PROMPTING

Veerav Chebrolu

M.Sc. Scientific Computing (8LP)

5 Dec 2023

Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing

Pengfei Liu
Carnegie Mellon University
pliu3@cs.cmu.edu

Weizhe Yuan
Carnegie Mellon University
weizhey@cs.cmu.edu

Jinlan Fu
National University of Singapore
jinlanjonna@gmail.com

Zhengbao Jiang
Carnegie Mellon University
zhengbaj@cs.cmu.edu

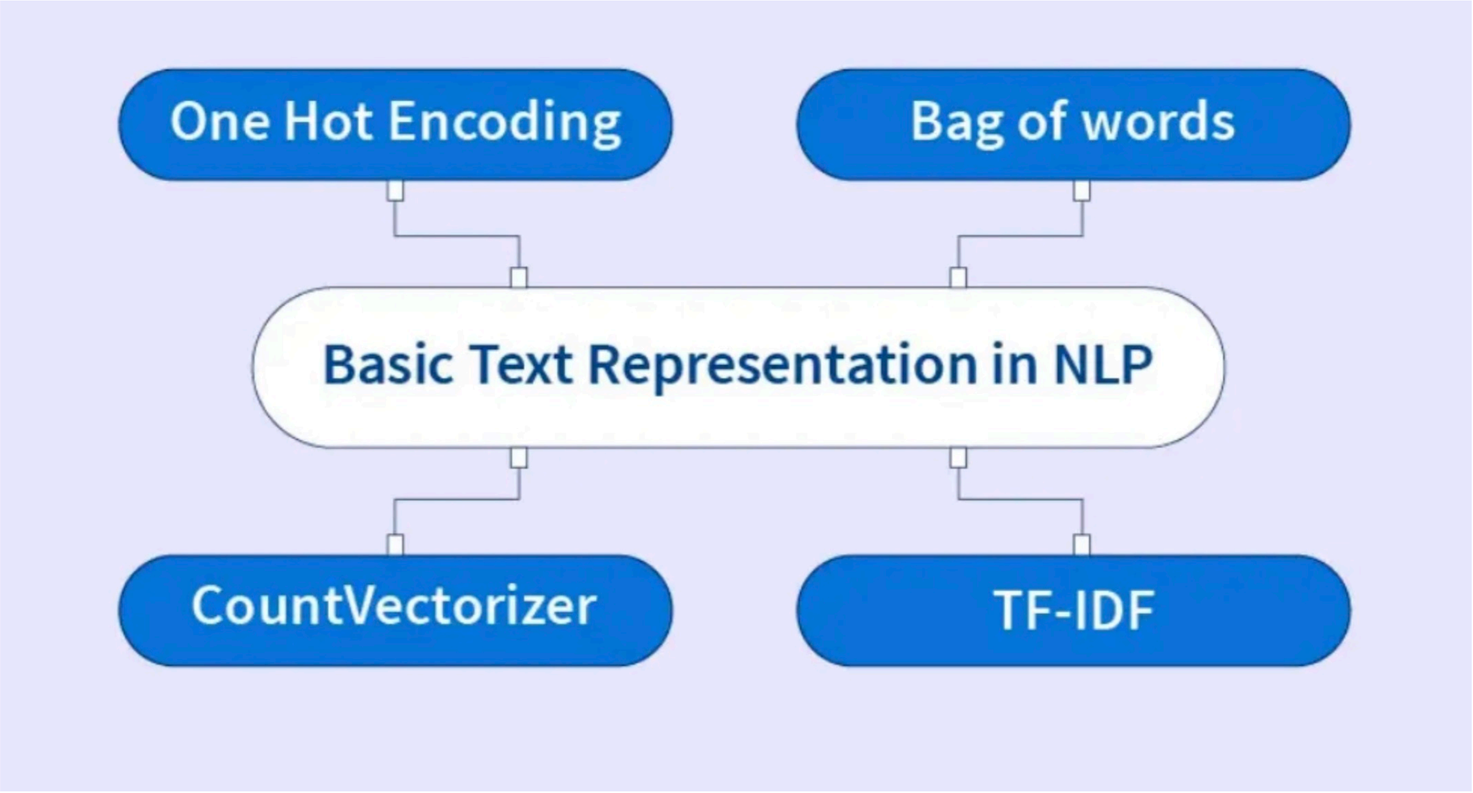
Hiroaki Hayashi
Carnegie Mellon University
hiroakih@cs.cmu.edu

Graham Neubig
Carnegie Mellon University
gneubig@cs.cmu.edu

2150 Citations

Paradigms of NLP Progress

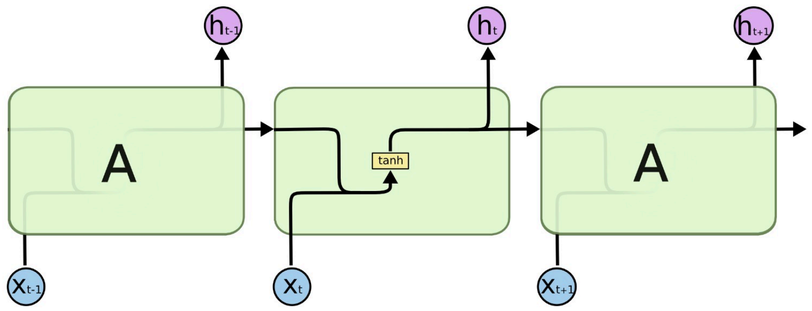
Feature Engineering (upto 2015)



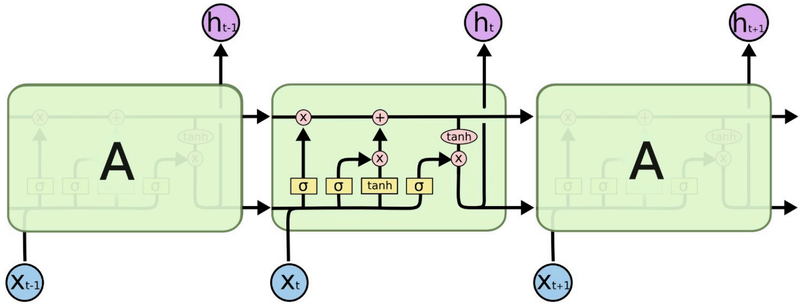
Manual Features → Support Vector Machines / Conditional Random Fields

Paradigms of NLP Progress

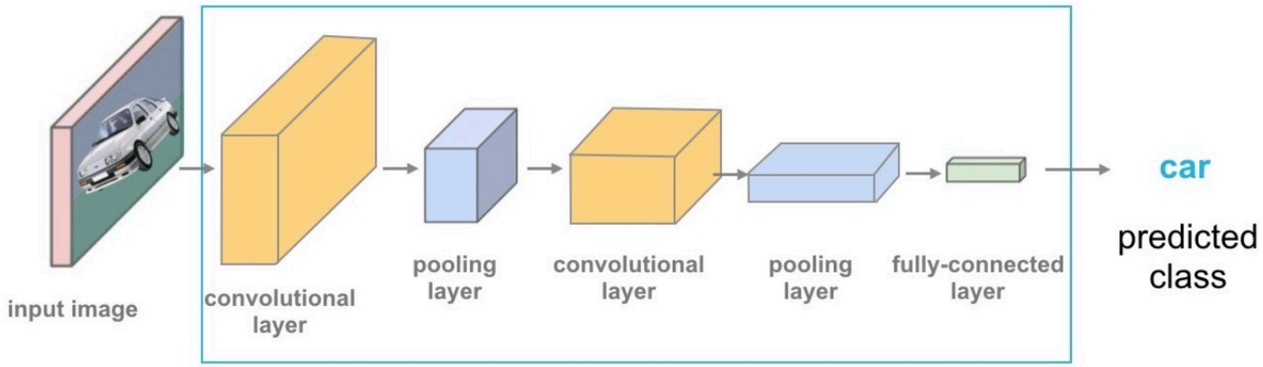
Architecture Engineering (2013-2018)



