Natural Language Processing

Adaptation

Kevin Lin – UC Berkeley

April 5, 2023

Adaptation

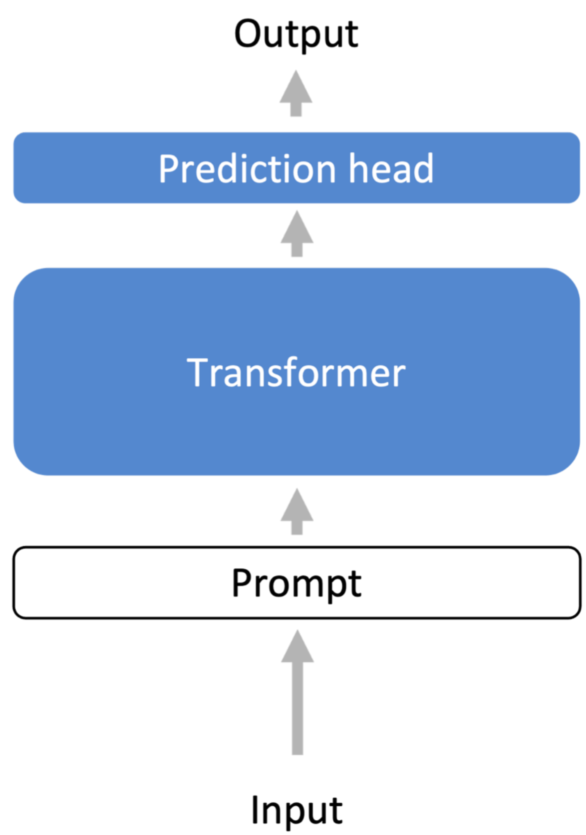
Challenges

- Task Format
  - **Premise**: I have never seen an apple that is not red.
  - **Hypothesis**: I have never seen an apple.
  - **Correct output**: Not entailment
- Harm Mitigation / Control
- Domain Shift
  - e.g. biomedical text, legal text
- Temporal Shift
  - Updated facts, news, etc.

Methods

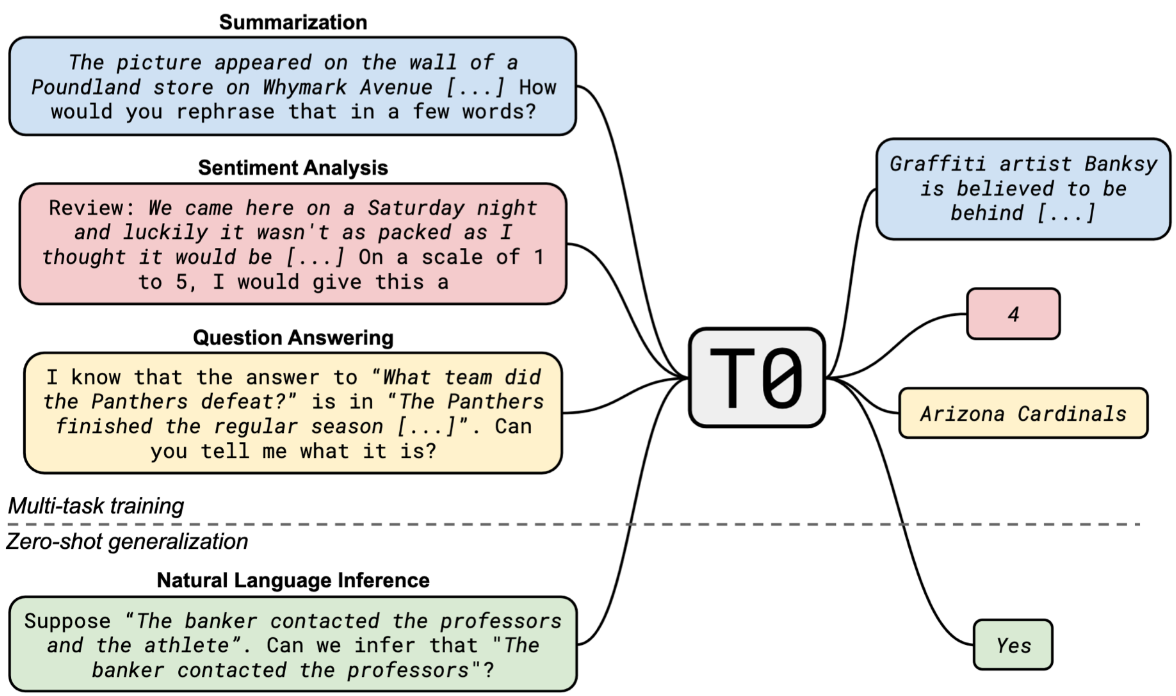
- Finetuning
- Lightweight Finetuning
- Prompting

