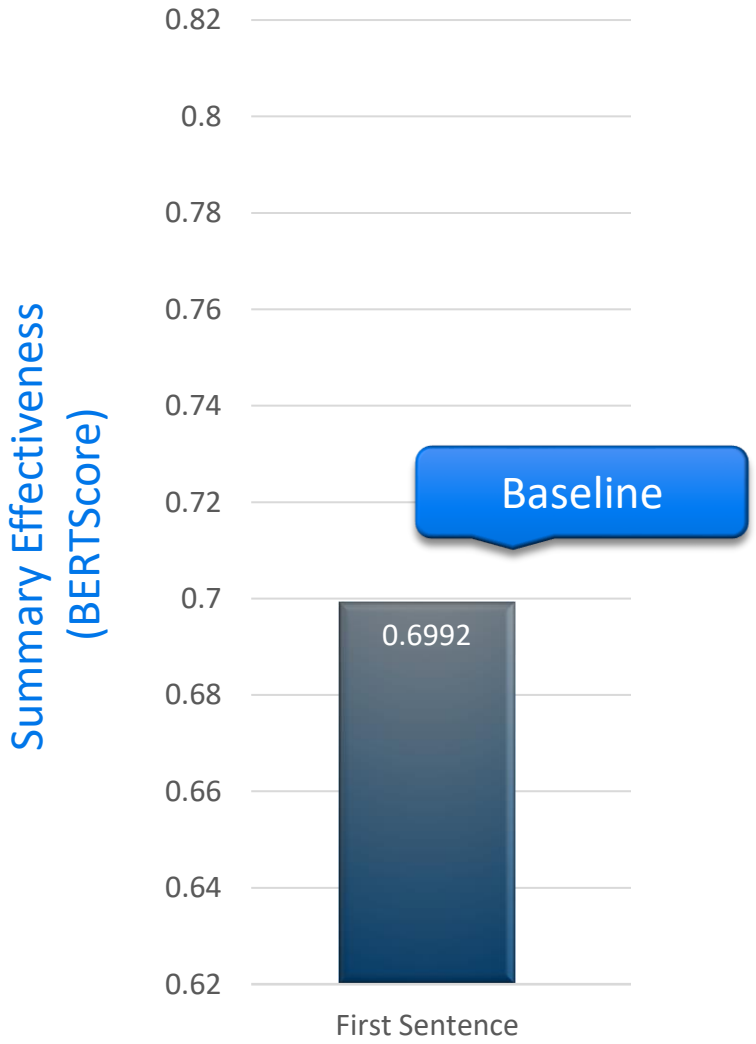
$$L = \begin{cases} e^{1-\frac{c}{r}} & if\ c > r, \\ e^{1-\frac{r}{c}} & if\ c \leq r \\ 0 & if\ c = 0 \end{cases}$$

where c and r represents the length of the generated and reference summary ,respectively

