

# Tutor-ICL: Guiding Large Language Models for Improved In-Context Learning Performance

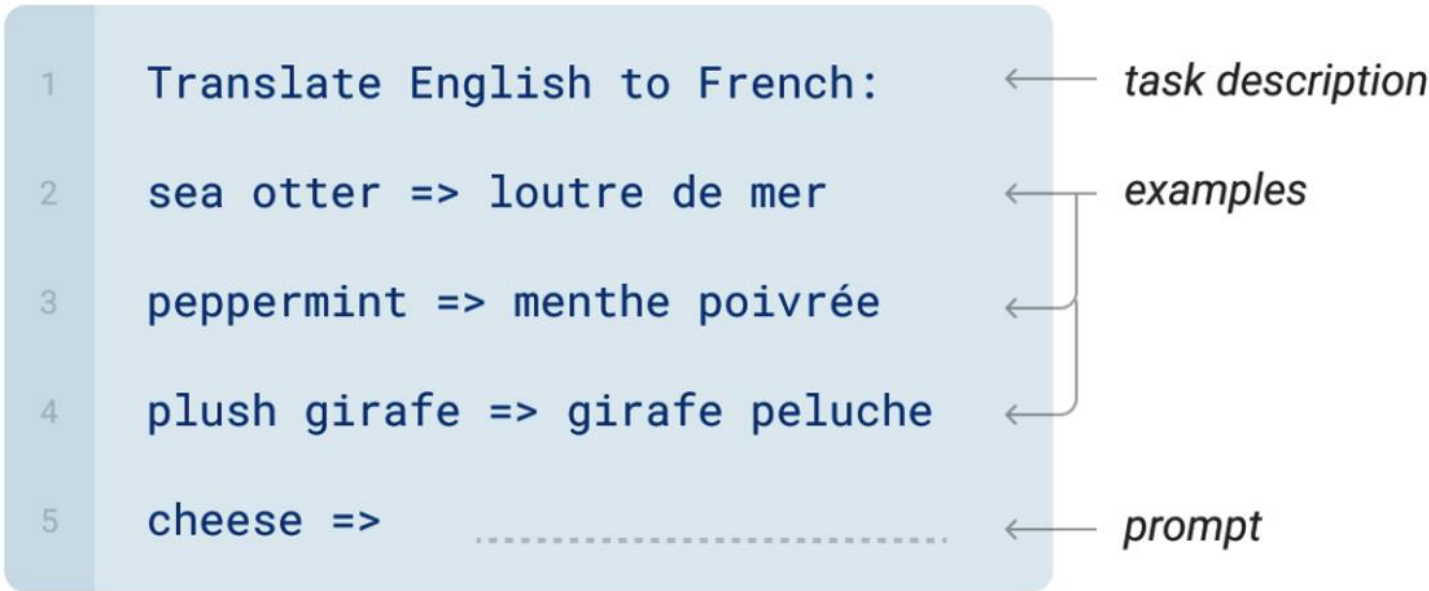
Ikhyun Cho<sup>1</sup>, Gaeul Kwon<sup>2</sup>, and Julia Hockenmaier<sup>1</sup>

<sup>1</sup>Siebel School of Computing and Data Science, University of Illinois at Urbana-Champaign, <sup>2</sup>Department of Computer Science, Grinnell College