Finetune



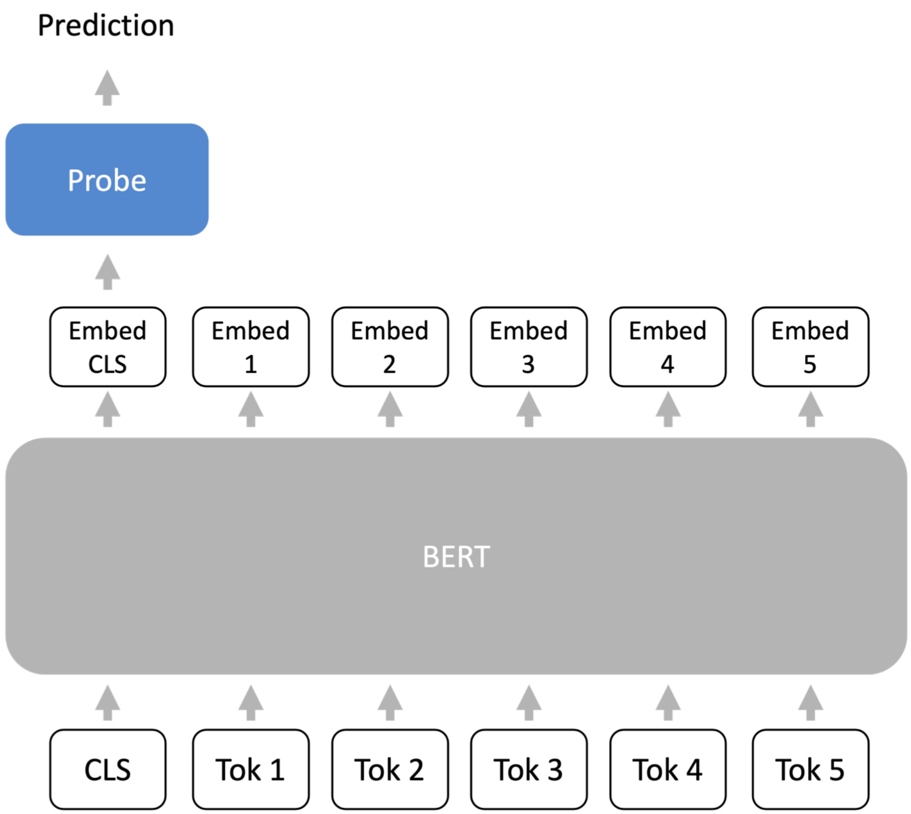
Finetuning for Zero-Shot / Instructions

- Multi-task Prompted Training Enabled Zero-Shot Task Generalization (Sanh et al., 2022)



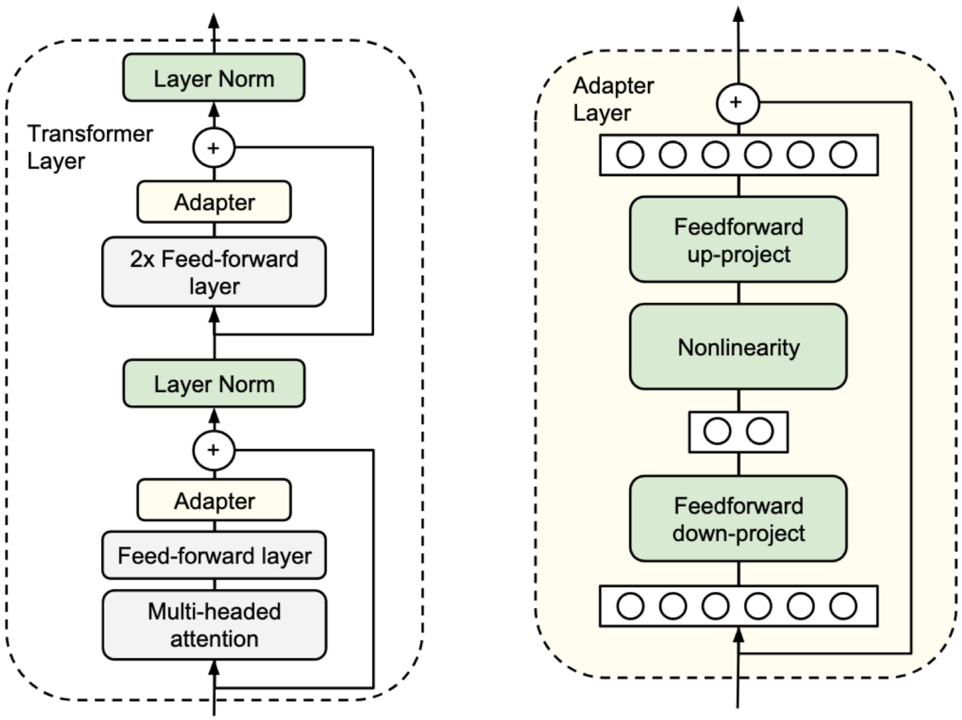
Lightweight Finetuning

- Probing



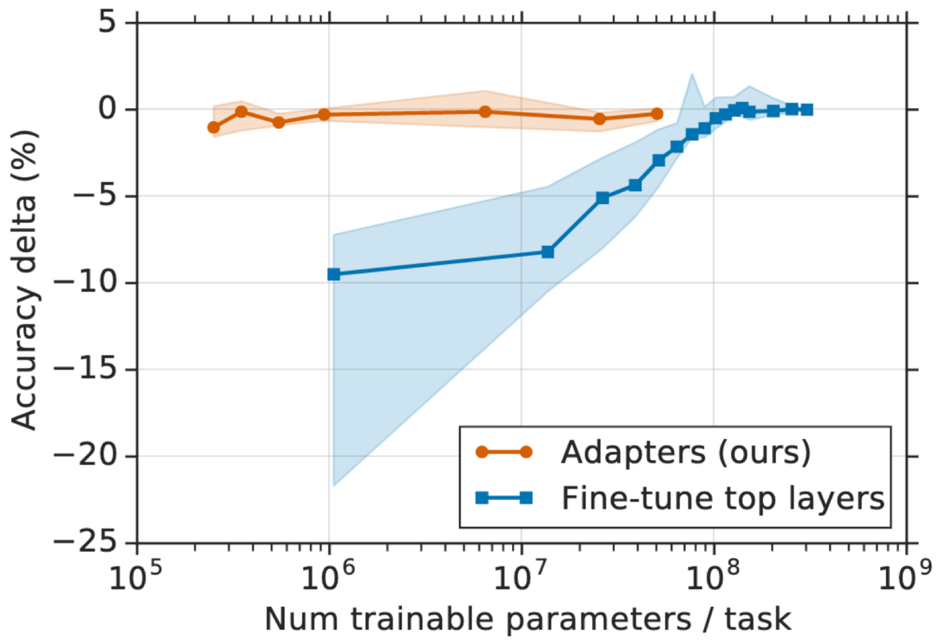
Lightweight Finetuning

- Adaptors: Parameter-Efficient Transfer Learning for NLP (Houlsby et al., 2019)