Results

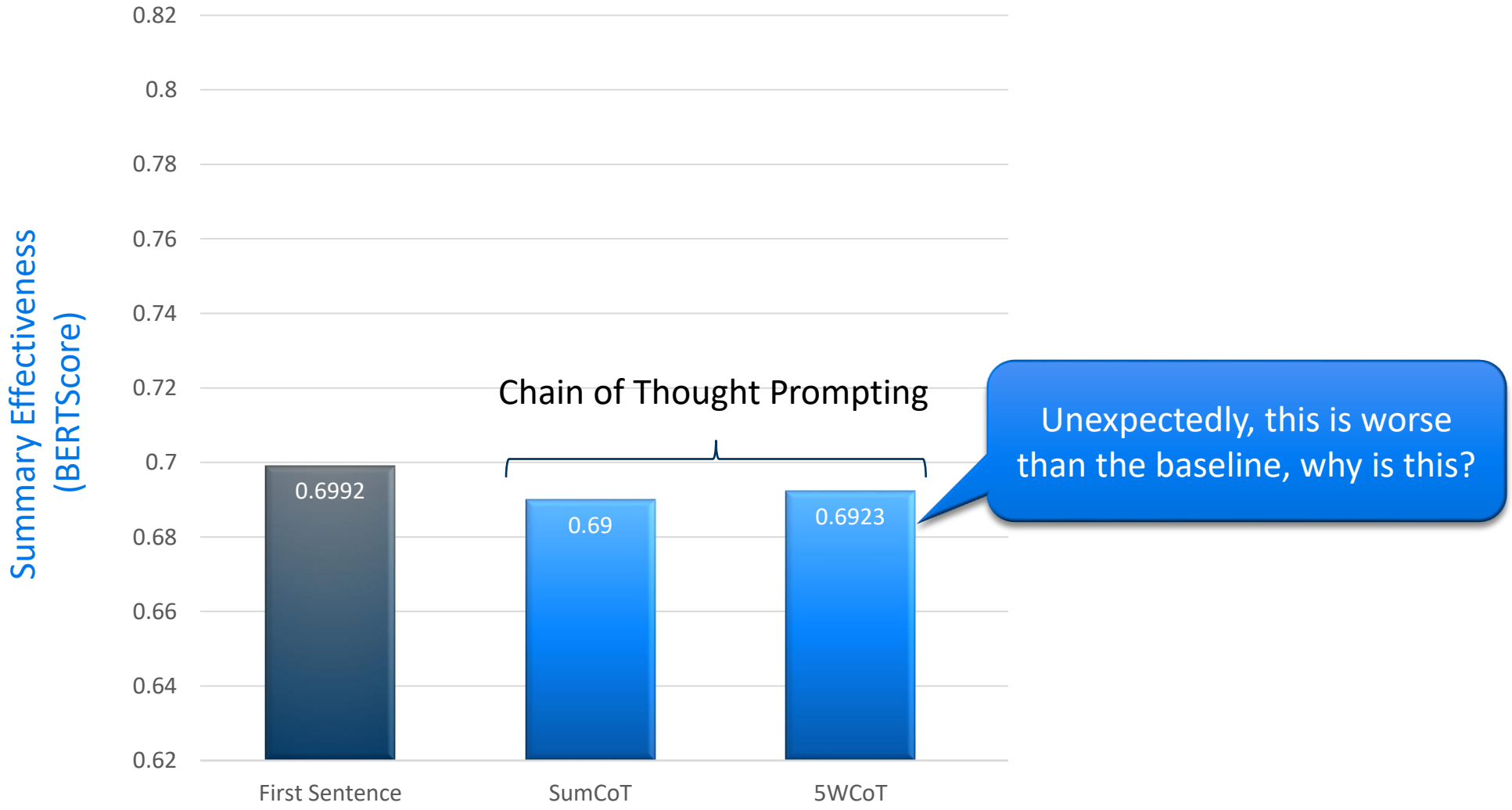
How Effective are our Summaries?

Let’s start with a classically strong baseline…



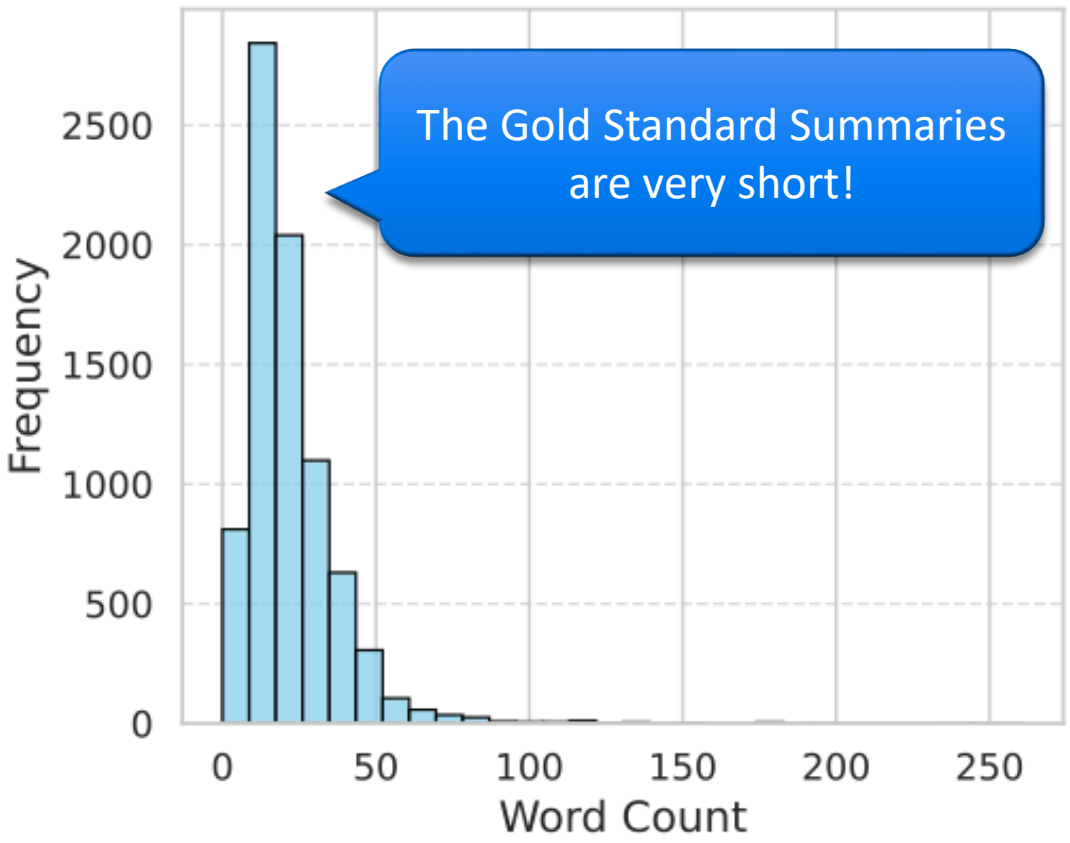
- First Sentence
  - The first few sentences often serves as a concise summary, highlighting the key aspects of the story for readers
  - Frequently used baseline in news summarization task
  - Strong baseline achieved BERTScore of ~0.7