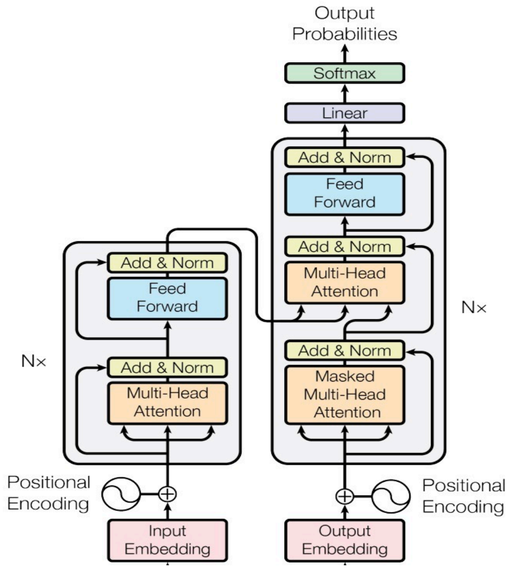
Recurrent Neural Networks



Long Short Term Memory



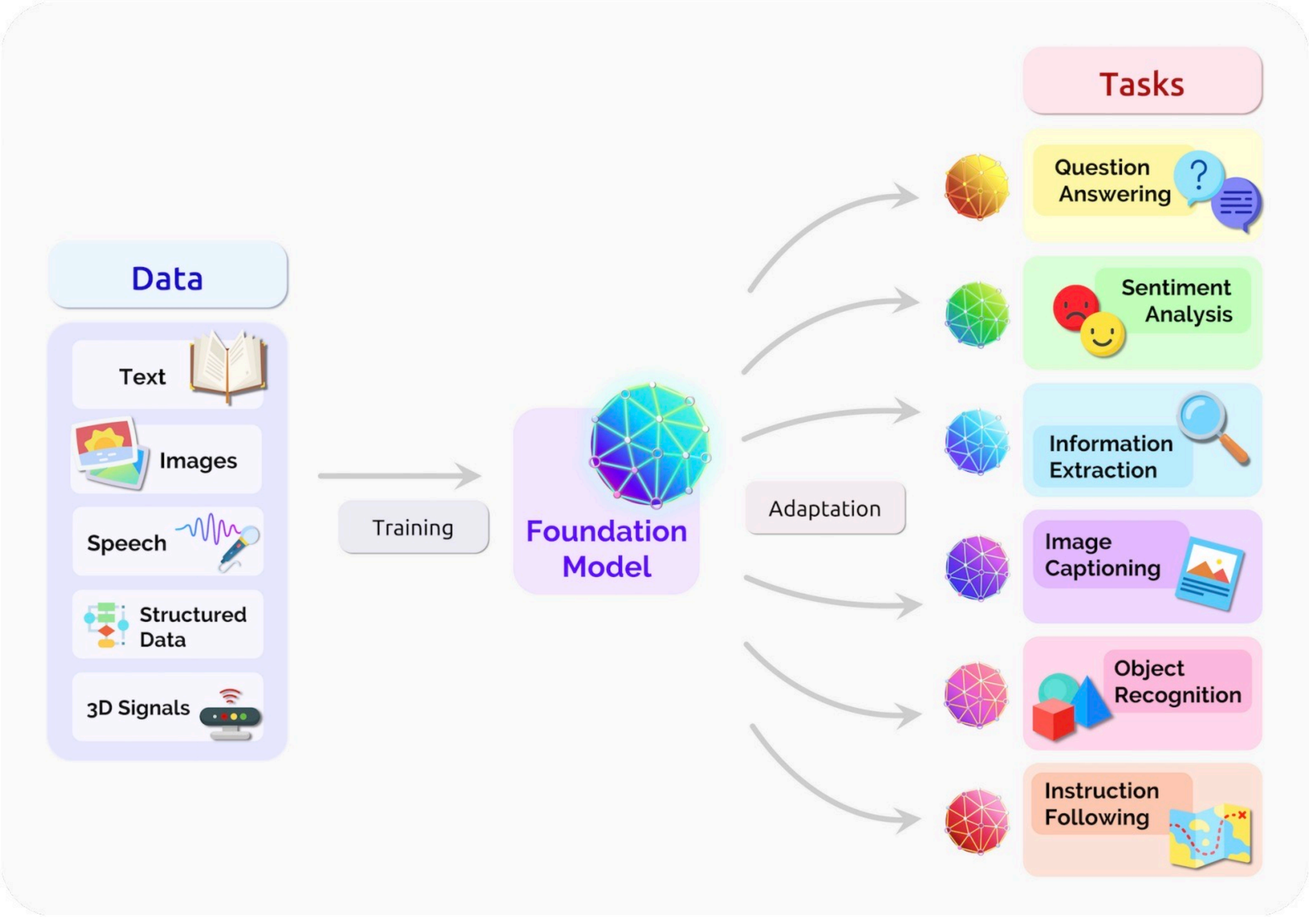
Convolutional Neural Networks



Transformers

Paradigms of NLP Progress

Objective Engineering (2017-Now)

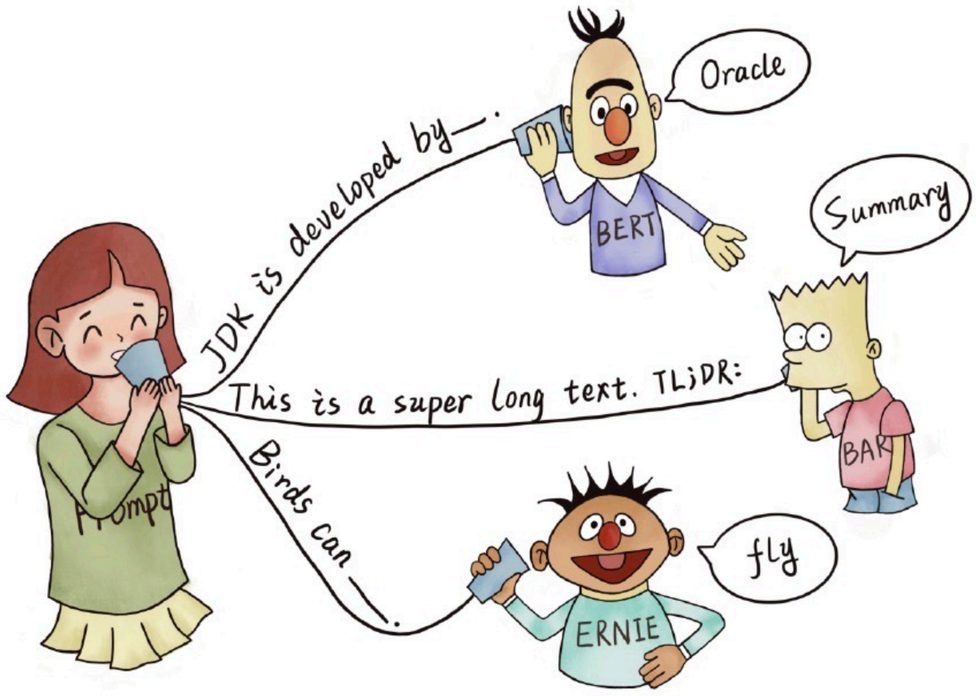


Paradigm : Pre-Training & Fine Tuning

Paradigms of NLP Progress

Prompt Engineering (2019-Now)

Encourage a Pre-trained model to make particular predictions by providing a "prompt" specifying the task to be done



Paradigm : Pre-Train, Prompt, Predict

What is the general workflow of prompting?

Prompt Addition

$$\mathbf{x}' = f_{prompt}(\mathbf{x})$$

Sentiment Classification

Input: x = "I love this movie."

Template: [x] Overall, it was a [z] movie.

Prompting: x' = "I love this movie. Overall, it was a [z] movie."

Cloze Prompt

Machine Translation

Input: x = "I love this movie."

Template: English: [x] Hindi: [z]

Prompting: x' = "English: I love this movie. Hindi: [z]"

Prefix Prompt

Answer Prediction

$$\hat{z} = \underset{z \in Z}{search} \quad P(f_{fill}(x', z); \theta)$$

Input: x = "I love this movie."

Template: [x] Overall, it was a [z] movie.

Prompting: x' = "I love this movie. Overall, it was a [z] movie."

Predicting: x' = "I love this movie. Overall, it was a fantastic movie."

Filled Prompt

boring

Answered Prompt

Answer Mapping

Predicting: x’ = “I love this movie. Overall, it was a fantastic movie.”

