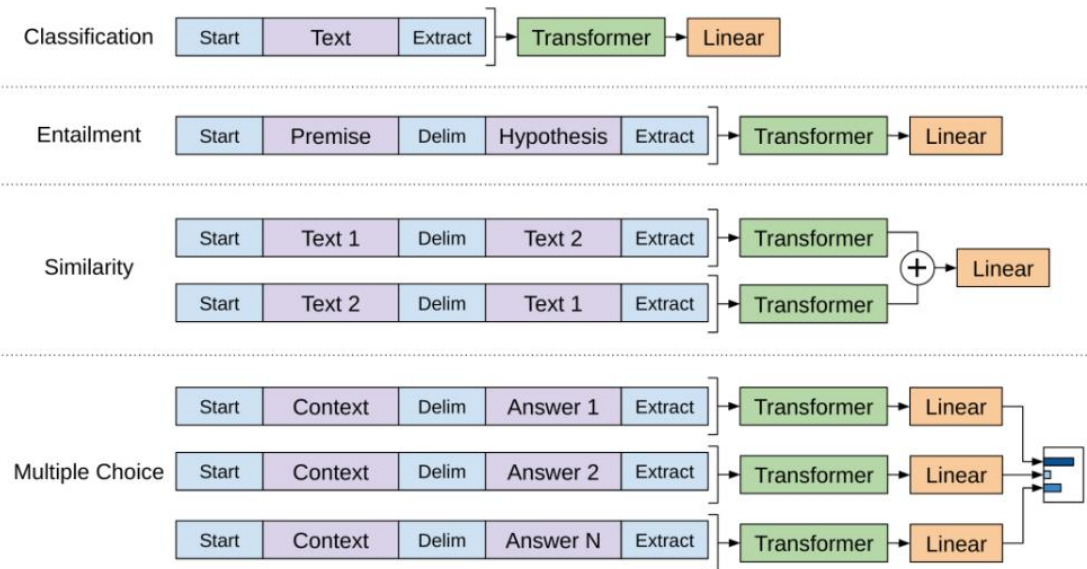
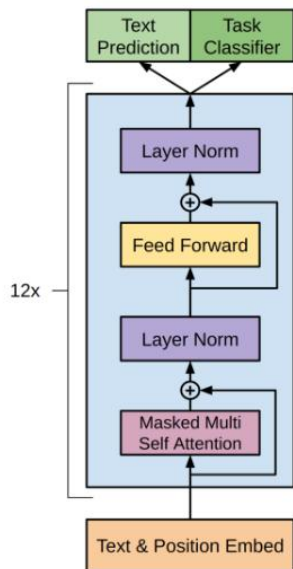
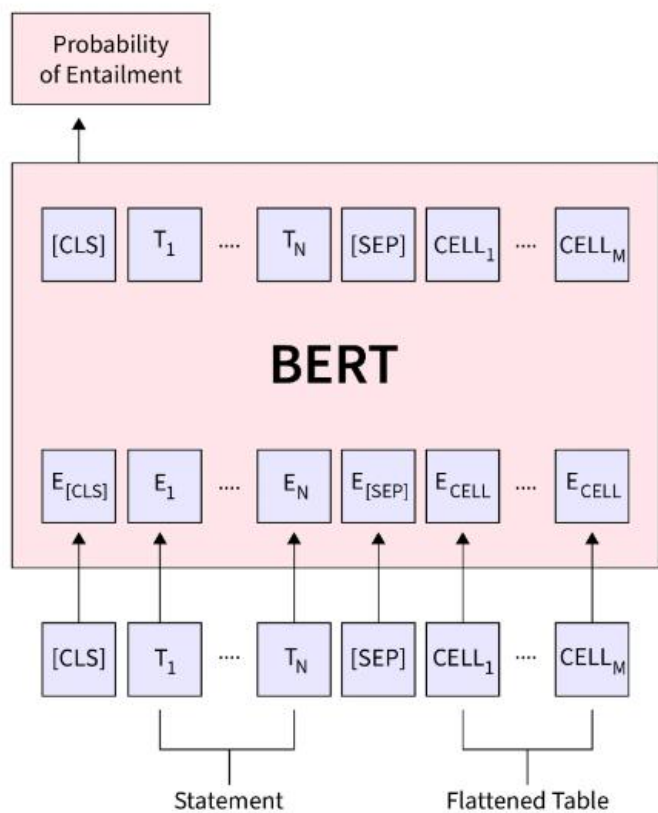


A conceptual image showing a young child wearing a pilot's cap and goggles, sitting on the shoulder of a large, white, humanoid robot. The robot is looking towards a large, glowing world map in the background, which is set against a blue sky with streaks of light. The child is pointing towards the map. The title 'Prompt Tuning' is overlaid on the right side of the image.

# **Prompt Tuning**

# 课前准备

- 基于BERT或GPT完成一个Fine-tune实验



# 目录

- Prompt-based fine-tuning
  - Pre-train, fine-tune
  - Prompt
  - In-context learning
  - Prompt-based fine-tuning
- Prompting
  - Workflow of prompting
  - Prompt template
  - Verbalizer

# 目录

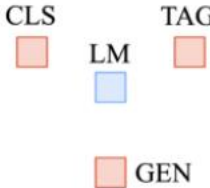
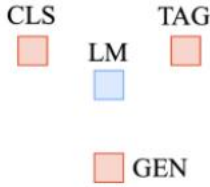
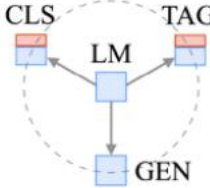
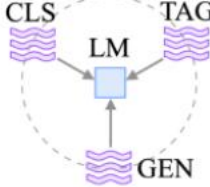
- **Prompt-based fine-tuning**

- Pre-train, fine-tune
- Prompt
- In-context learning
- Prompt-based fine-tuning

- Prompting

- Workflow of prompting
- Prompt template
- Verbalizer

# Pre-train, fine-tune to pre-train, prompt, predict

Paradigm	Engineering	Task Relation
a. Fully Supervised Learning (Non-Neural Network)	Feature (e.g. word identity, part-of-speech, sentence length)	
b. Fully Supervised Learning (Neural Network)	Architecture (e.g. convolutional, recurrent, self-attentional)	
c. Pre-train, Fine-tune	Objective (e.g. masked language modeling, next sentence prediction)	
d. Pre-train, Prompt, Predict	Prompt (e.g. cloze, prefix)	