Approach 1: Chain of Thought



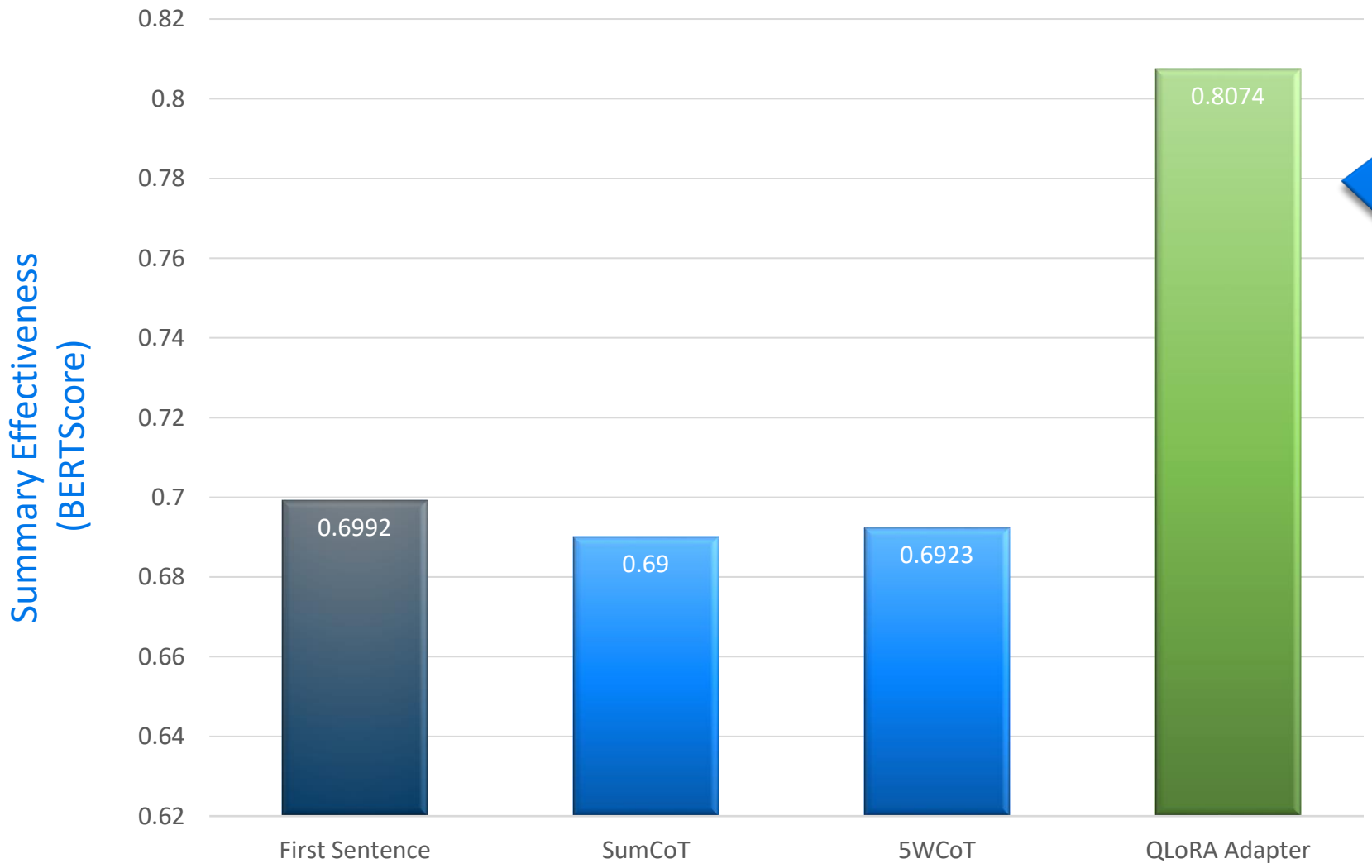
Is the task really Article Summarization?