Mapping: fantastic => Positive

Design Considerations for Prompting

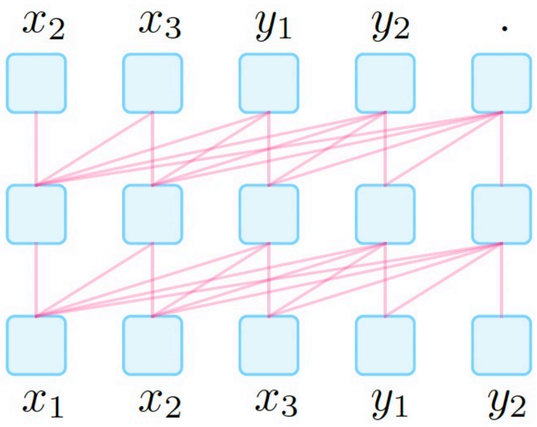
Pre- Trained model Choice

Prompt Engineering

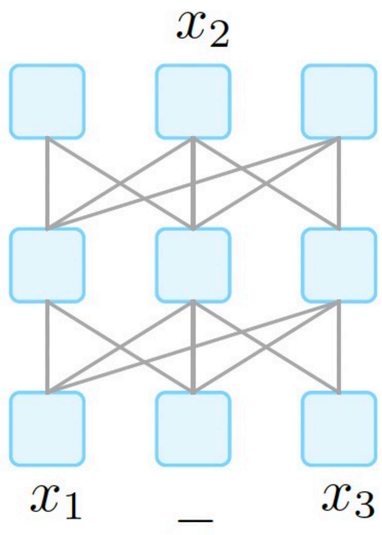
Answer Engineering

Prompt based training strategies

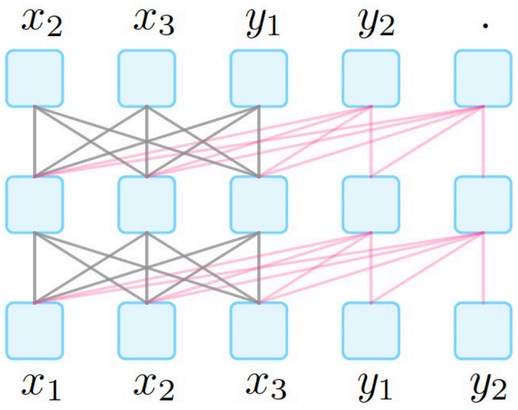
Pre Trained Language Models



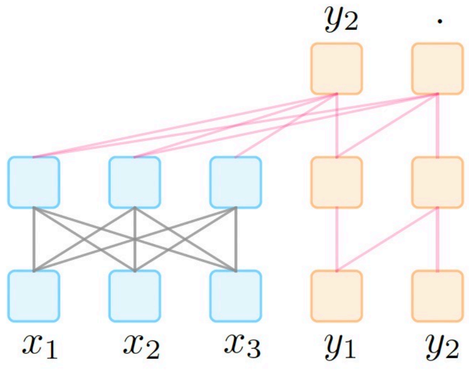
Left-Right LM



Masked LM

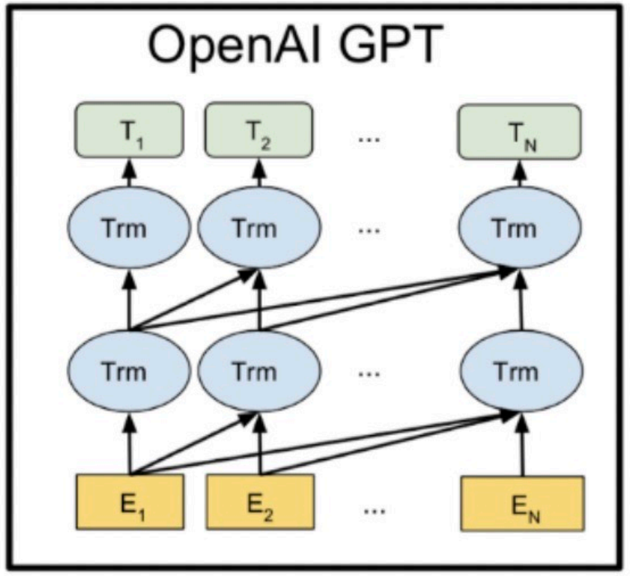


Prefix LM



Encoder - Decoder

Left-Right LM



GPT 3
(Brown et al. 2020)

$$P(\mathbf{x}) = P(x_1) \times \ldots P(x_n | x_1 \ldots x_{n-1})$$

Auto-regressive

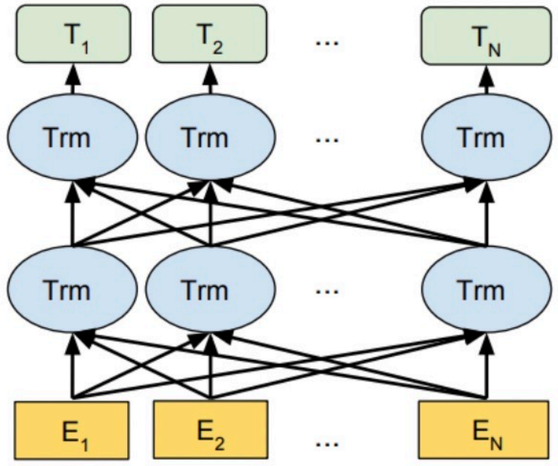
Objective : Standard Language Modelling

Data : CommonCrawl (1T words)

Roles in Prompting

- Earliest architecture chosen for prompting
- Usually used with Prefix prompts
- Parameters of PLMs are fixed.
- Suitable for Natural Language generation.

Masked LM



BERT
(Devlin et al. 2018)

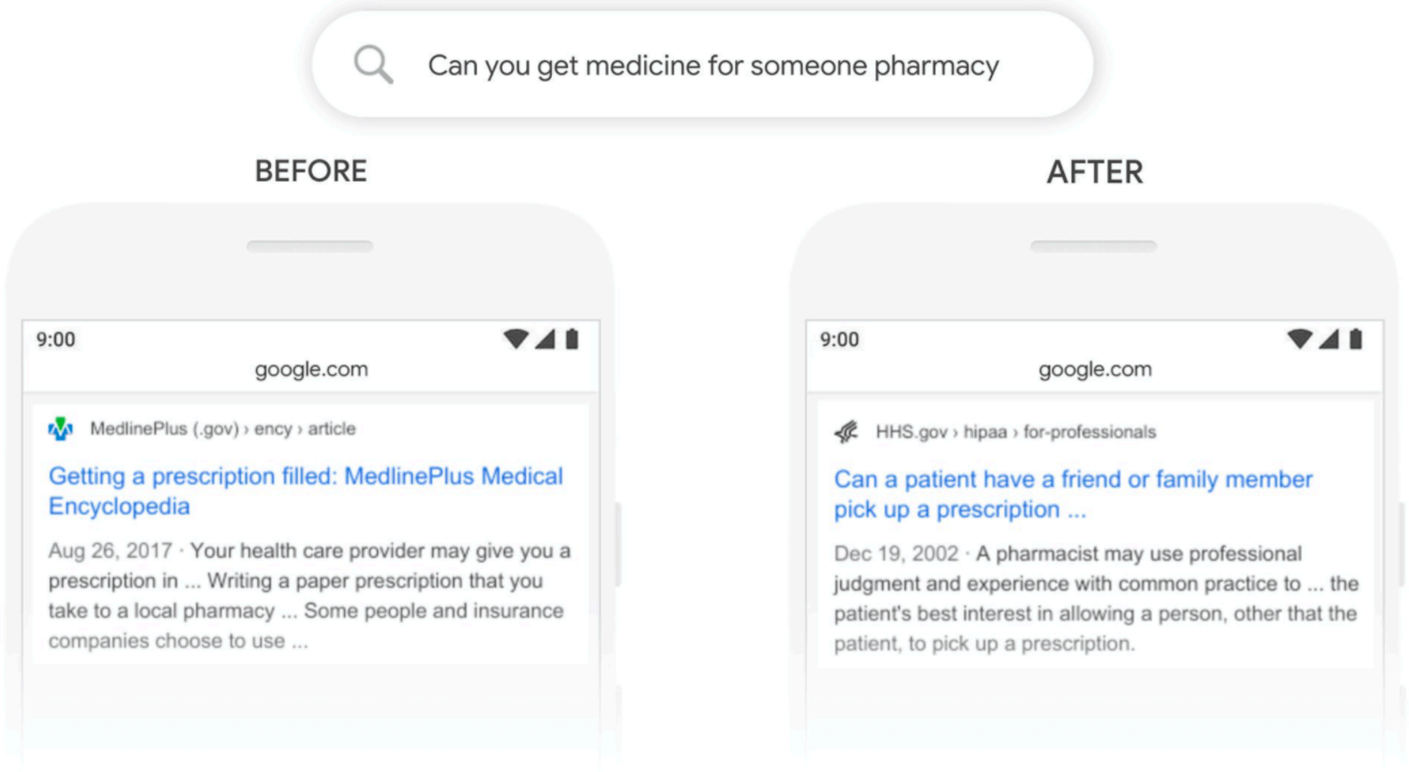
$$P(x_i \mid x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