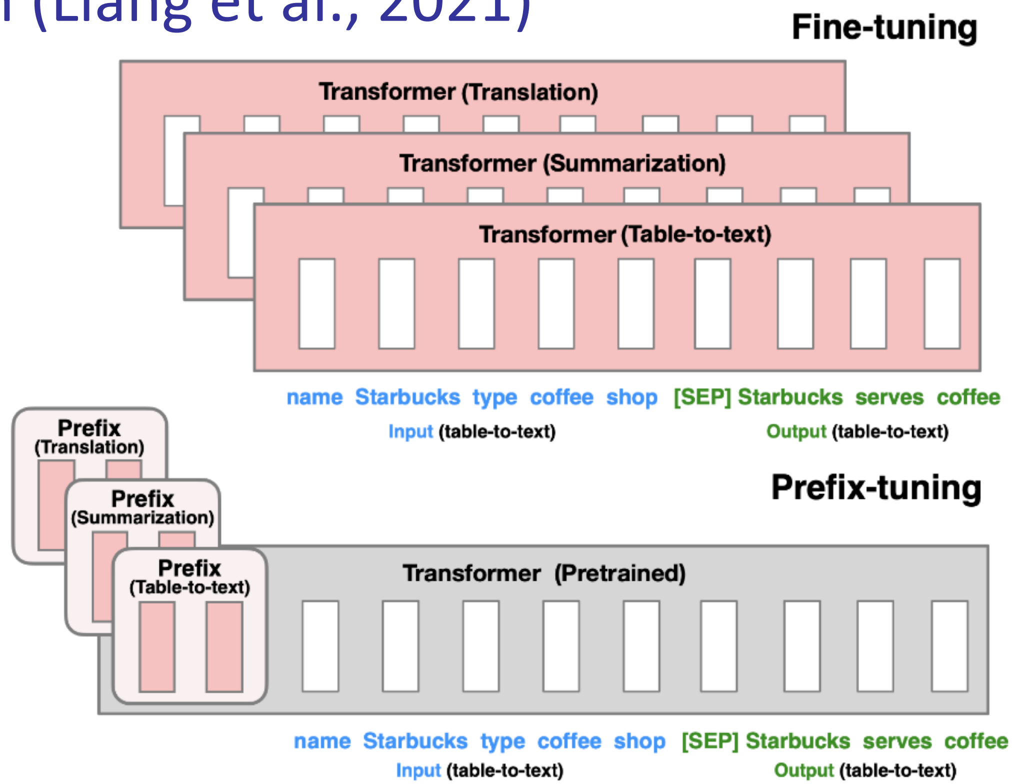
Lightweight Finetuning

- Adaptors: Parameter-Efficient Transfer Learning for NLP (Houlsby et al., 2019)



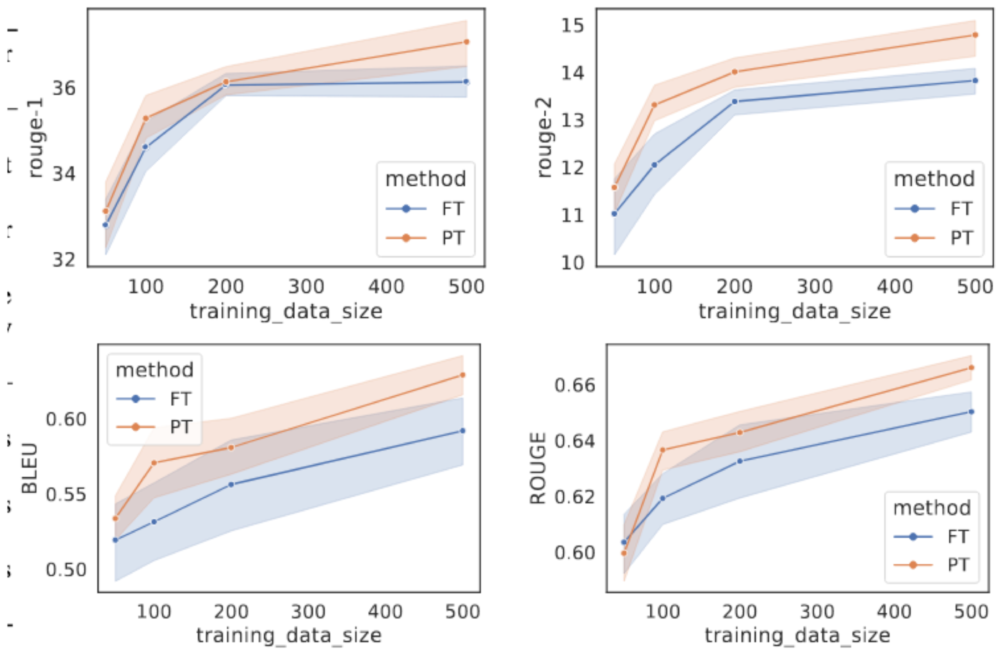
Prefix Tuning

- Prefix Tuning: Optimizing Continuous Prompts for Generation (Liang et al., 2021)