# Pre-train, fine-tune: PLM

GPT



文字接龍

how are \_\_\_\_\_

BERT



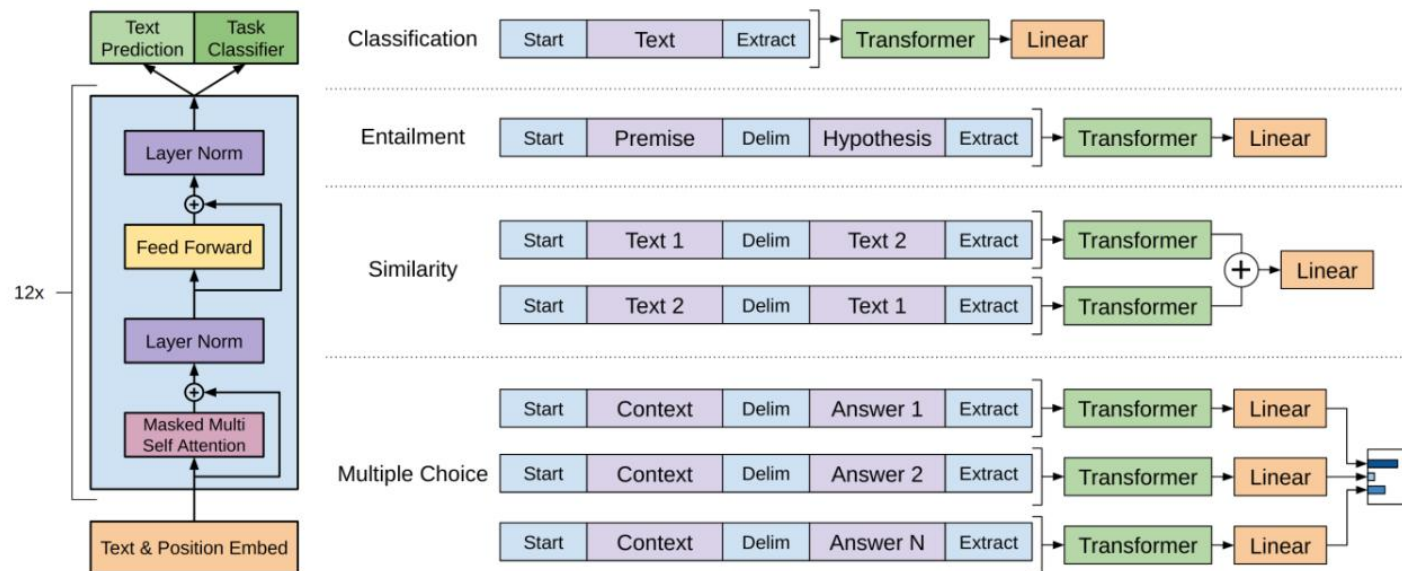
文字填空

how are you

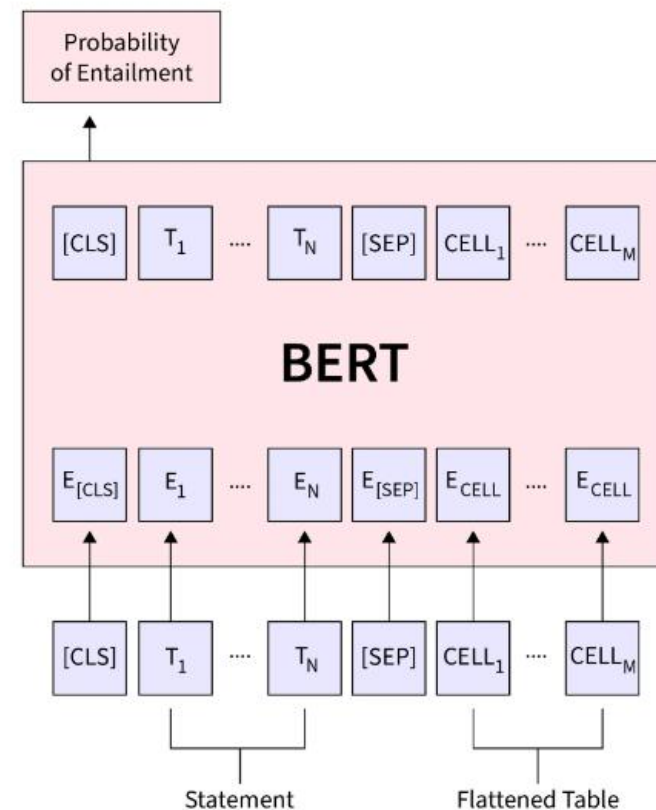


# Pre-train, fine-tune: PLM fine-tuning

## GPT



## BERT

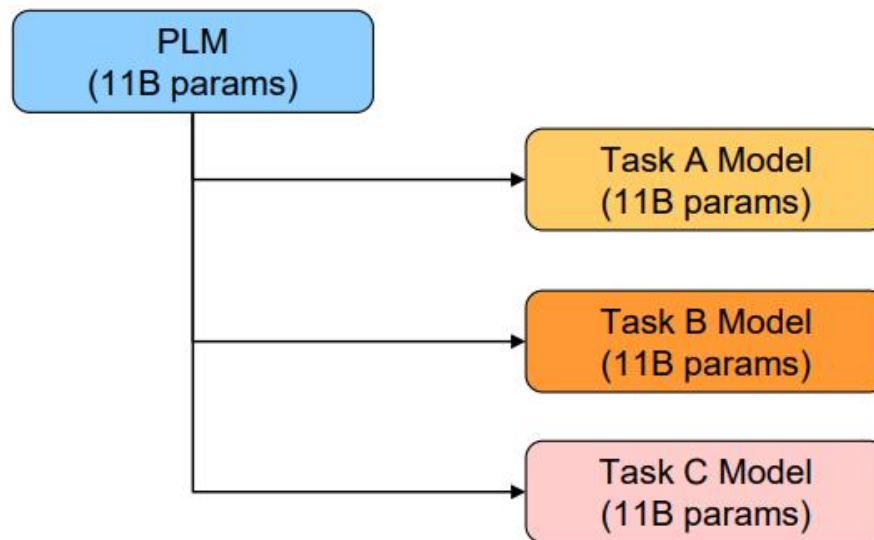


# Pre-train, fine-tune: challenges

- **Data Scarcity:** Downstream annotated data may not be large.

Task	MNLI	QQP	QNLI	SST-2	CoLA	STS-B	MRPC	RTE
Size	391K	363K	108K	67K	8.5K	5.7K	3.5K	2.5K

- **High training cost and space requirement:** Each task requires a copy of a large model.

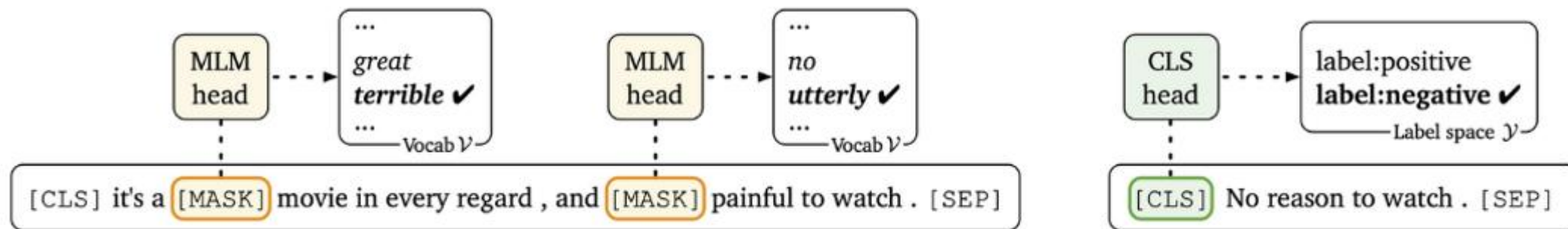




# Pre-train, fine-tune: challenges

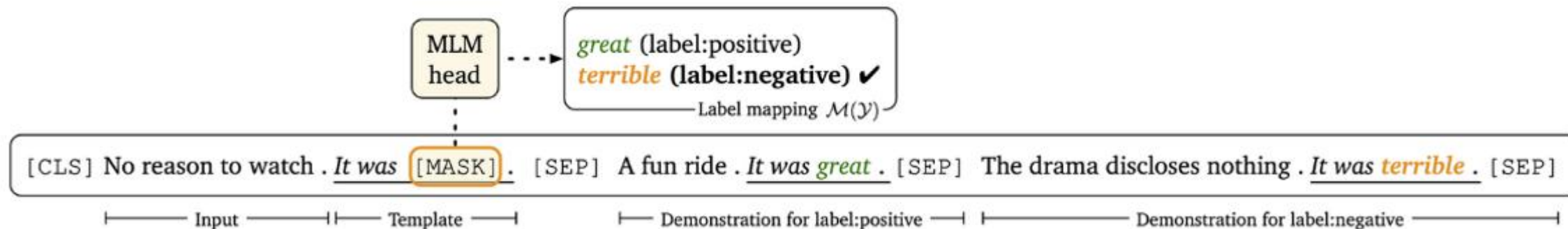
- The gap between the pre-training stage and the downstream task can be significant.