Objective : Masked word prediction + next sentence prediction

Data : Book Corpus + English Wikipedia

Roles in Prompting

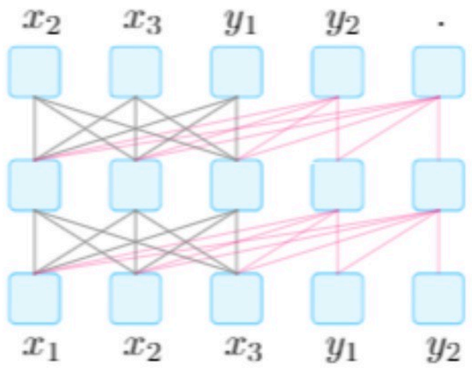
- Usually used with Cloze prompts
- Suitable for Natural Language Understanding, which should be reformulated into cloze tasks



Post BERT, google understands that for “someone” relates to picking up prescription for someone else

https://blog.google/products/search/search-language-understanding-bert/

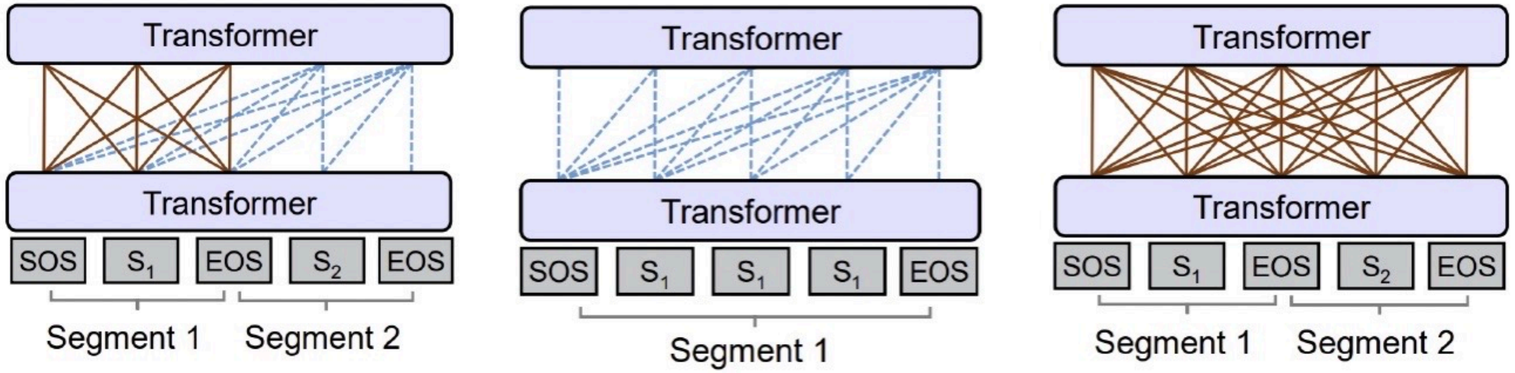
Prefix Language Model



Use a single transformer
but different mask mechanisms

UniLM (Dong et al. 2019)

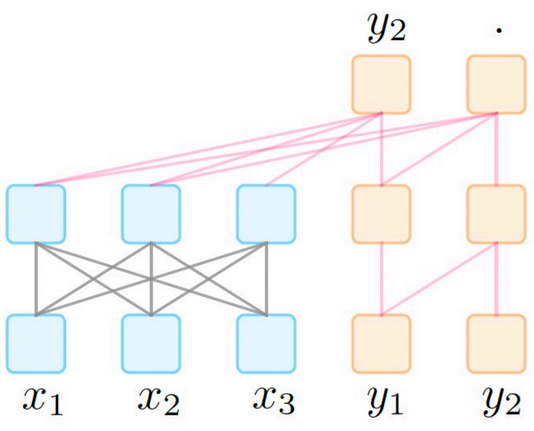
Combination of Left-to Right and Masked LM



Objective : three types of LM, shared parameters

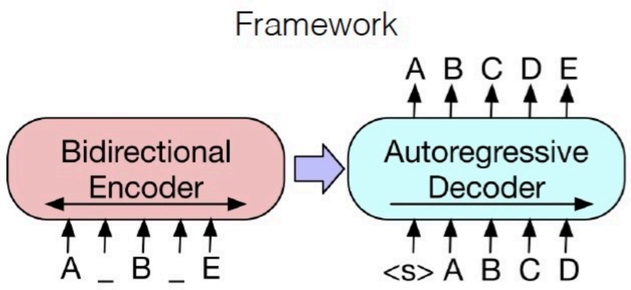
Data : English Wikipedia and Book Corpus

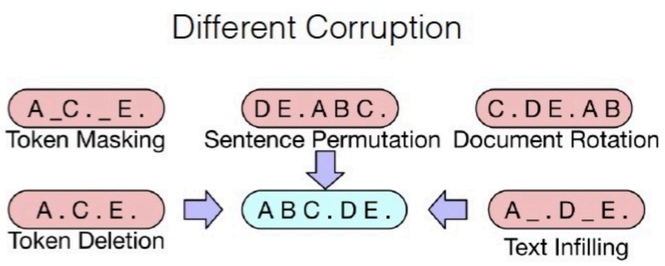
Encoder - Decoder LM



Use a two transformers and two different mask mechanisms

BART (Lewis et al. 2019)





Objective : Reconstruct corrupted original sentences

Data : Book Corpus, CC News, WebText, stories

Application in Prompting

Conditional Text generation (t*ranslation, summarization*)

Generation like tasks (*information extraction, Question-Answering*)

Traditional Formulation vs Prompt Formulation

Input: x = "I love this movie"

Predicting: y = Positive

Input: x = "I love this movie"

Template: [x] Overall, it was a [z] movie

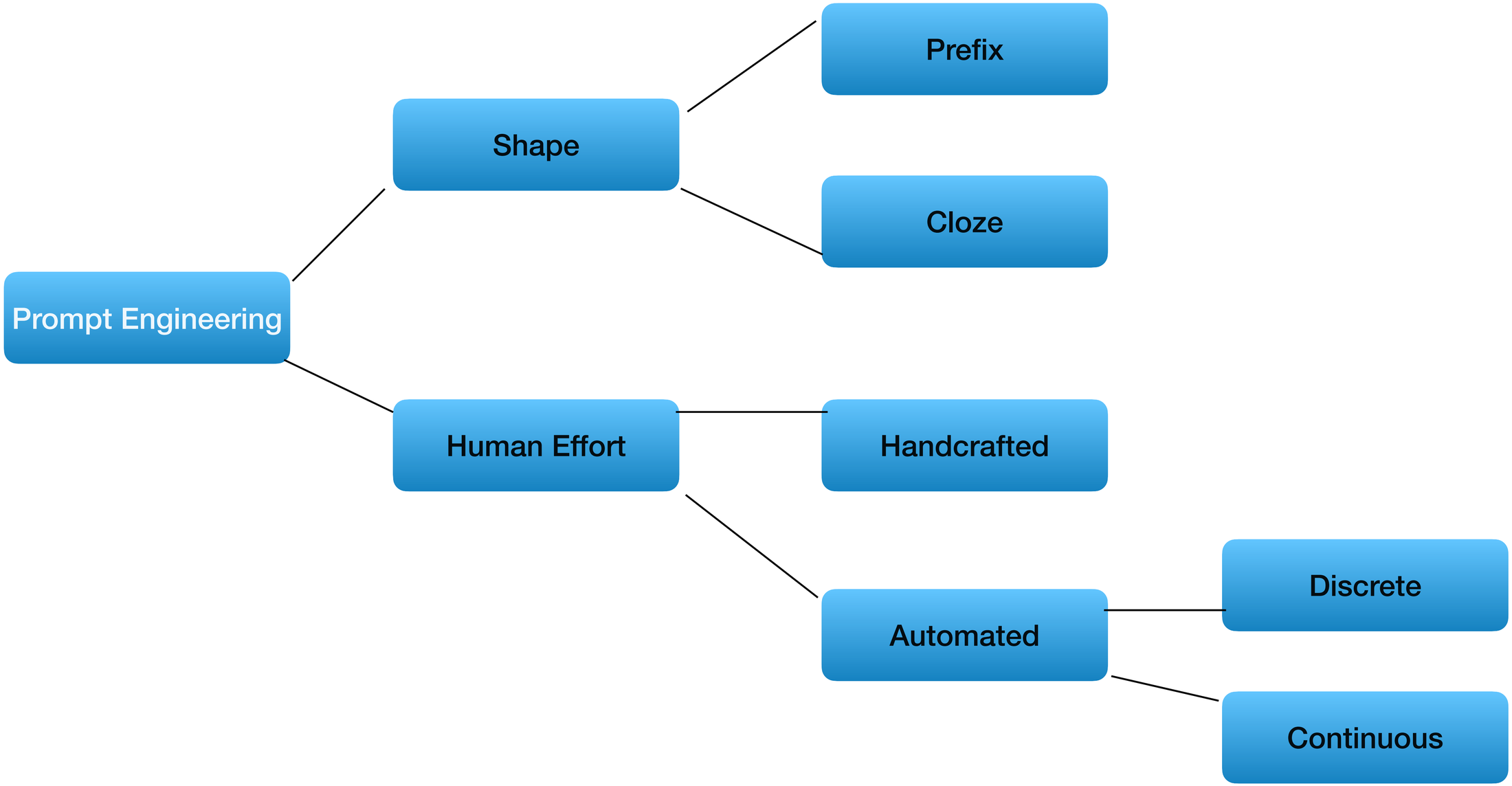
Prompting: x' = "I love this movie. Overall it was a [z] movie."