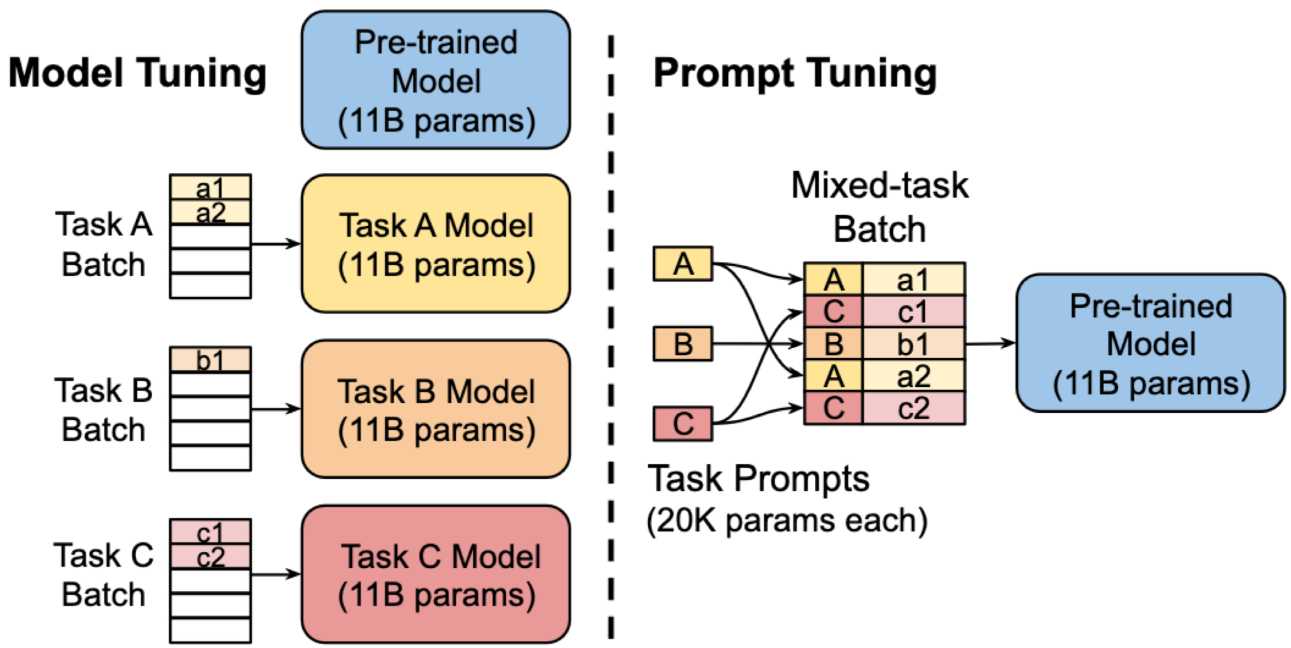
Prefix Tuning

- Prefix Tuning: Optimizing Continuous Prompts for Generation (Liang et al., 2021)



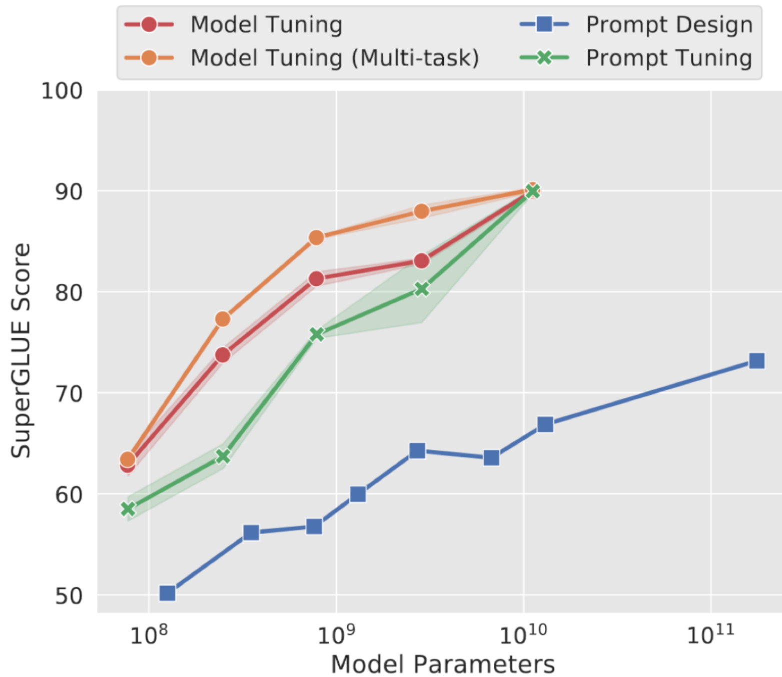
Lightweight Finetuning

- Prompt Tuning: The Power of Scale for Parameter-Efficient Prompt Tuning (Lester et al., 2021)



Lightweight Finetuning

- Prompt Tuning: The Power of Scale for Parameter-Efficient Prompt Tuning (Lester et al., 2021)