## Why Prompts?



(a) MLM pre-training

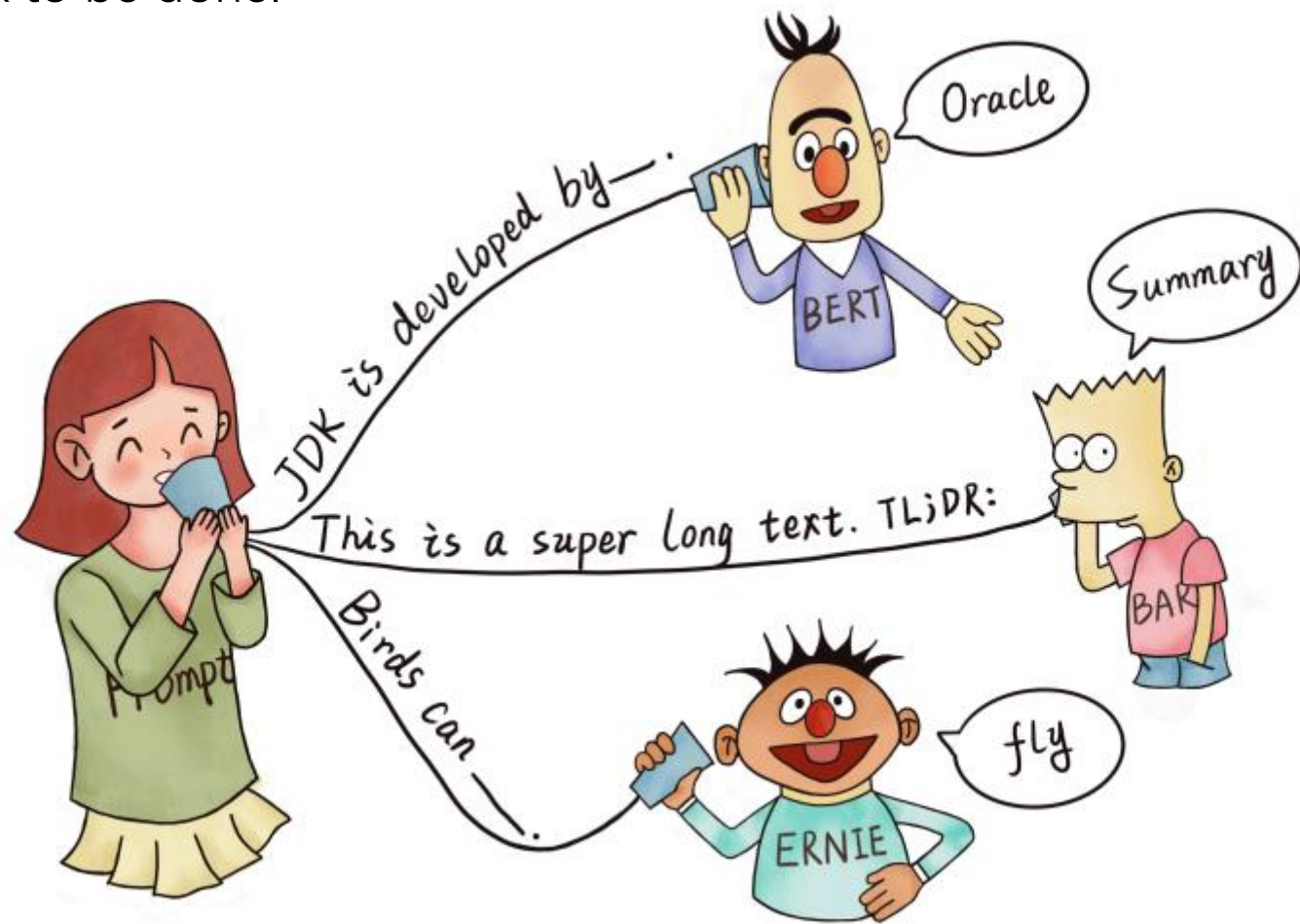
(b) Fine-tuning



(c) Prompt-based fine-tuning with demonstrations (our approach)

# Prompt

Encouraging a pre-trained model to make particular predictions by providing a "prompt" specifying the task to be done.



# In-context learning

## Zero-Shot

```
1 Translate English to French:
2 cheese => .....
```

← task description

← prompt (提示)

## One-Shot

```
1 Translate English to French:
2 sea otter => loutre de mer
3 cheese => .....
```

← task description

← example

← prompt

natural language instruction and/or  
a few task demonstrations

## Few-Shot

```
1 Translate English to French:
2 sea otter => loutre de mer
3 peppermint => menthe poivrée
4 plush girafe => girafe peluche
5 cheese => .....
```