Predicting: x' = "I love this movie. Overall it was a fantastic movie."

Mapping (answer -> label):
fantastic => Positive

How to define a suitable prompt template ?

Prompt Template Engineering



Representative methods for prompt selection

Prompt Mining (Jiang et al. 2020)

Mine Prompts given a set of questions/answers

Middle Word

Elon Musk was born in South Africa. → [x] was born in [y].

Dependency based

The capital of France is Paris. → capital of [x] is [y].

Select template having highest frequency in corpus

How can we know what language models know?. In ACL

Representative methods for prompt selection

Prompt Paraphrasing (Jiang et al. 2020)

Paraphrase an existing prompt to get other candidates

Example : Back translation with beam search

[x] shares a border with [y]. → en-de model → de-en model → [x] adjoins [y].

[x] has a common border with [y].

Select template which achieves highest accuracy in dev set.

How can we know what language models know?. In ACL

Careful prompt design leads to upto 8.5% increase in fact retrieval accuracy

Prompts	
manual	DirectX *is developed by* y_{man}
mined	y_{mine} *released the* DirectX
paraphrased	DirectX *is created by* y_{para}

Top 5 predictions and log probabilities

	y_{man}		y_{mine}		y_{para}	
1	Intel	-1.06	Microsoft	-1.77	Microsoft	-2.23
2	Microsoft	-2.21	They	-2.43	Intel	-2.30
3	IBM	-2.76	It	-2.80	default	-2.96
4	Google	-3.40	Sega	-3.01	Apple	-3.44
5	Nokia	-3.58	Sony	-3.19	Google	-3.45

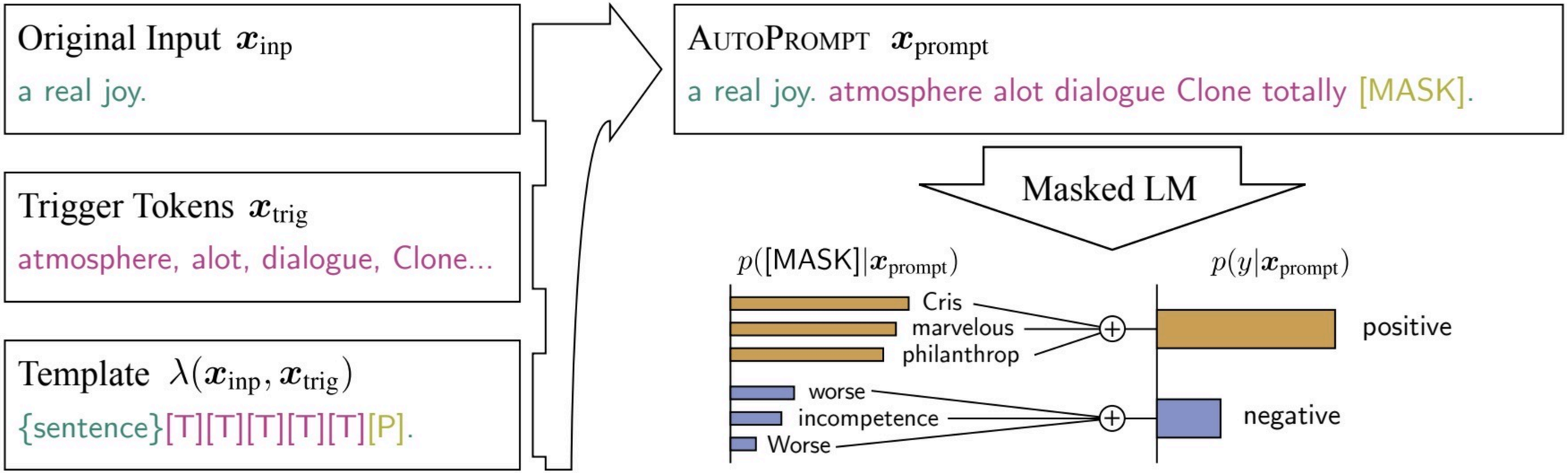
BERT model

How can we know what language models know?. In ACL

Representative methods for prompt selection

Gradient Based Search (Shin et al. 2020)

Automatically optimise arbitrary prompts based on existing words



Prompts generated are not sensible, but they over perform previous methods.

AutoPrompt: Eliciting knowledge from language models with automatically generated prompts. In (EMNLP)

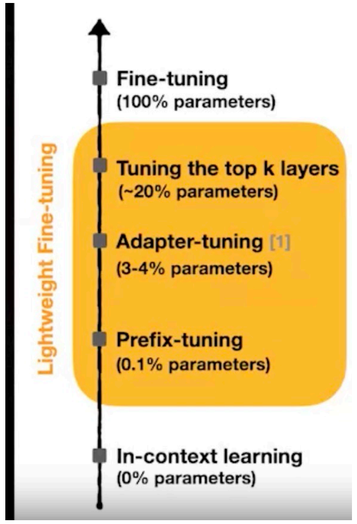
Representative methods for prompt selection

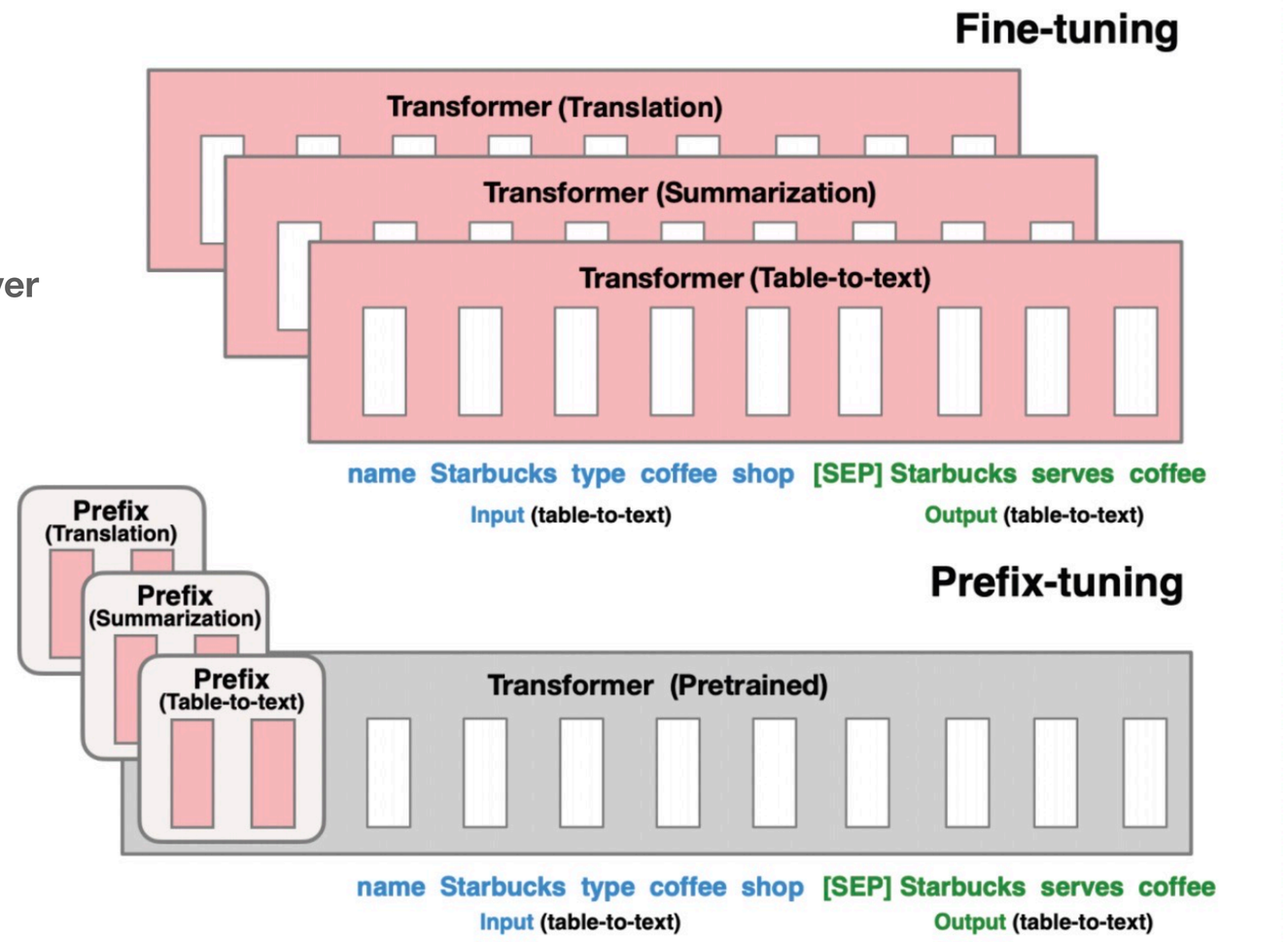
Prefix/ Prompt Tuning (Li et al. 2021, Lester et al. 2021)