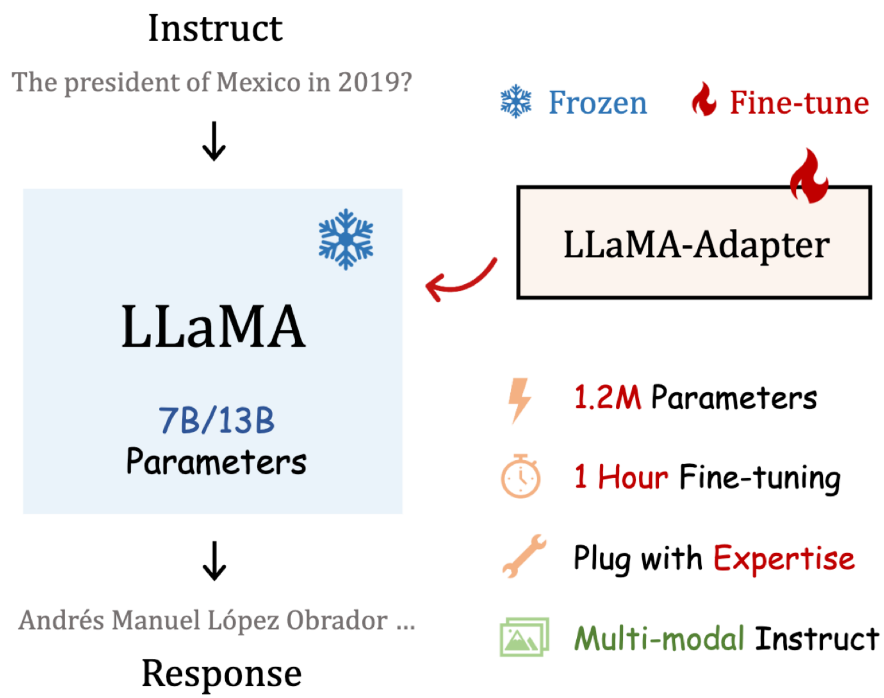
Lightweight Finetuning

- Lightweight finetuning improves OOD results

Dataset	Domain	Model	Prompt	Δ
SQuAD	Wiki	94.9 $\pm$0.2	94.8 $\pm$0.1	$-$0.1
TextbookQA	Book	54.3 $\pm$3.7	**66.8** $\pm$2.9	+12.5
BioASQ	Bio	77.9 $\pm$0.4	**79.1** $\pm$0.3	+1.2
RACE	Exam	59.8 $\pm$0.6	**60.7** $\pm$0.5	+0.9
RE	Wiki	88.4 $\pm$0.1	**88.8** $\pm$0.2	+0.4
DuoRC	Movie	**68.9** $\pm$0.7	67.7 $\pm$1.1	$-$1.2
DROP	Wiki	**68.9** $\pm$1.7	67.1 $\pm$1.9	$-$1.8

Lightweight Finetuning

- LLaMa-Adaptor: Efficient Finetuning of Language Models with Zero-Init Attention



Prompting

- Zero-Shot

```
Text: i'll bet the video game is a lot more fun than the film.
Sentiment:
```

copy

Prompting

- ## Few-Shot

```
Text: (lawrence bounces) all over the stage, dancing, running, sweating, mopping his face and g
Sentiment: positive

Text: despite all evidence to the contrary, this clunker has somehow managed to pose as an actu
Sentiment: negative

Text: for the first time in years, de niro digs deep emotionally, perhaps because he's been sti
Sentiment: positive

Text: i'll bet the video game is a lot more fun than the film.
Sentiment:
```

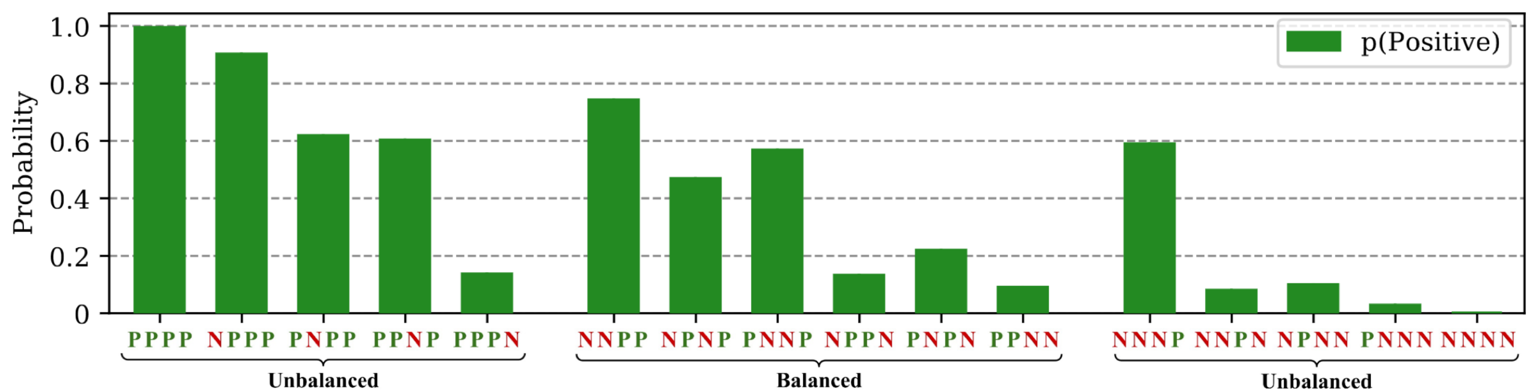
Prompting

- Calibrate Before Use: Improving Few-Shot Performance of Language Models (Zhao et al., 2021)

```
Text: (lawrence bounces) all over the stage, dancing, running, sweating, mopping his face and g
Sentiment: positive

Text: despite all evidence to the contrary, this clunker has somehow managed to pose as an actu
Sentiment: negative

Text: for the first time in years, de niro digs deep emotionally, perhaps because he's been st
Sentiment: positive

Text: i'll bet the video game is a lot more fun than the film.
Sentiment:
```

