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

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, Graham Neubig

Time: 2021-12-18
Reporter: Xiachong Feng

Background

Authors

Pengfei Liu
Postdoc CMU
PhD Fudan

Weizhe Yuan
PhD Stu CMU
MS CMU

Jinlan Fu
Postdoc NUS
PhD Fudan

Handsome Pic

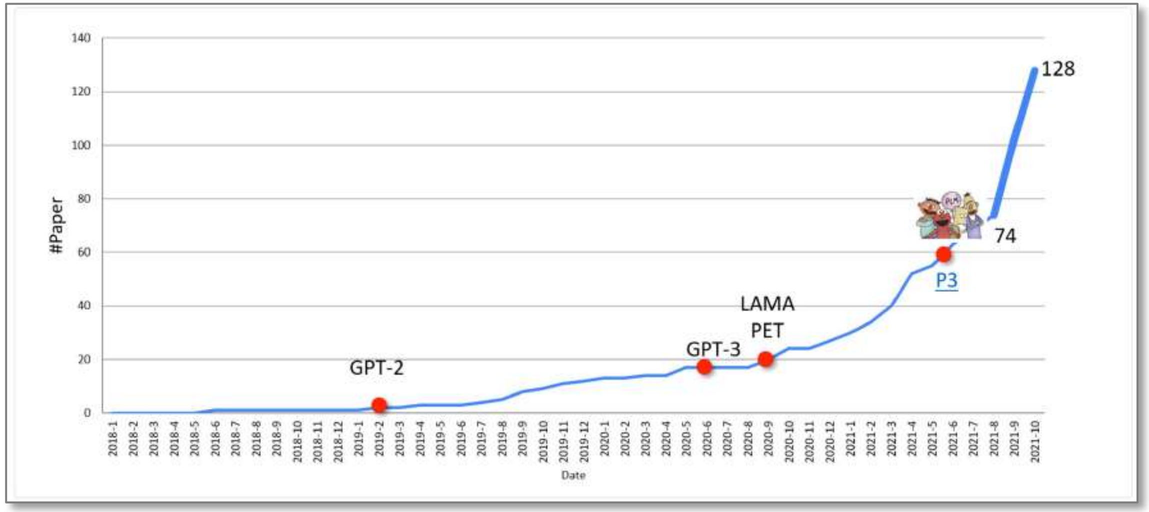
Zhengbao Jiang
PhD Stu CMU

Handsome Pic

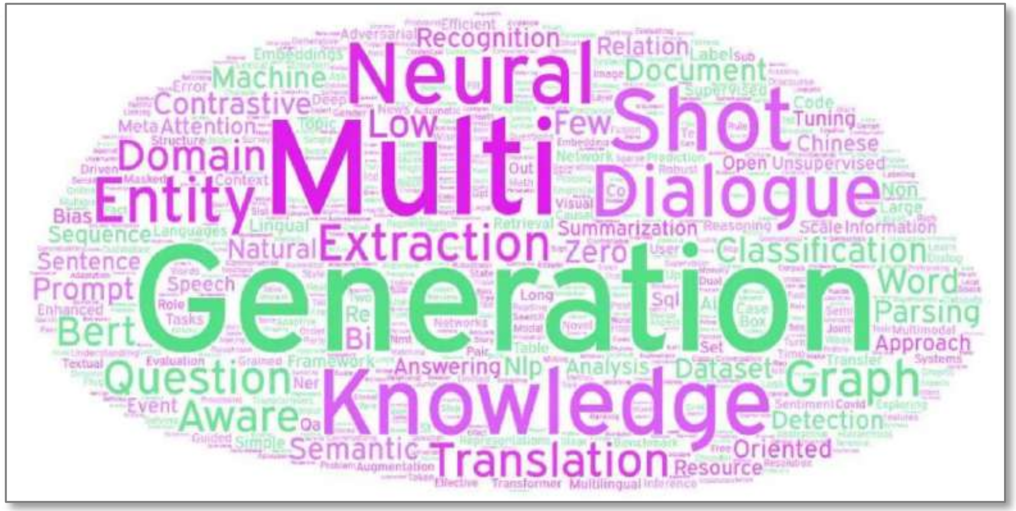
Hiroaki Hayashi
PhD Stu CMU

Graham Neubig
A.P. CMU

Trending of Prompt-based Papers



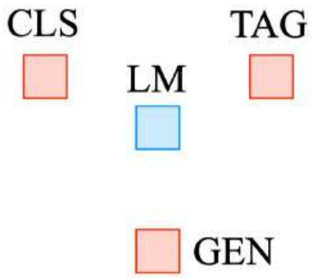
http://pretrain.nlpedia.ai/



ARR Nov.