## In-Context Learning (ICL) in LLMs

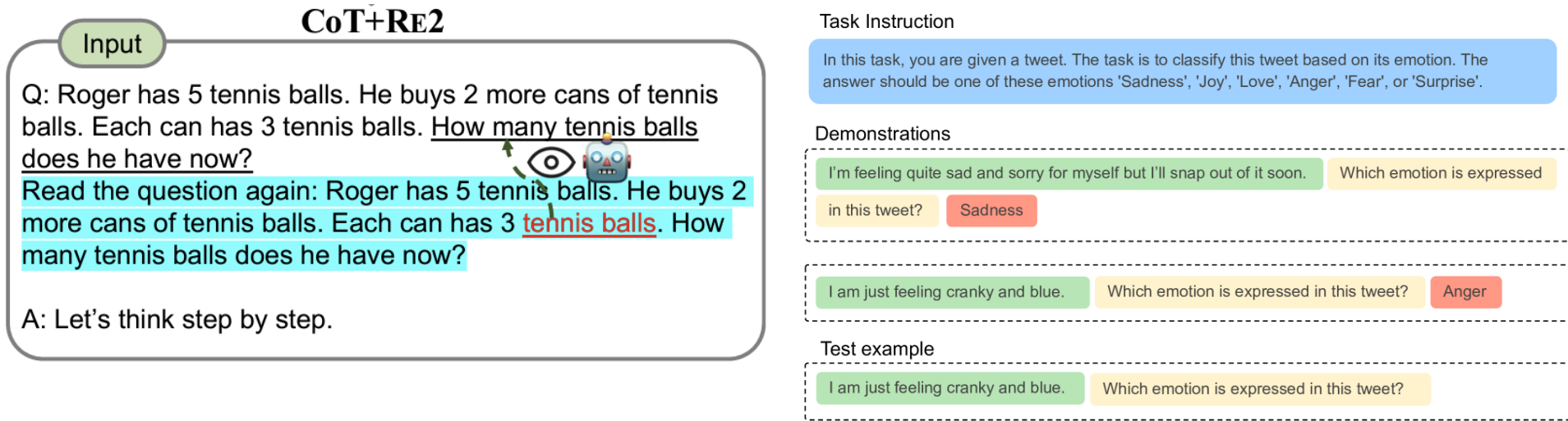
### 1. In-Context Learning with LLMs Works



LLMs can solve new tasks by learning from the examples on the fly, without further training.