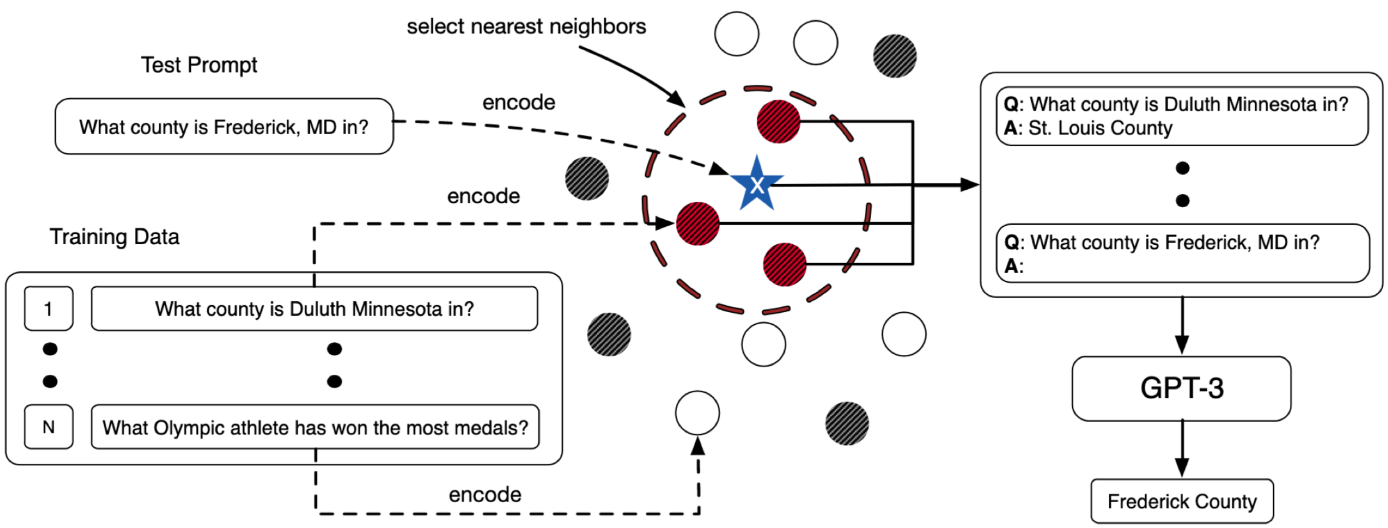
Prompting

- Calibrate Before Use: Improving Few-Shot Performance of Language Models (Zhao et al., 2021)
- Majority label bias
- Recency
- Common token



Example Selection

- K-nearest neighbor clustering
- What Makes Good In-Context Examples for GPT-3? (Liu et al., 2021)



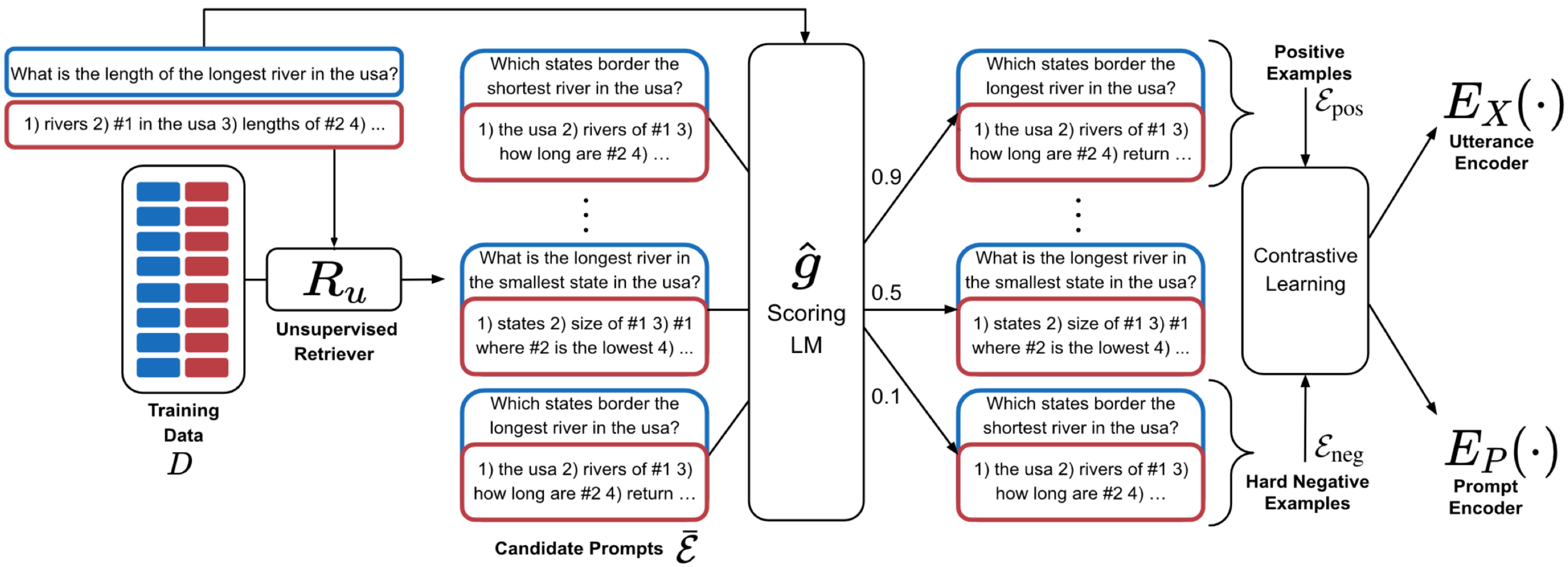
Example Selection

- Diversity-based selection
- Selective Annotation Makes Language Modeling Better Few-Shot Learners (Su et al., 2022)