← task description

← examples

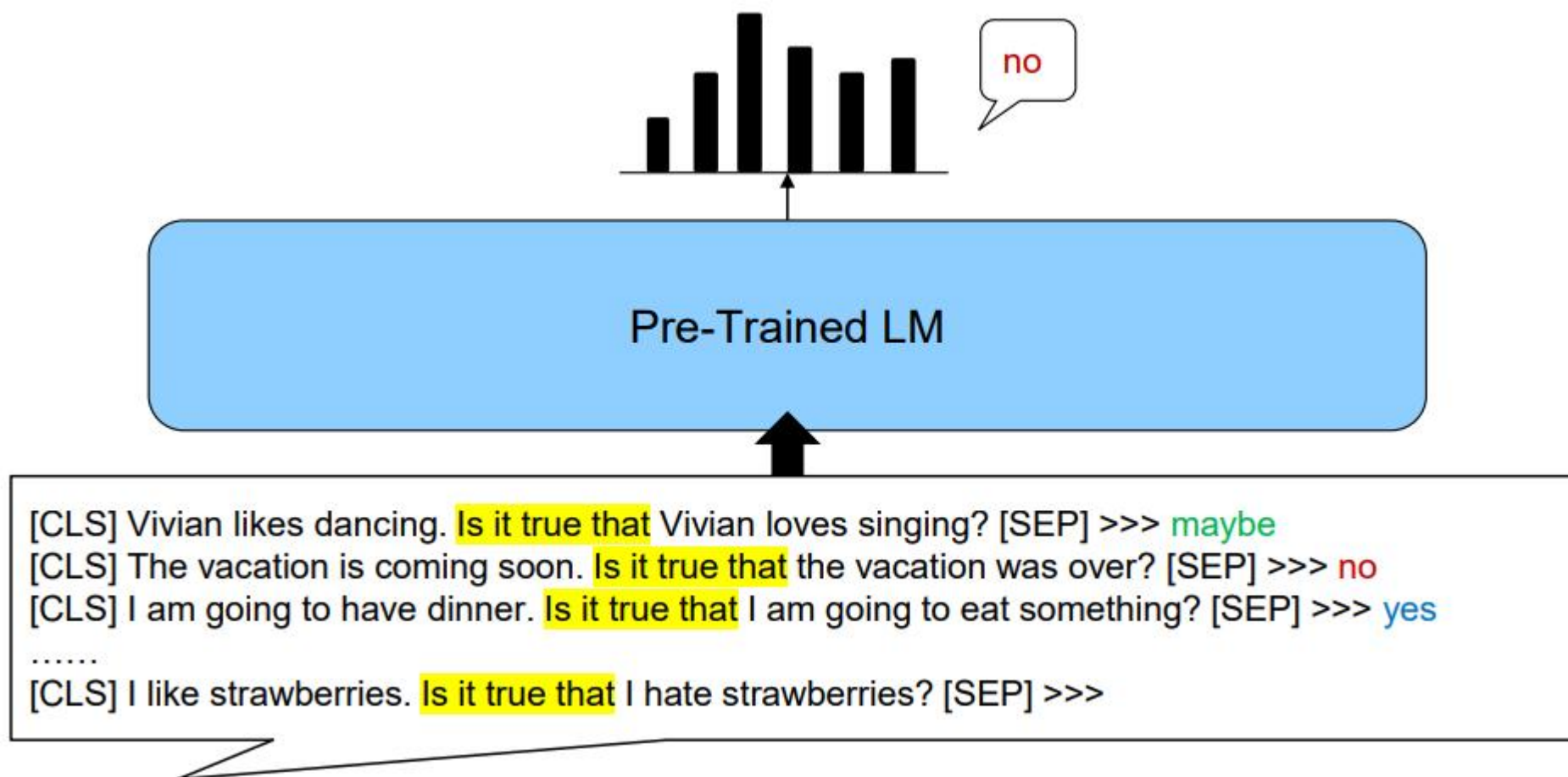
← prompt

no parameter tuning

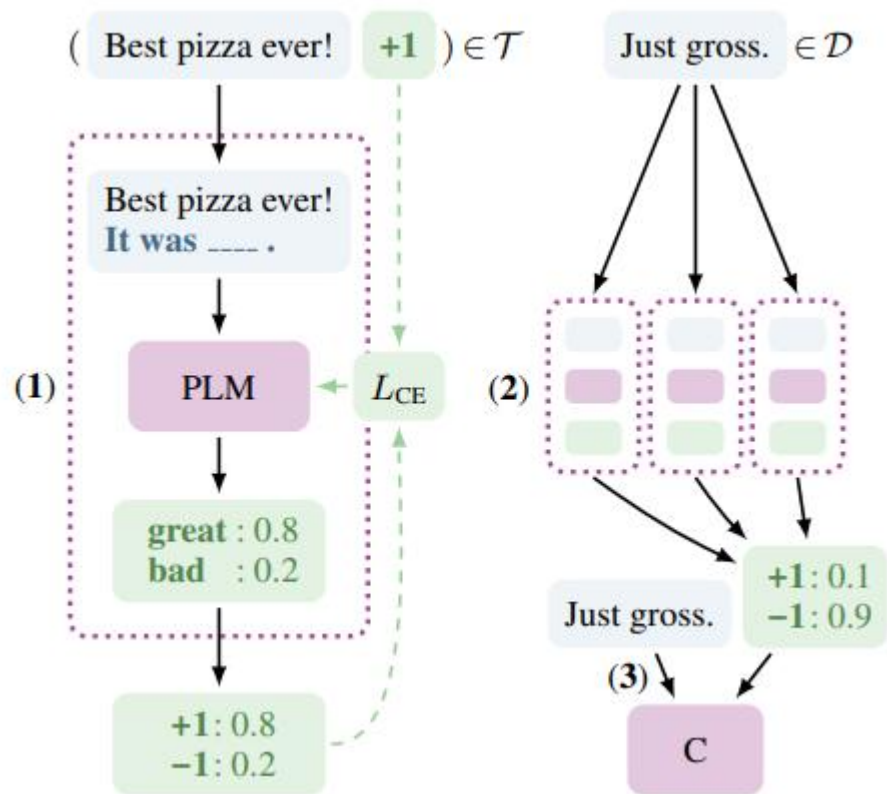
rely on the performance of pre-  
trained language model

# Prompt-based fine-tuning

Convert data into natural language prompts



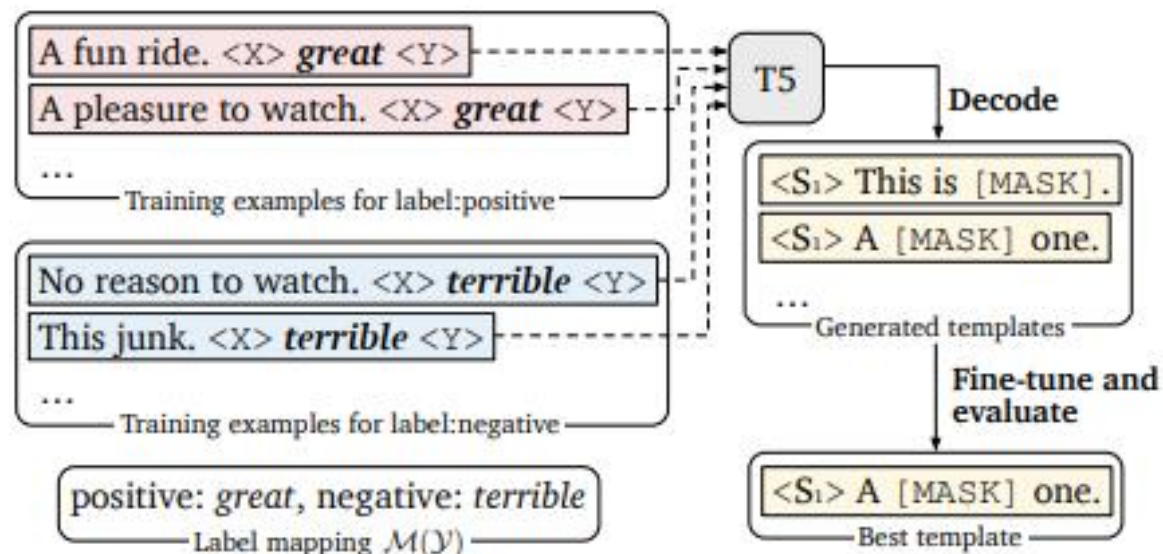
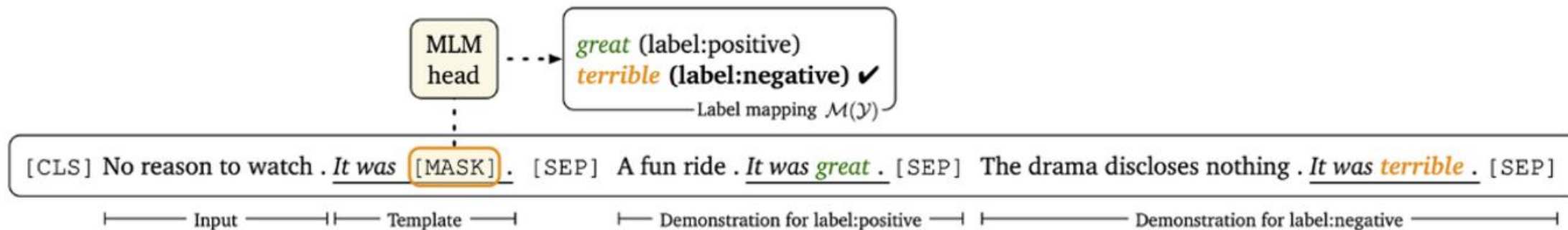
# Hard Prompt: PET



1. Fine-tune a separate pretrained language model (PLM) for each pattern on a small labeled training set
2. The ensemble of trained models annotates unlabeled data.
3. A classifier is trained on the resulting soft-labeled dataset.