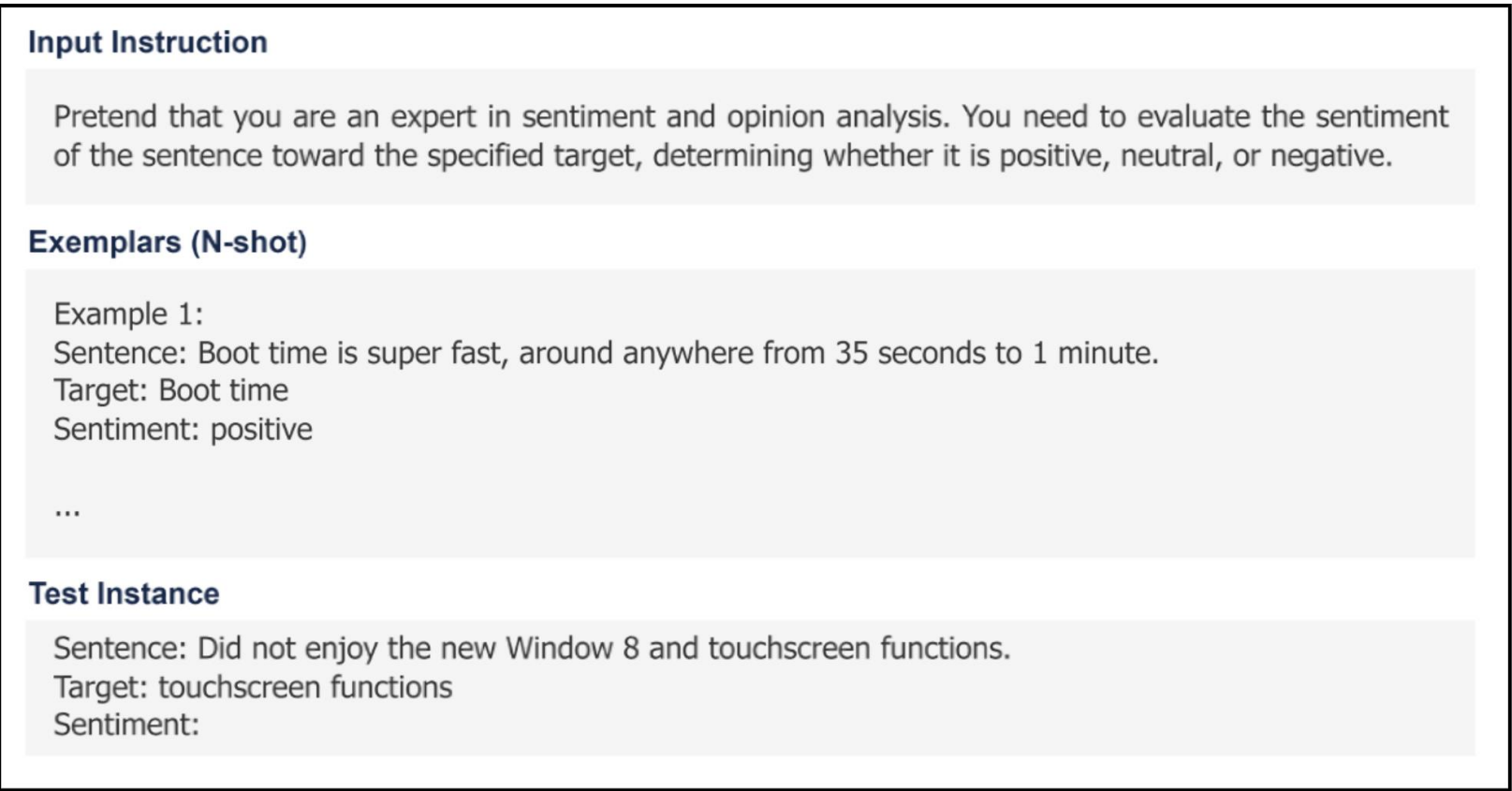
### 2. Template Design Matters in ICL

Several recent studies show that simple modifications to the template can boost ICL performance.



Repeating Questions (Avg. +2.2% in arithmetic)      Inserting Inline Instructions (Avg. +12.5% in classifications)

### 3. Our Approach



Standard Template