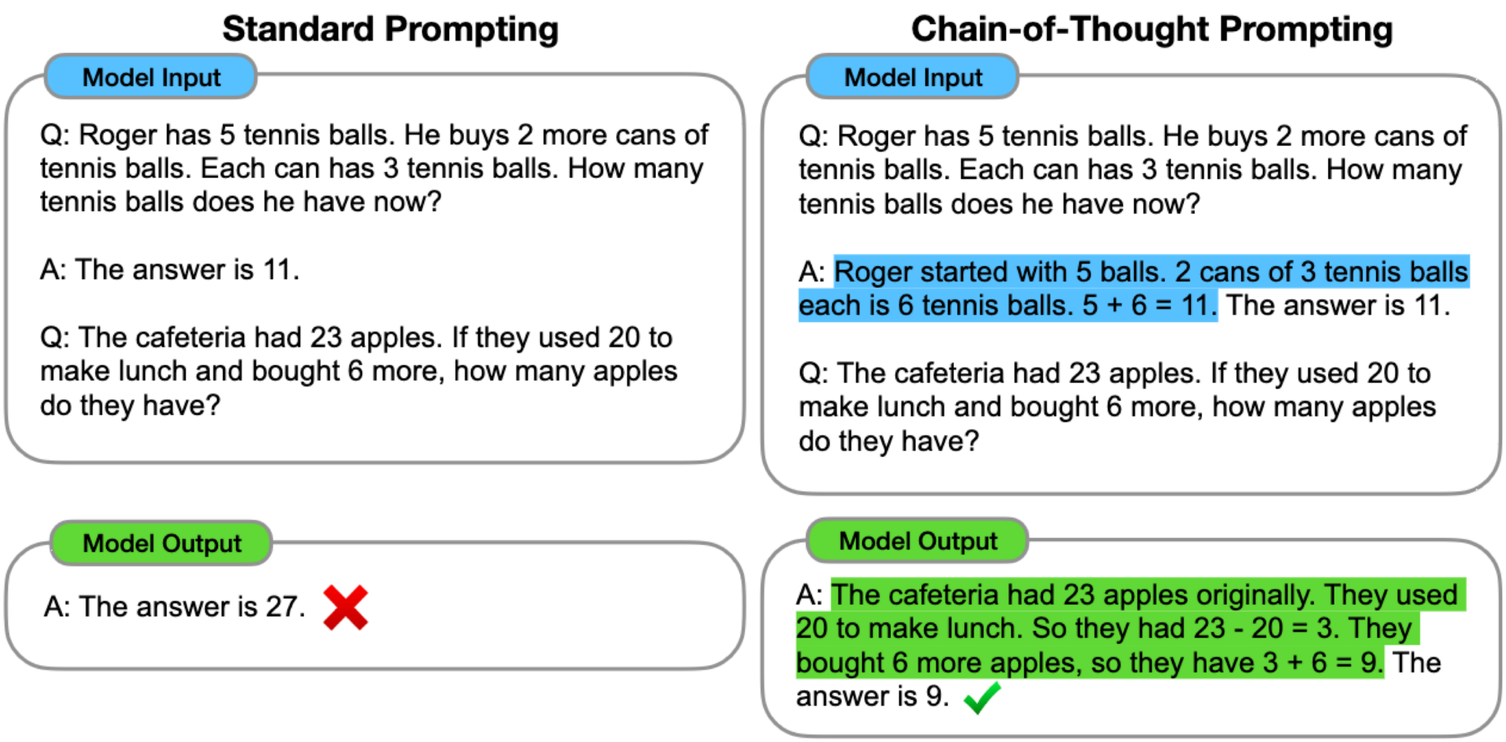
$$\mathrm{score}(u) = \sum_{v \in \{v \mid (v,u) \in E, v \in \mathcal{U}\}} s(v), \quad \text{where } s(v) = \rho^{-|\{\ell \in \mathcal{L} \mid (v,\ell) \in E\}|}, \quad \rho > 1$$

Example Selection

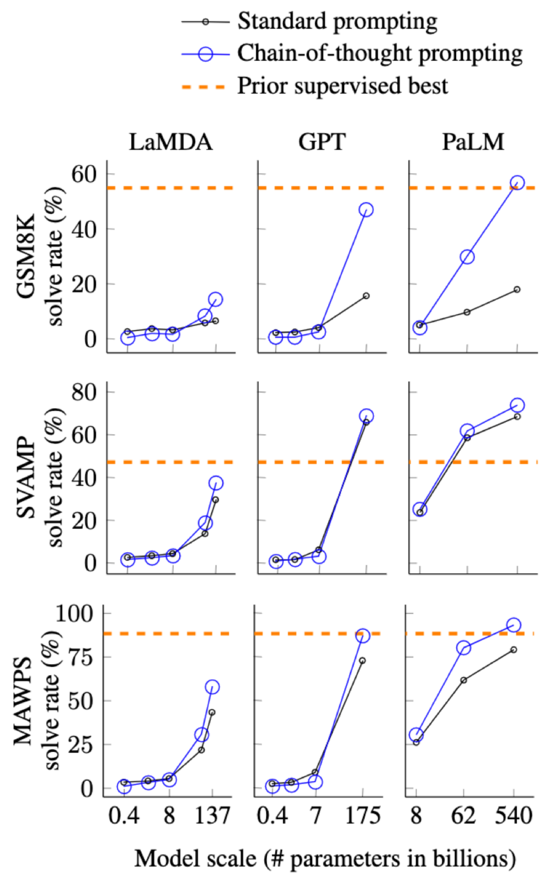
- Learning To Retrieve Prompts for In-Context Learning (Rubin et al., 2022)