# Hard Prompt: LM-BFF



Prompt-based fine-tuning

Automatic generation of Templates

# Hard Prompt: Problems

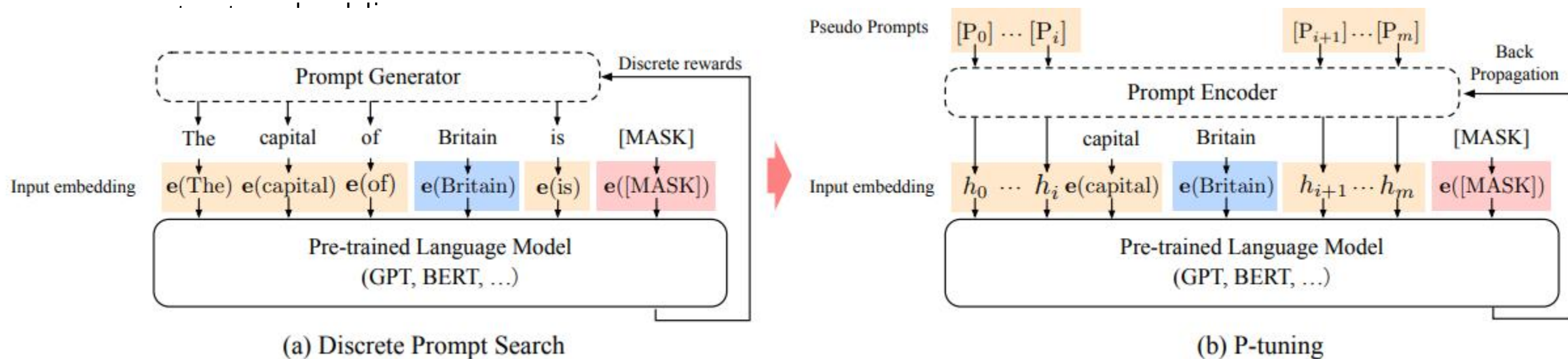
- Task description is error-prone and requires human involvement.
- The effectiveness of a prompt is limited by how much conditioning text can fit into the model's input.

Prompt	P@1
[X] is located in [Y]. ( <i>original</i> )	31.29
[X] is located in which country or state? [Y].	19.78
[X] is located in which country? [Y].	31.40
[X] is located in which country? In [Y].	51.08



# Soft Prompt: P-Tuning

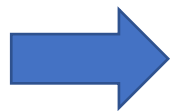
- Automatically search prompts in the continuous space
- Replace the input embeddings of pre-trained language models with its differential