Optimize the embeddings of the prompt

Prompt tuning optimizes only the “embedding” layer

Prefix tuning optimizes prefix of all layers

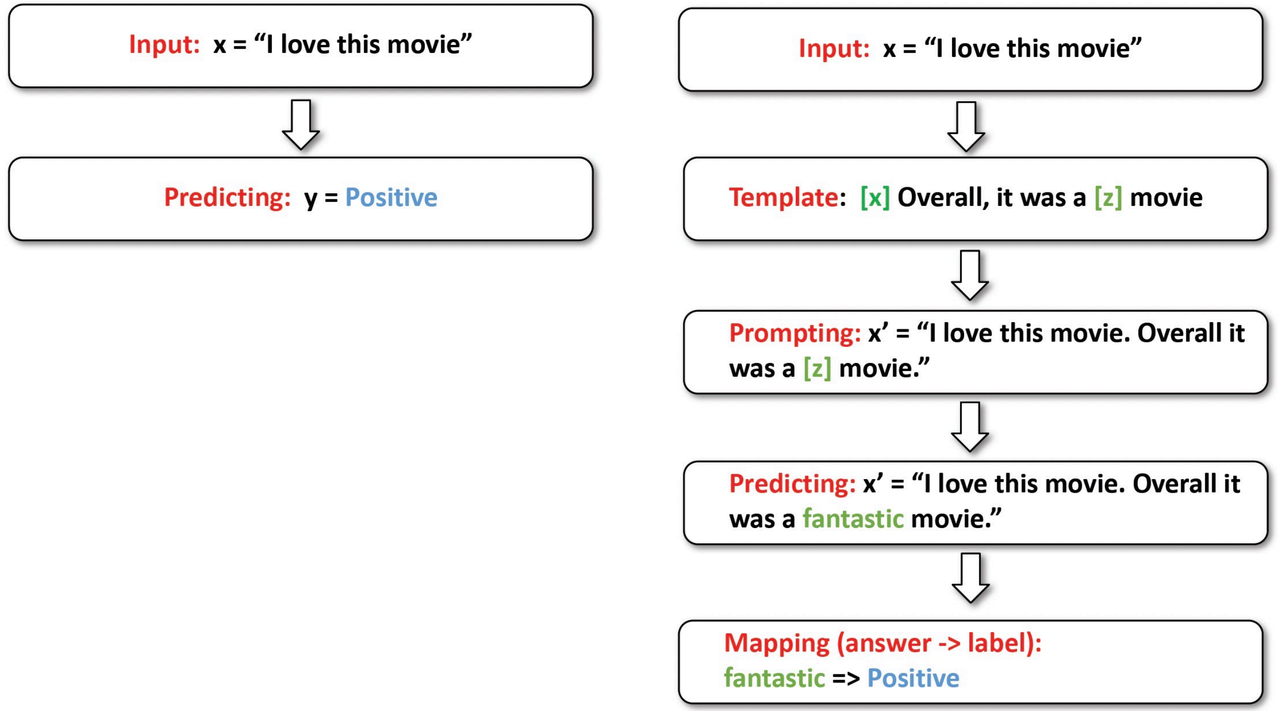




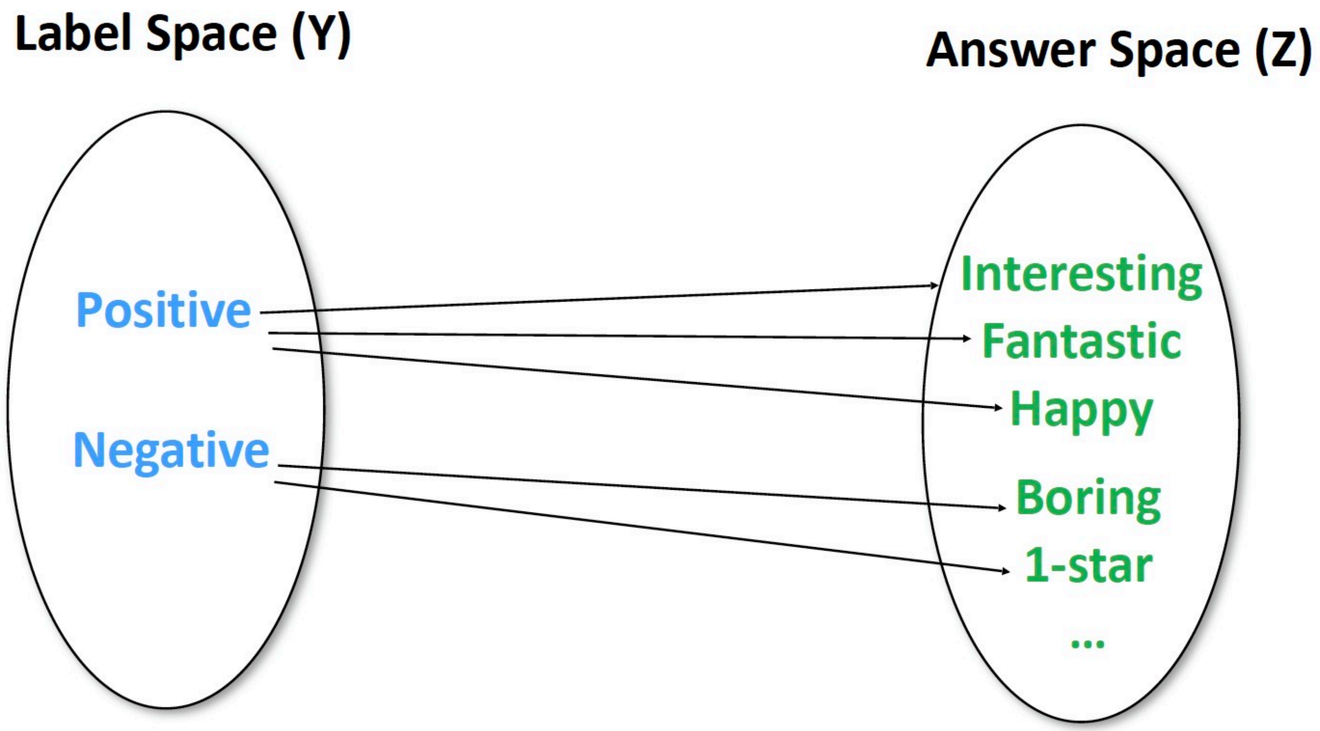
Answer Engineering

We have reformulated the task!, so we should also reformulate "ground truth label"