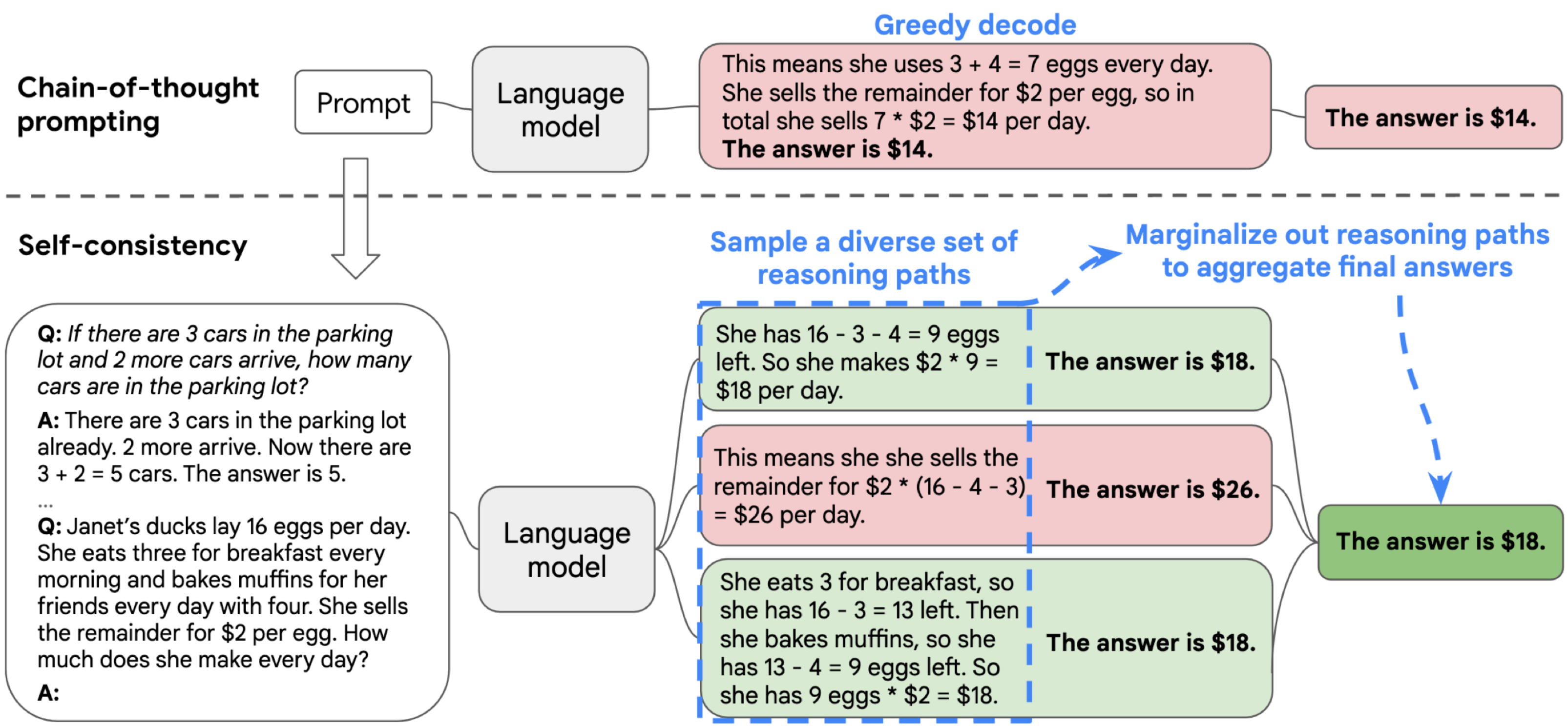
Chain-of-Thoughts



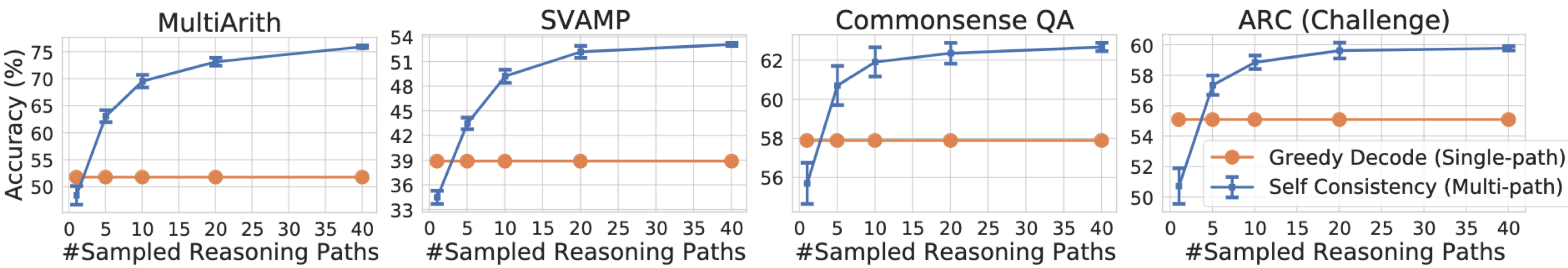
Chain-of-Thoughts



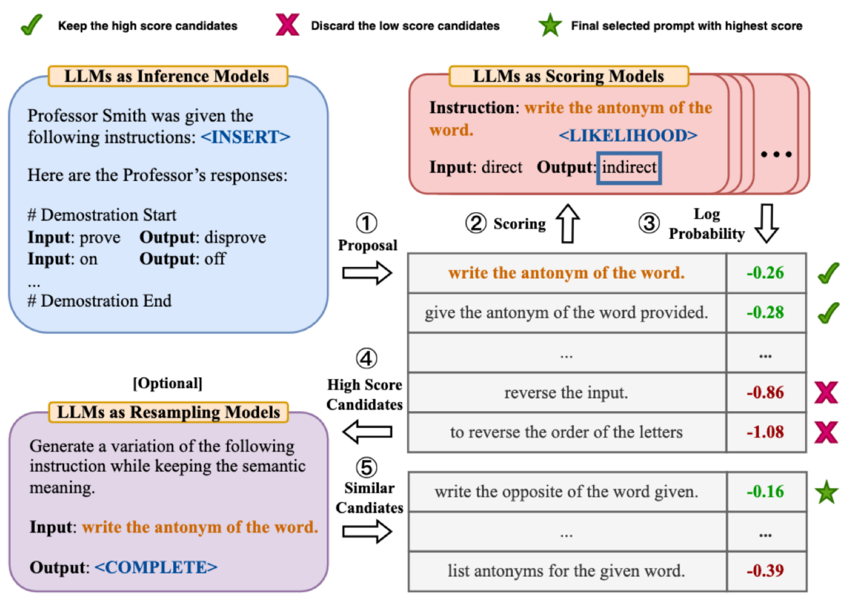
Self-Consistency



Self-Consistency



Automatic Prompting



(a) Automatic Prompt Engineer (APE) workflow

- AutoPrompt; Shin et al., 2020
- Automatic Prompt Engineer; Zhou et al., 2022

Automatic Chain-of-thoughts