**Prompt:** The capital of ... is ... .

**Context:** Britain

**Target:** MASK



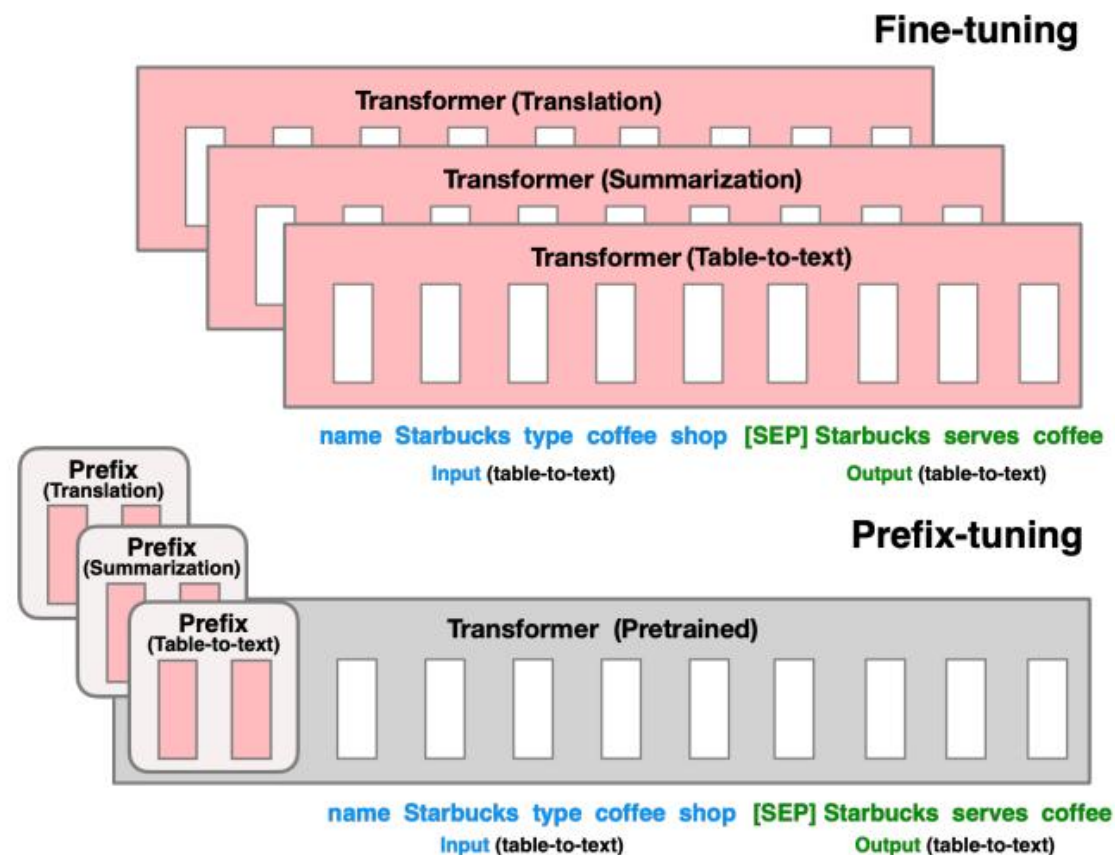
**Prompt:** trainable embedding tensors

**Context:** input embedding from pretrained embedding layer

**Target:** input embedding from pretrained embedding layer

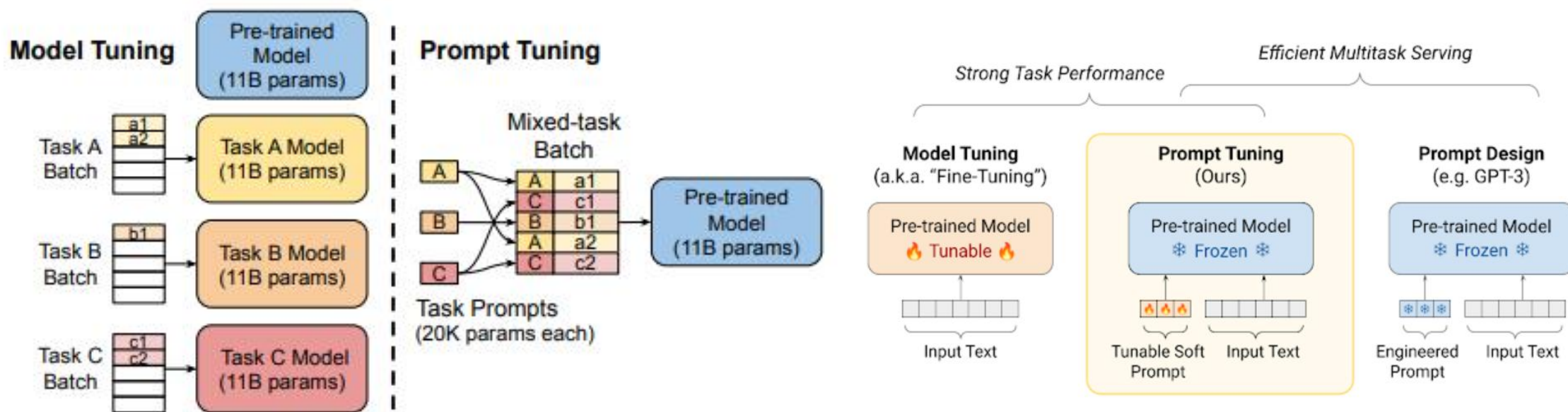
# Soft Prompt: Prefix Tuning

- One prefix activation per downstream task prepended to **each layer in the encoder stack**, including the input layer
- Freeze the Transformer parameters and only optimize the prefix



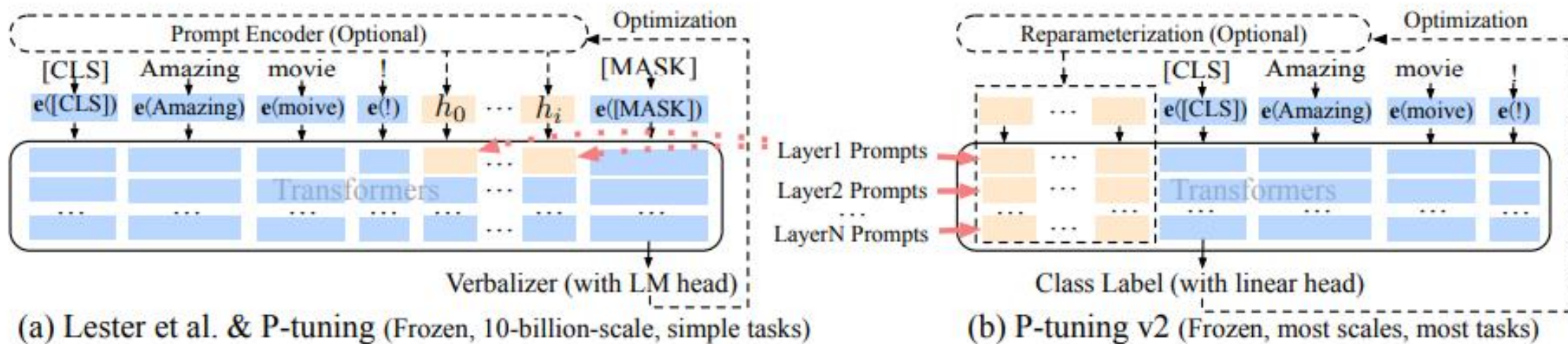
# Soft Prompt: (Soft) Prompt Tuning

- Allow an additional  $k$  tunable tokens per downstream task to be prepended to **the input text**
- No intermediate-layer prefixes or task-specific output layers
- Freeze the entire pre-trained model and only optimize the embedding layer



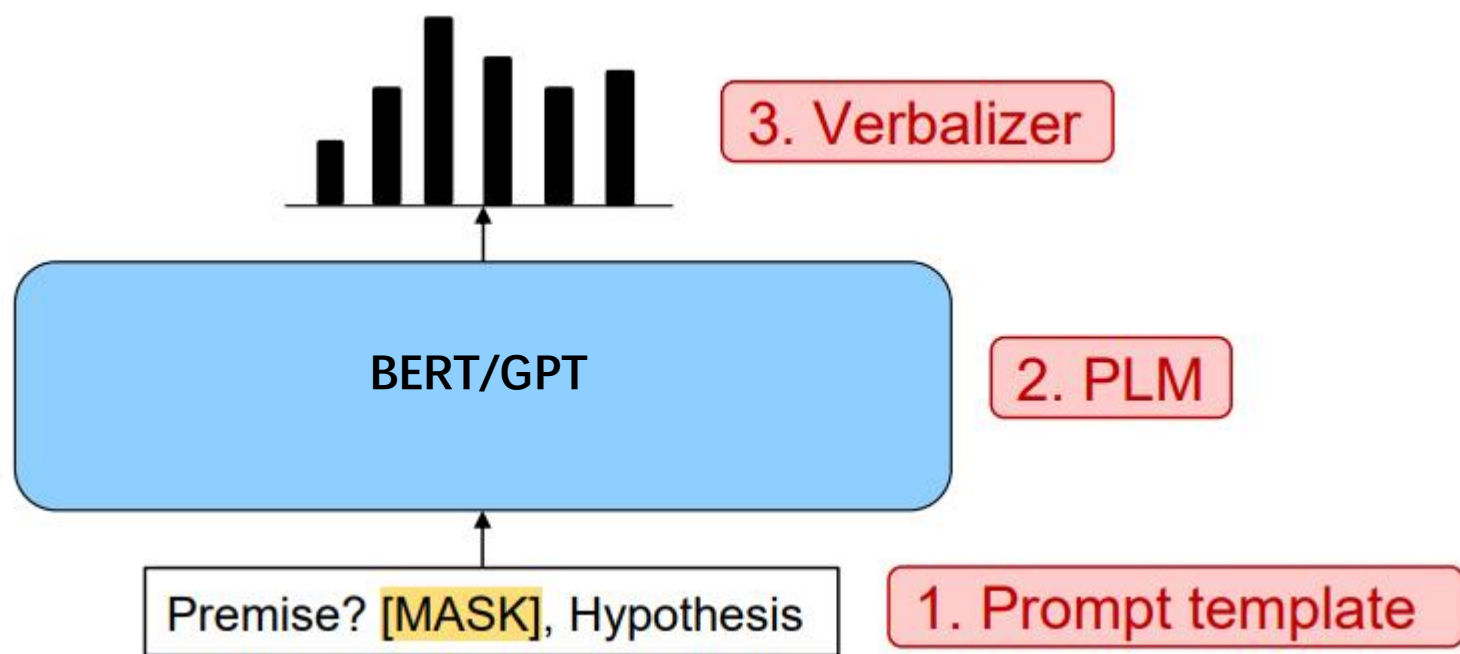
# Soft Prompt: P-Tuning v2

- Address the challenges in P-tuning
  - Number of tunable parameters is limited to sequence length
  - The input embeddings have relatively indirect impact on model predictions
- Prompts in different layers are added as prefix tokens