We improve ICL classification performance by adding three novel components to the standard template.

## Experimental Settings

### Tasks and Datasets

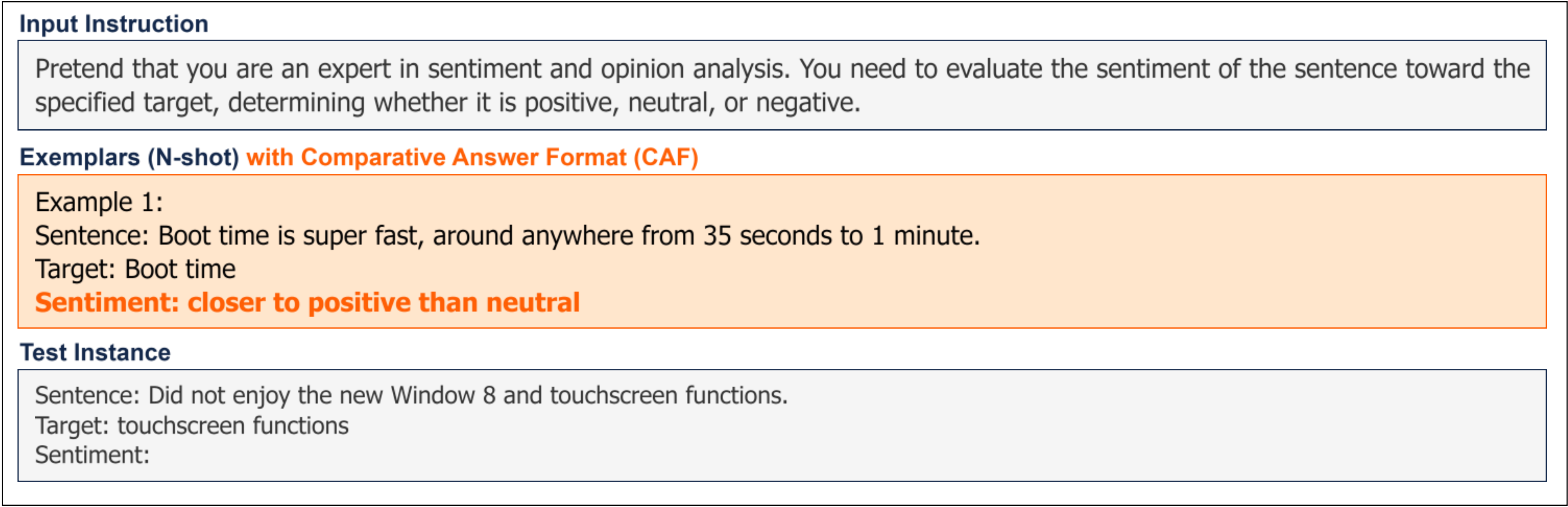
1. Aspect-Based Sentiment Classification (SemEval-14-Laptops/Restaurants)
2. News Topic Classification (AGNews)
3. Question Type Classification (TREC-QC)
4. Emotion Classification in Dialogues (EmoC)