Definition: Search for Answer space & a Map

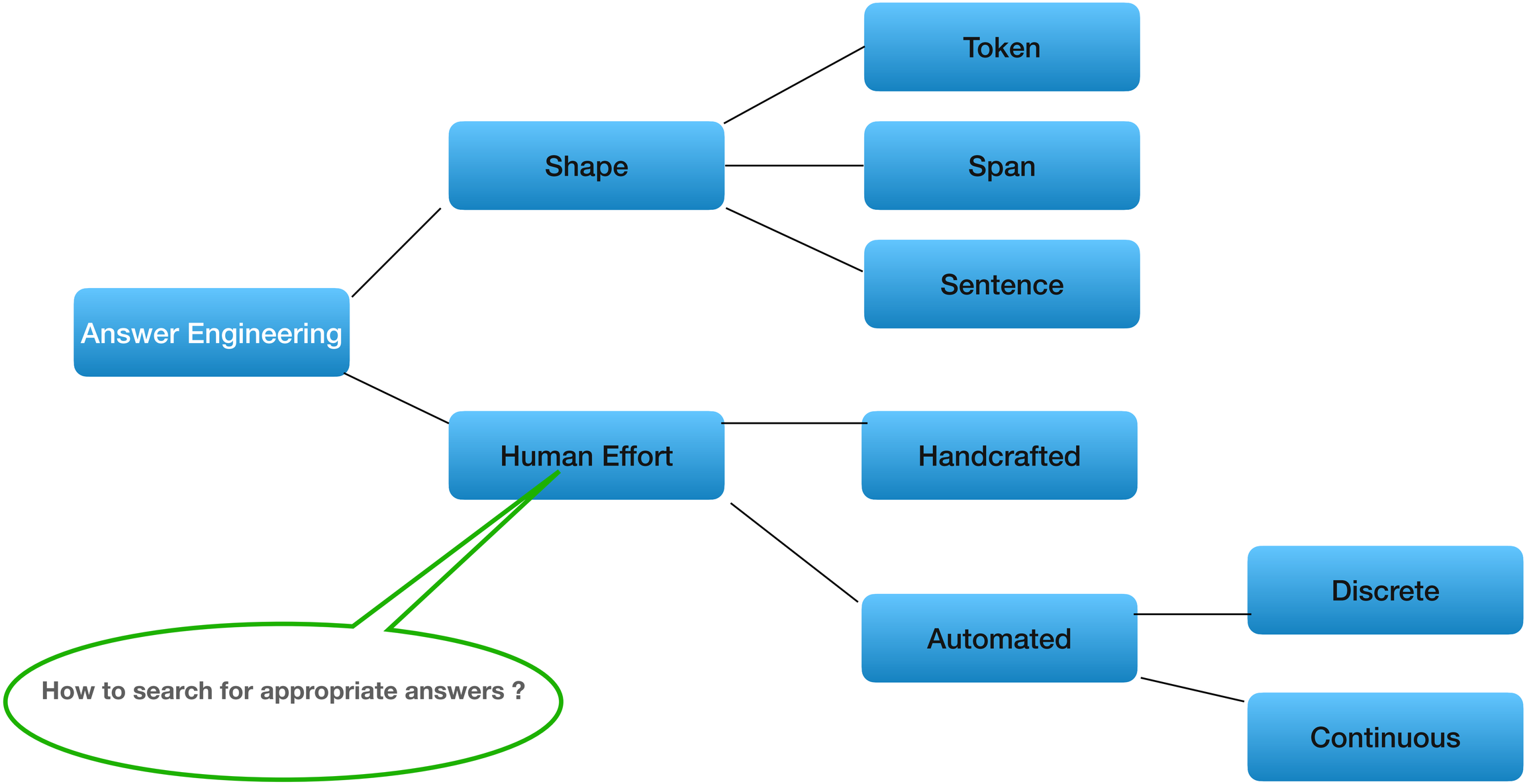


Traditional Formulation vs Prompt Formulation



Ideally, we want Many to One mapping from Answer space to Label Space

Design of Prompt Answer



Answer Shape

Type	Task	Input ([X])	Template	Answer ([Z])
Text CLS	Sentiment	I love this movie.	[X] The movie is [Z].	great fantastic ...
	Topics	He prompted the LM.	[X] The text is about [Z].	sports science ...
	Intention	What is taxi fare to Denver?	[X] The question is about [Z]	quantity city ...
Text-span CLS	Aspect Sentiment	Poor service but good food.	[X] What about service? [Z].	Bad Terrible ...
Text-pair CLS	NLI	[X1]: An old man with ... [X2]: A man walks ...	[X1]? [Z], [X2]	Yes No ...
Tagging	NER	[X1]: Mike went to Paris. [X2]: Paris	[X1] [X2] is a [Z] entity.	organization location ...
Text Generation	Summarization	Las Vegas police ...	[X] TL;DR: [Z]	The victim ... A woman ... ...
	Translation	Je vous aime.	French: [X] English: [Z]	I love you. I fancy you. ...

token

Token or span

sentences

Answer Search

Answer Paraphrasing (Jiang et al. 2020)

Paraphrase to expand initial answer space

Example : Back translation with beam search

$< z', y >$ → en-de model → de-en model → $para(z')$

Select answer space which achieves highest accuracy in dev set.

How can we know what language models know?. In ACL

Answer Search

Prune-then-Search

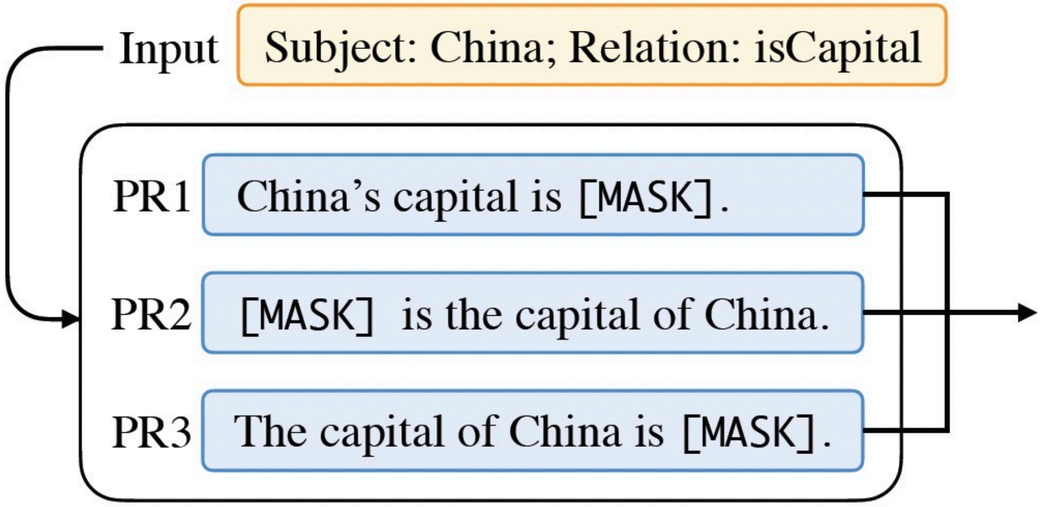
Generate several candidate answers using Paraphrasing

Device an algorithm that searches on this pruned space to select final set of answers

Multi- Prompt learning

Prompt Ensembling

Definition : Use multiple unanswered prompts for an input at inference time to make predictions