# 课后作业

- 基于BERT/GPT完成一个使用prompt fine-tuning的作业

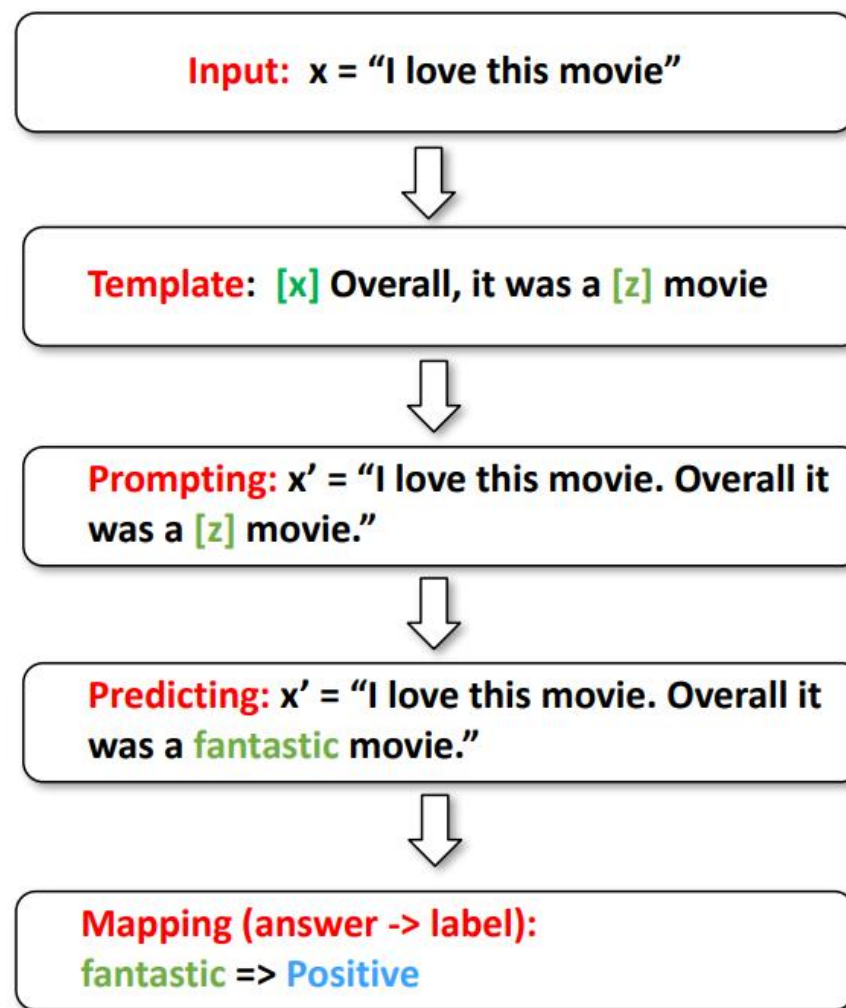
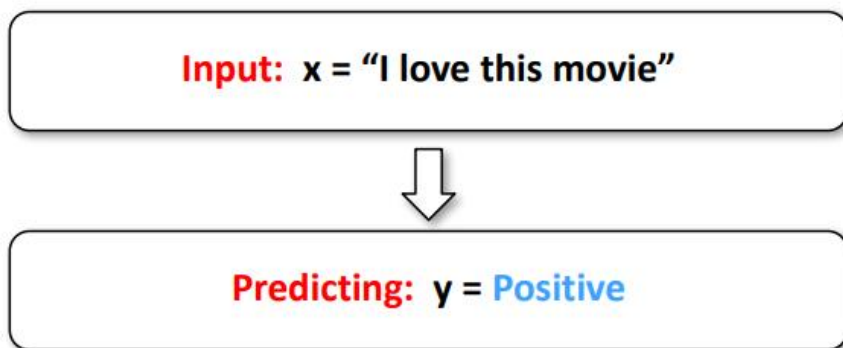


# 目录

- Prompt-based fine-tuning
  - Pre-train, fine-tune
  - Prompt
  - In-context learning
  - Prompt-based fine-tuning
- **Prompting**
  - Workflow of prompting
  - Prompt template
  - Verbalizer

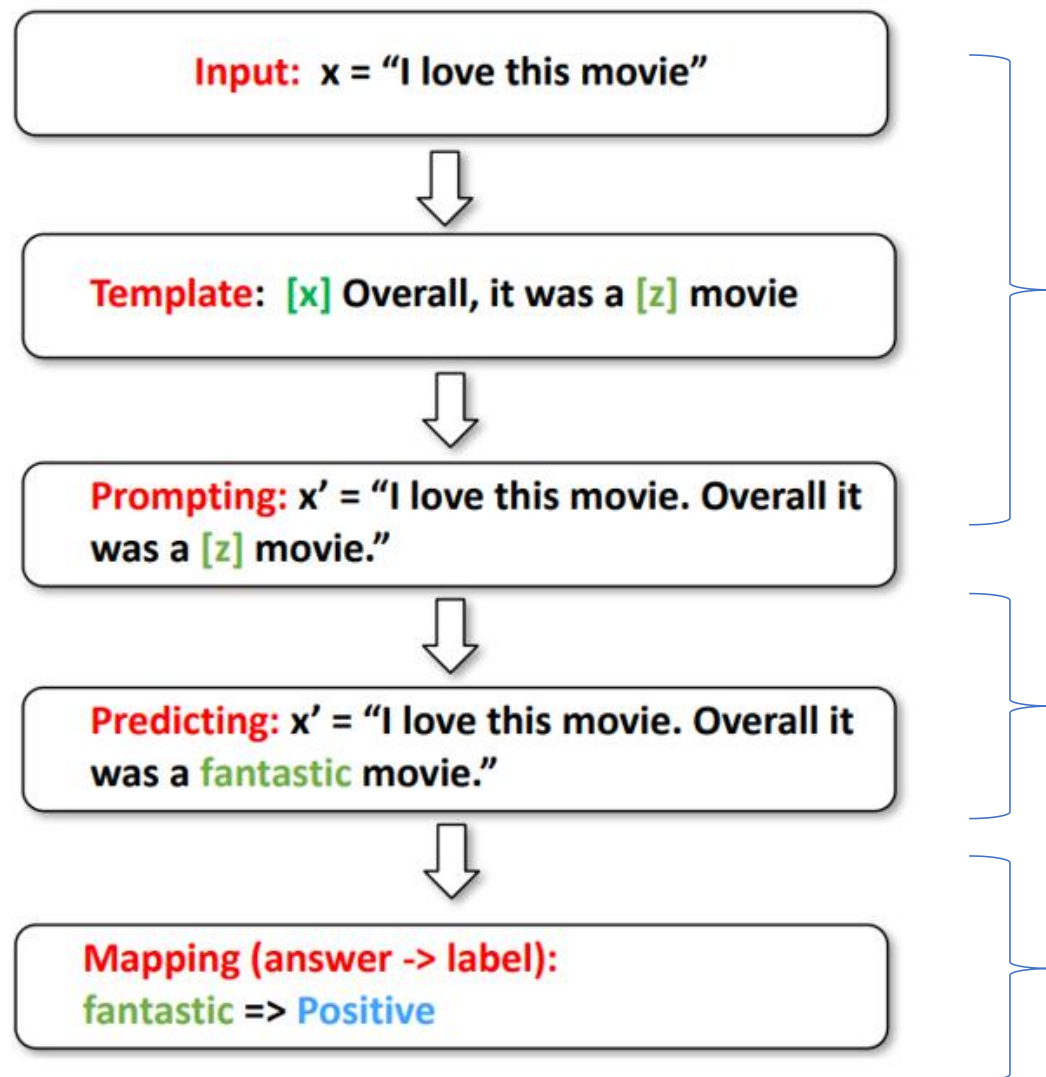


# Traditional formulation vs prompt formulation





# Workflow of prompting



## Prompt Addition

Given input x, we transform it into prompt x' through two steps:

- Define a **template** with two slots, one for input [x], and one for the answer [z]
- Fill in the input slot [x]