### Models

1. Encoder-Decoder LLMs (Flan-T5-XL (3B), XXL (11B))
2. Decoder Only LLMs (Llama2-7B, 13B, Llama3-8B-Instruct)

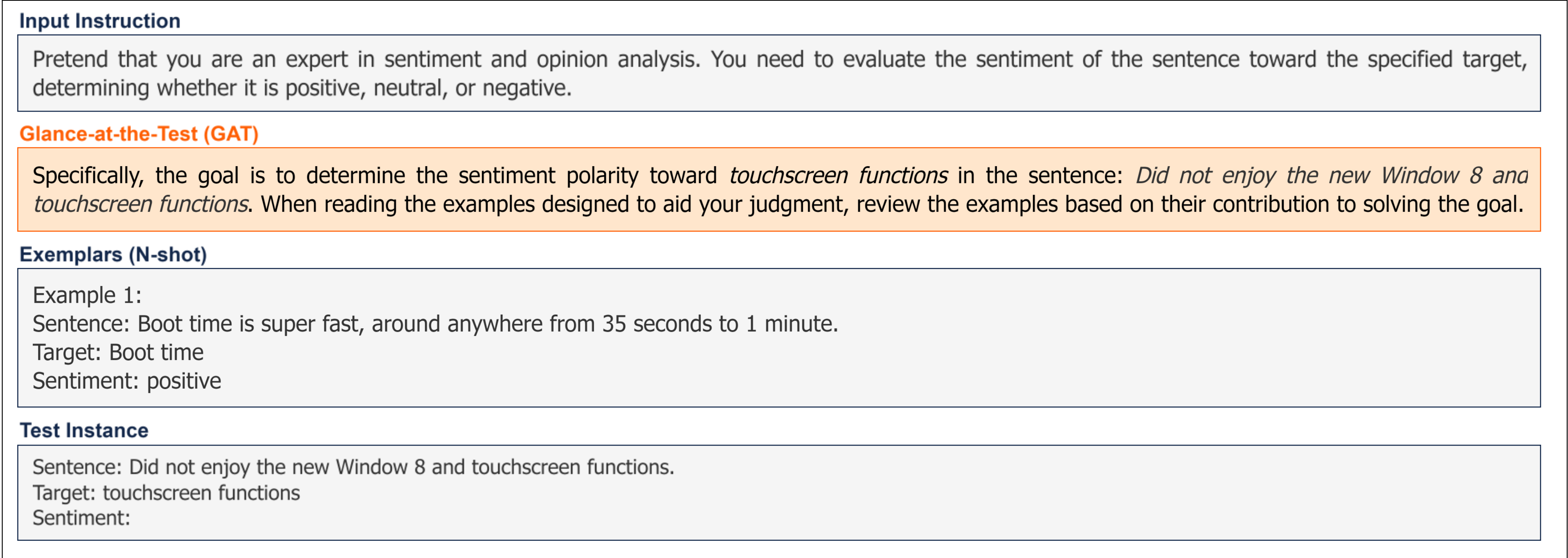
## Tutor-ICL

### 1. Comparative Answer Format (CAF)



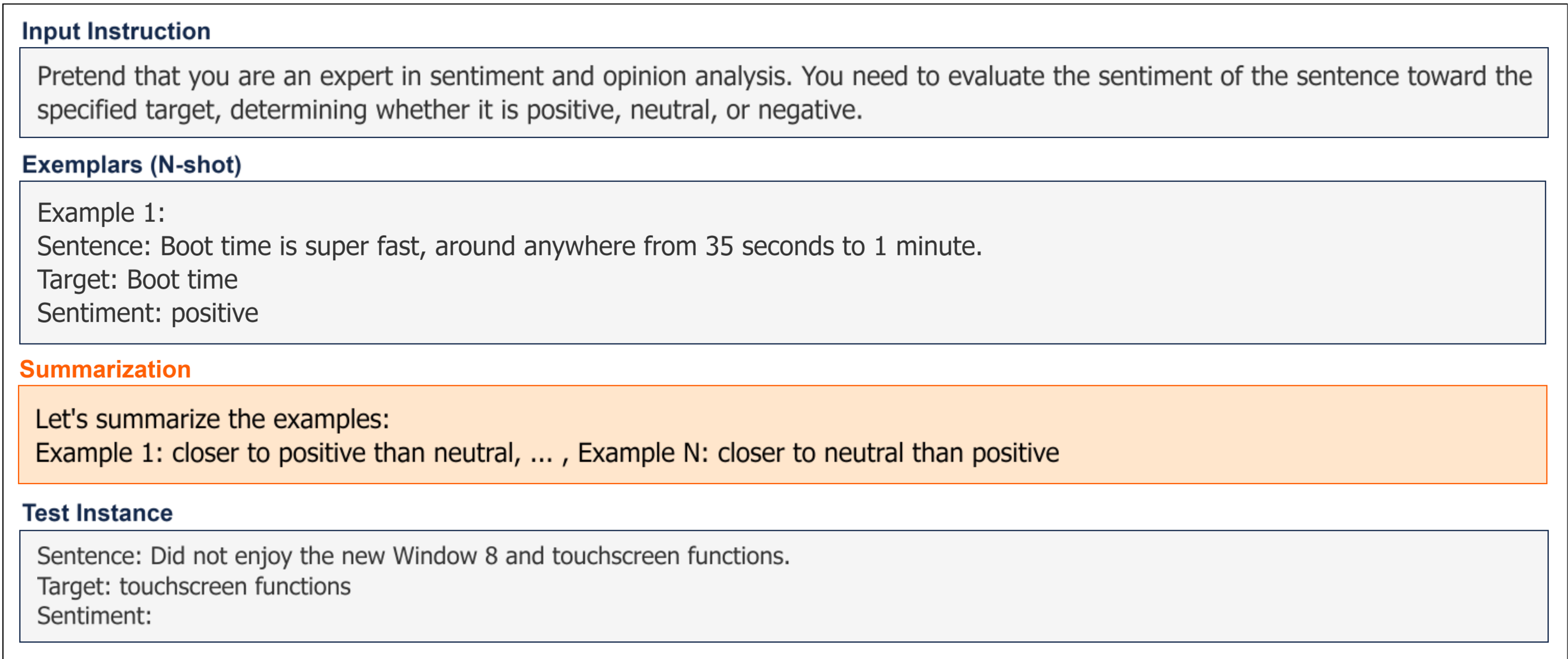
- Comparative answer format encourages deeper thinking from various answer perspectives.