It is too difficult to answer multiple questions in ~20 words



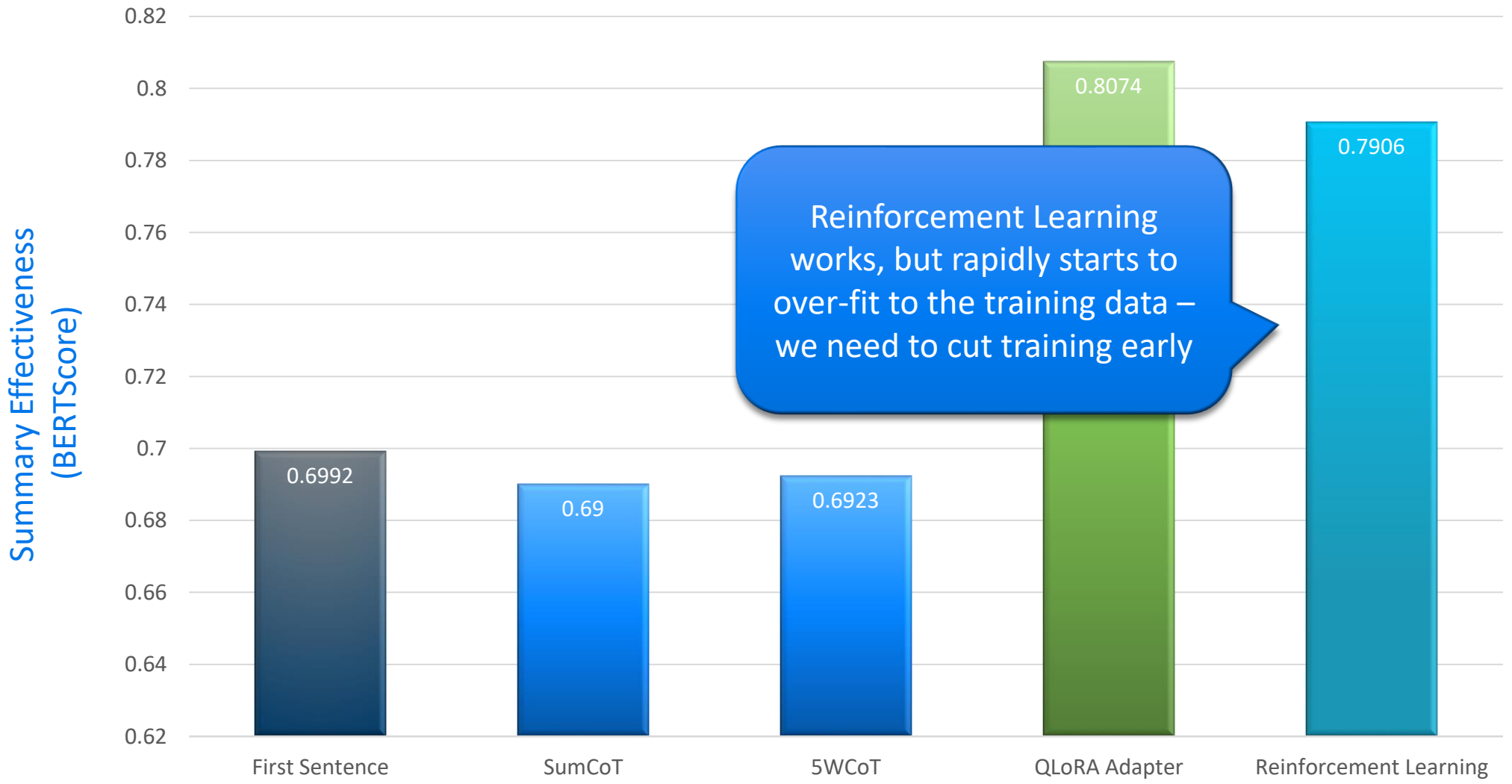
The task is closer to **headline generation** than article summarization

Approach 2: Adapter Training



The trained adapter works very well – it learns how to concisely summarize the topic of the article
(and not much else)

Approach 3: Reinforcement Learning



Main Conclusions

- Instruction tuning is essential to align the response of LLM with the desired summaries (for this task formulation)

- The reward function based on similarity to the gold standard lacks the granularity needed to provide effective signals when training large language models

But… how useful are the generated summaries?

- Let's consider the goal of news summary…
  - Provide the "target user" with the key facts and narrative of the news content
  - Provide coverage of the main entities and important facts
  - Should include enough contextual information for assistant "target users'" decision making

- The generated summaries do not fit well with these goals!
  - Aiming to generate short headlines (~17 words per summary)

Either human or AI system

- Suggestions for enhancing the task:
  - More explicitly model the information needs of the news consumer
  - Consider longer summaries
  - Incorporate human assessors into the evaluation stage

Questions?