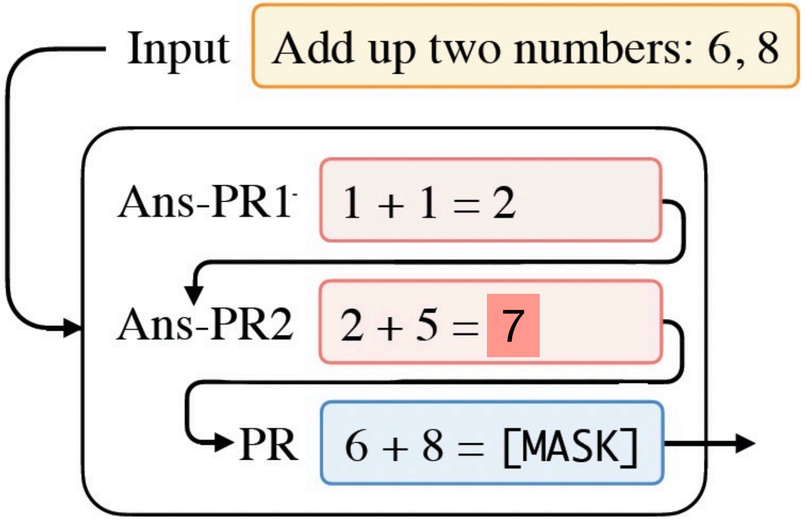
Uniform/Weighted Average

Majority Voting

Multi- Prompt learning

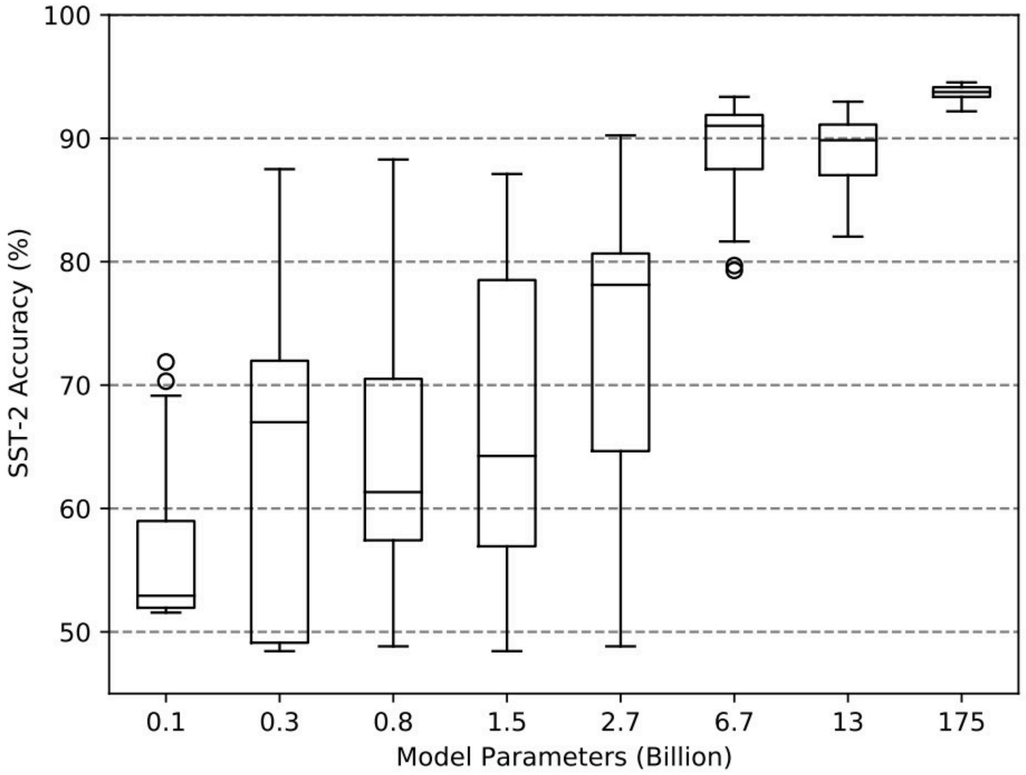
Prompt Augmentation/ InContext Learning

Help the model by providing additional answered prompts

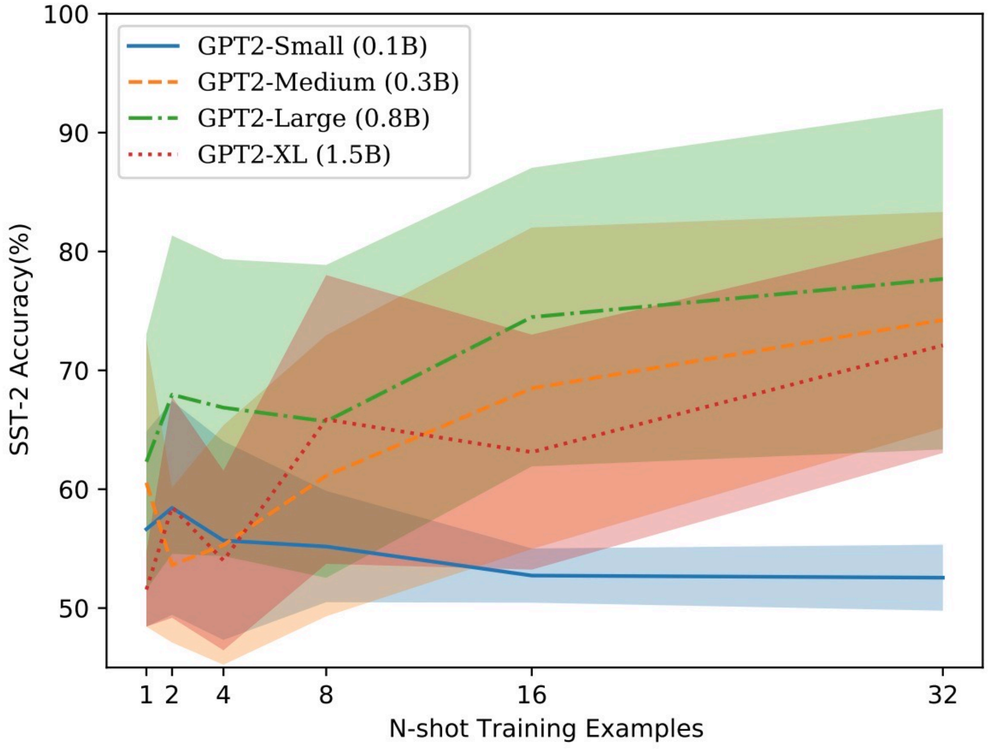


- Selection of answered prompts
- Ordering of answered prompts (Yao et. al 2021)

Prompt ordering Matters!



Four prompts, 24 permutations



Increasing number of prompts does not help !

Fantastically Ordered Prompts and Where to Find Them: Overcoming Few-Shot Prompt Order Sensitivity, Yao et al (2022)

Chain of Thought Prompting

Wei et. al (2022)

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ✗

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had 23 - 20 = 3. They bought 6 more apples, so they have 3 + 6 = 9. The answer is 9. ✓

Large Language Models are zero shot reasoners

Kojima et. al (2022)

(a) Few-shot

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?
A: The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

(Output) The answer is 8. ✗

(b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are 16 / 2 = 8 golf balls. Half of the golf balls are blue. So there are 8 / 2 = 4 blue golf balls. The answer is 4. ✓

(c) Zero-shot

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: The answer (arabic numerals) is

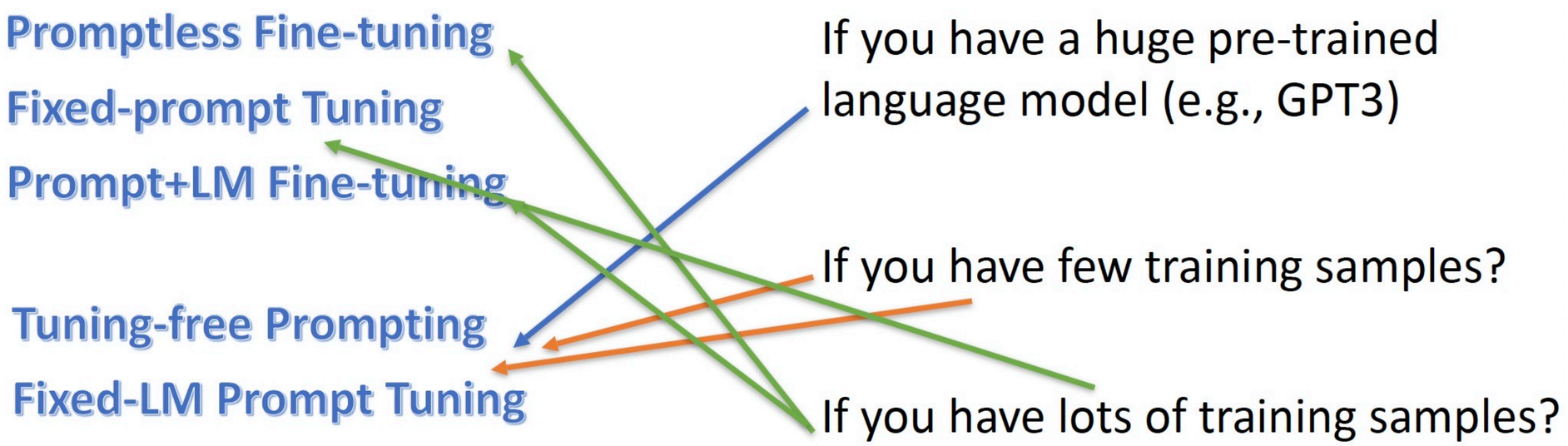
(Output) 8 ✗

(d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: ***Let's think step by step.***

(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓

Prompt based training : Parameter perspective

Strategy	LM Params Tuned	Additional Prompt Params	Prompt Params Tuned	Examples
Promptless Fine-Tuning	Yes	N/A	N/A	BERT Fine-tuning
Tuning-free Prompting	No	No	N/A	GPT-3
Fixed-LM Prompt Tuning	No	Yes	Yes	Prefix Tuning
Fixed-prompt LM Tuning	Yes	No	N/A	PET
Prompt+LM Fine-tuning	Yes	Yes	Yes	PADA

Too Many, Difficult to select?