### 2. Glance-At-the-Test (GAT)



- Providing the test instance in advance allows LLMs to focus more on the relevant exemplars using the self-attention mechanism.

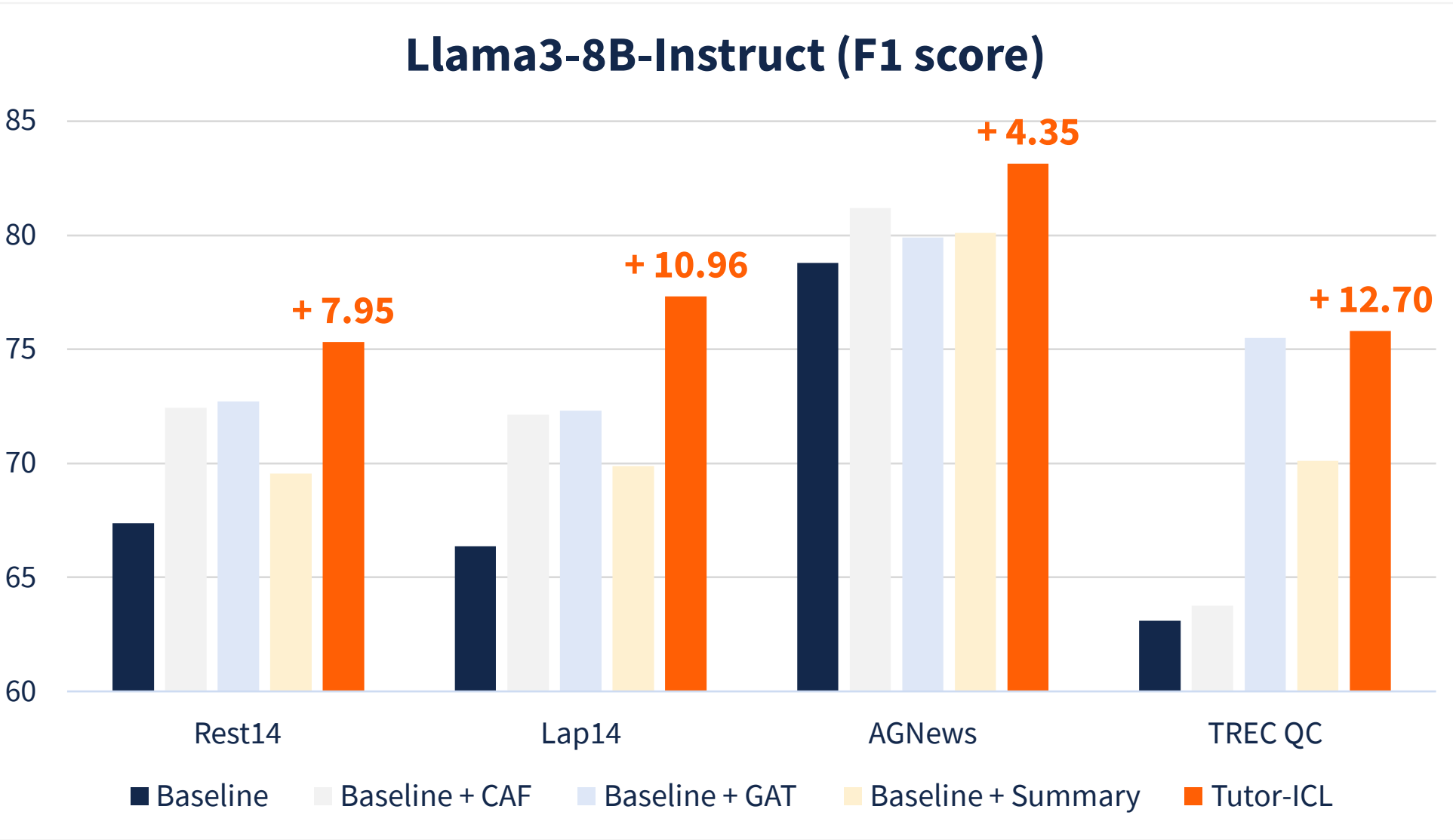
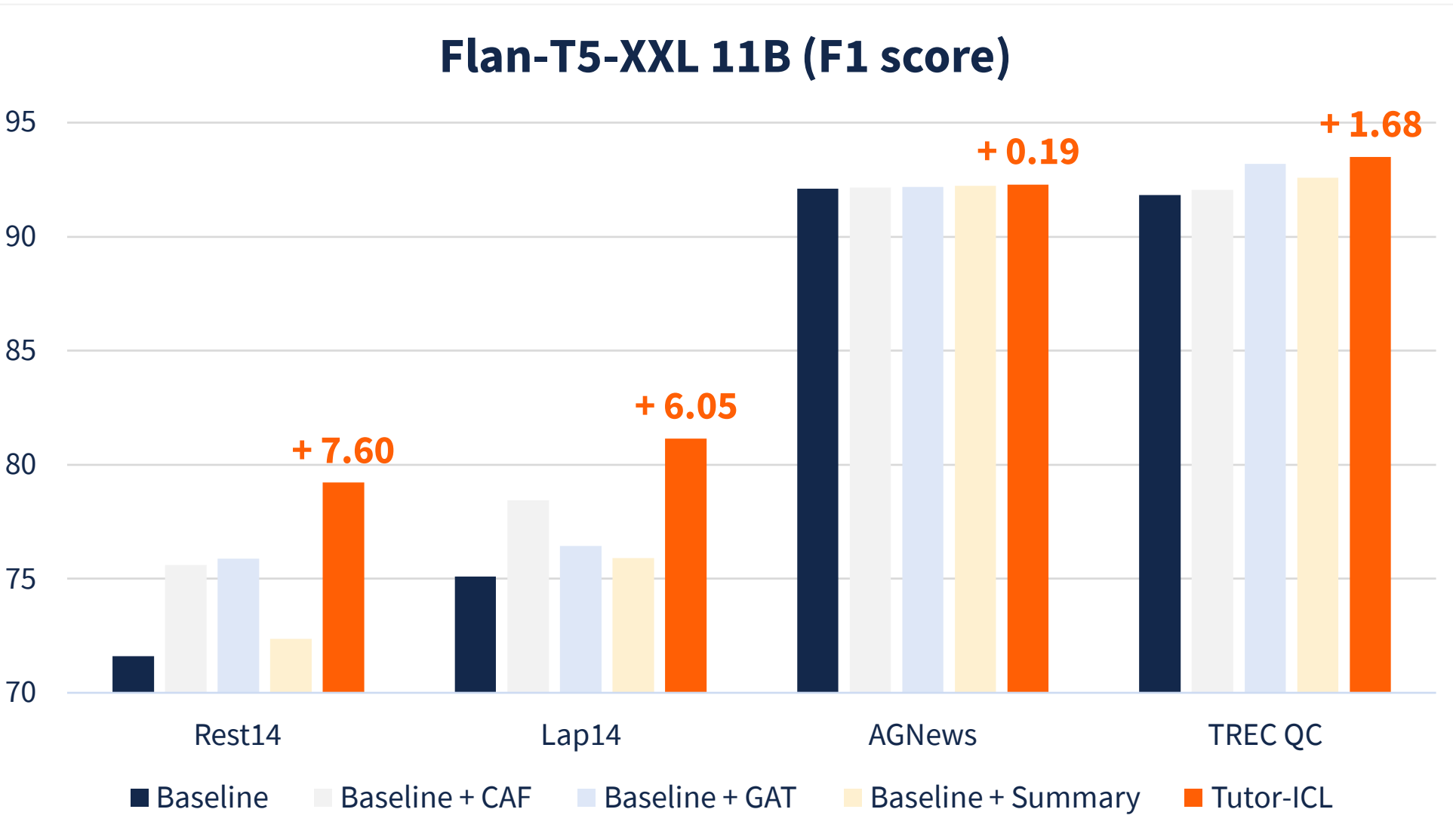
### 3. Summarization