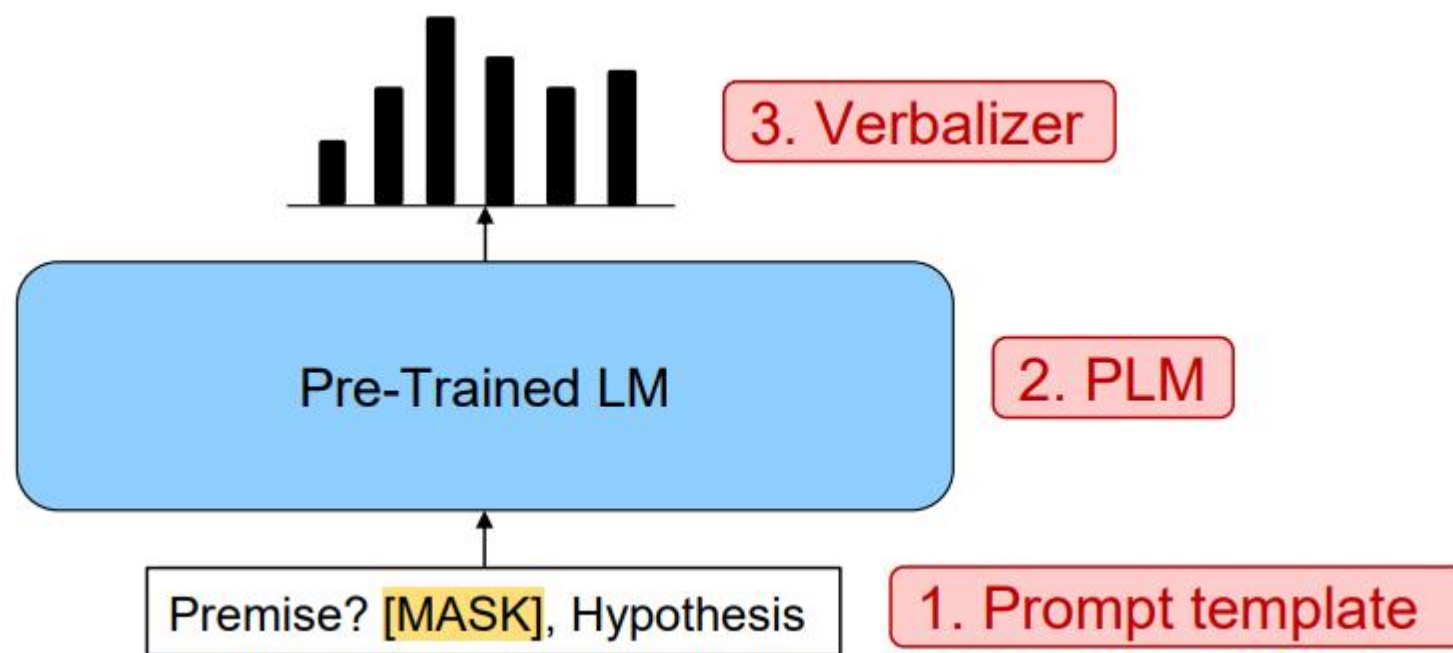
## Answer Prediction

Given a prompt, predict the answer [z]

## Mapping/Verbalizer

Given an answer, map it into a class label

# Workflow of prompting



# Prompt Template

Manually designed natural language input for a task

NLI sample datapoint

**Premise**

Vivian is Jolin's fans

**Hypothesis**

Vivian loves Jolin.

**Label**

0

0: "entailment"

1: "neutral"

2: "contradiction"



[CLS] Vivian is Jolin's fans? [MASK], Vivian loves Jolin.



Premise? [MASK], Hypothesis

prompt template

# Prompt Template

Design of prompt template depends on the task and the pre-trained language model

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. ...

Different prompt templates based on tasks

# Prompt Template

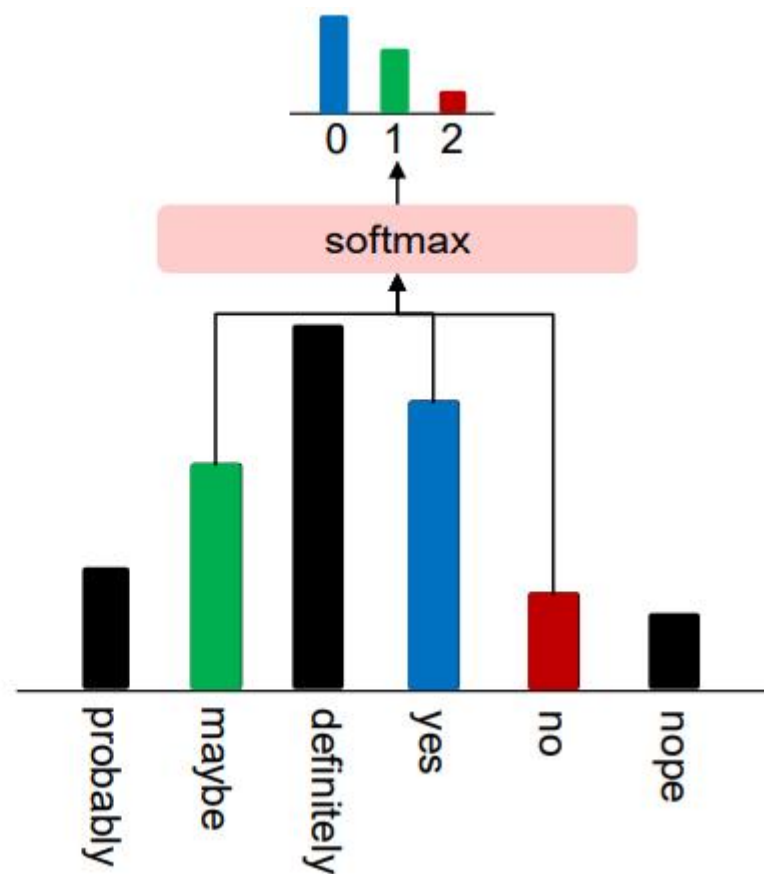
Design of prompt template depends on the task and the pre-trained language model

- **Cloze Prompt:** I love this movie. Overall it was a [z] movie
  - more suitable in downstream tasks using **masked LMs**
- **Prefix Prompt:** I love this movie. Overall this movie is [z]
  - generally used in **text generation tasks**

# Verbalizer

Mapping from predicted vocabulary to labels

0: "entailment" → yes  
1: "neutral" → maybe  
2: "contradiction" → no



# 课后作业

- 试用ChatGLM

### ChatGLM

Chatbot

你好

你好👋！我是人工智能助手 ChatGLM-6B，很高兴见到你，欢迎问我任何问题。

介绍一下华为

华为是一家总部位于中国深圳的科技公司，成立于 1987 年，是全球领先的电信设备制造商之一。华为的产品和服务涵盖了电信网络、智能手机、云计算、数字媒体、智能家居等多个领域，为全球 170 多个国家和地区的客户的产品和服务。

华为一直致力于推动技术创新和产业发展，坚持“自主创新、开放合作”的发展理念，已经成为全球最具影响力的科技公司之一。华为在技术创新方面取得了许多成就，例如在 5G 技术、人工智能、大数据等领域均有卓越的表现。

同时，华为也积极参与公益事业，致力于为社会做出贡献。华为在全球范围内拥有广泛的客户群体和合作伙伴，与多家知名高校和研究机构建立了合作关系，致力于推动技术创新和产业发展。

Input...

提交

Clear History

Maximum length2048

Top P0.7

Temperature0.95