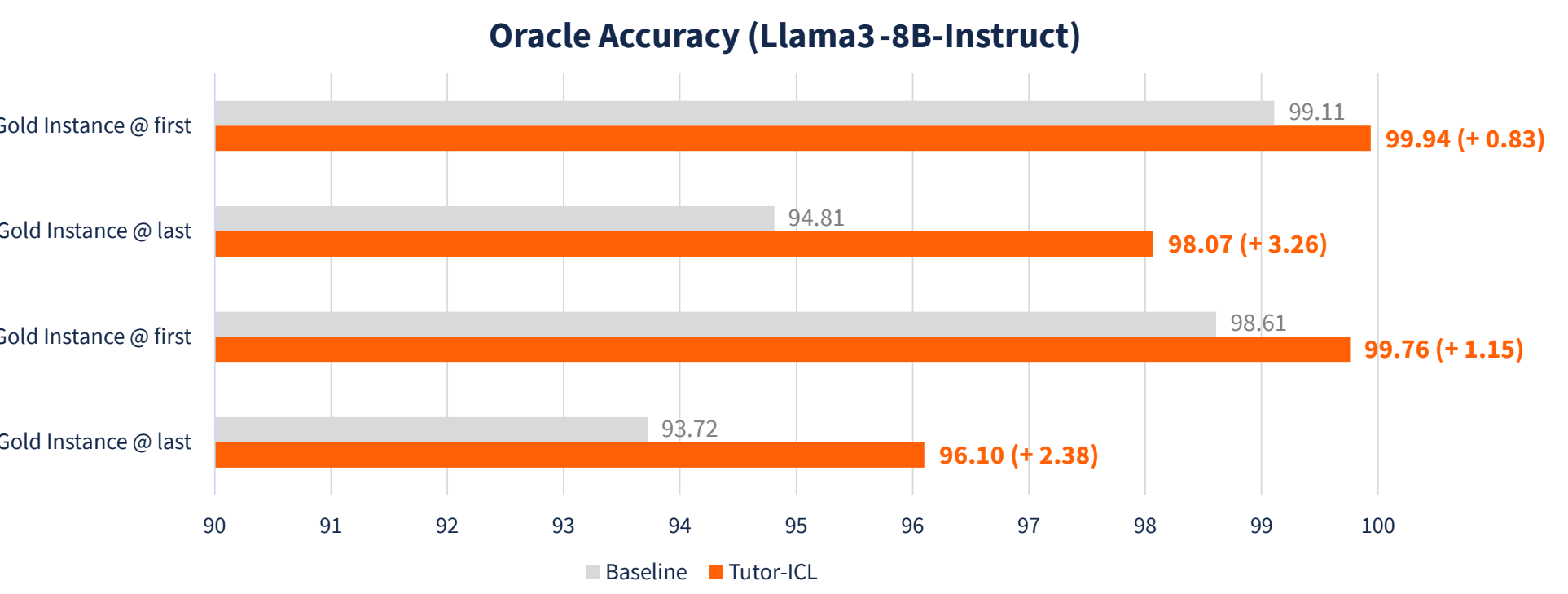
- Mimicking a common human practice of reviewing the materials to gain a deeper understanding.

## Results

### 1. Tutor-ICL achieves consistent improvements across various models and tasks:



### 2. With Tutor-ICL, LLMs make better use of exemplars:



We include the test instance & gold answer as one of the ICL exemplars. Tutor-ICL results in accuracy closer to 100%.

## Conclusion

We improve ICL classification performance by adding three new components to the standard ICL template:

1. Comparative Answer Format (CAF)
2. Glance-at-The-Test (GAT)
3. Summarization