Four Paradigms in NLP

Fully Supervised Learning
(Non-Neural Network)
Features Engineering
TF-IDF, POS, Length



Domain Adaptation to Summarize Human Conversations
ACL 2010

3.2 Features Used

We consider two sets of features for each sentence: a small set of conversational structure features, and a large set of lexical features.

Conversational features: We extract 24 conversational features from both the email and meetings domain, and which consider both emails and meetings to be conversations comprised of turns between multiple participants. For an email thread, a turn consists of a single email fragment in the exchange. Similarly, for meetings, a turn is a sequence of dialogue acts by the same speaker. The conversational features, which are described in detail in (Murray and Carenini, 2008), include sentence length, sentence position in the conversation and in the current turn, pause-style features, lexical cohesion, centroid scores, and features that measure how terms cluster between conversation participants and conversation turns.

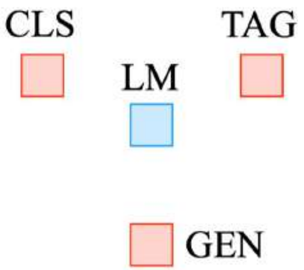
Lexical features: We derive an extensive set of lexical features, originally proposed in (Murray et al., 2010) from the AMI and BC3 datasets, and then compute their occurrence in the Enron corpus. After throwing out features that occur less than five times, we end up with approximately 200,000 features. The features derived are: character trigrams, word bigrams, POS tag bigrams, word pairs, POS pairs, and varying instantiation ngram (VIN) features. For word pairs, we extract the ordered pairs of words that occur in the same sentence, and similarly for POS pairs. To derive VIN features, we take each word bigram w_1,w_2 and further represent it as two patterns p_1,w_2 and w_1,p_2 each consisting of a word and a POS tag.

3.3 Classifier

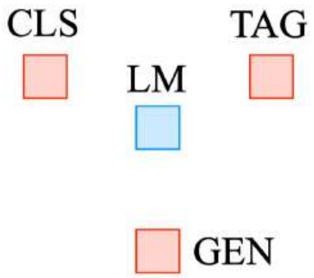
In all of our experiments, we train logistic regression classifiers using the *liblinear* toolkit[3]. This choice was partly motivated by our earlier summarization research, where logistic regression classifiers were compared alongside support

Four Paradigms in NLP

Fully Supervised Learning
(Neural Network)
Architecture Engineering
Conv, RNN, Self-Atten

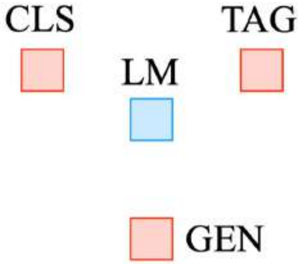


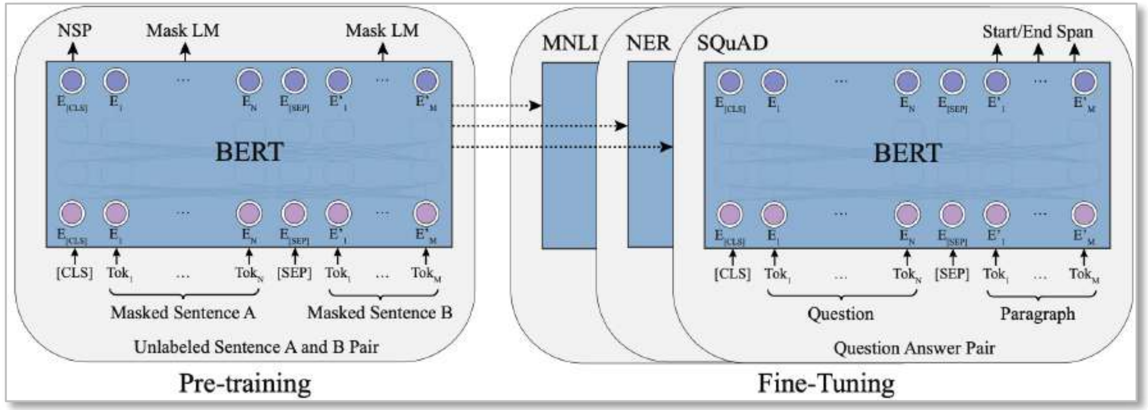
Fully Supervised Learning
(Non-Neural Network)
Features Engineering
TF-IDF, POS, Length



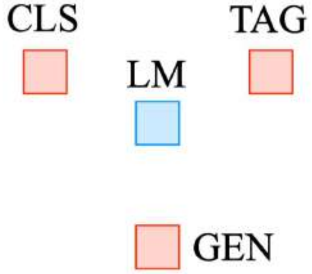
Four Paradigms in NLP

Fully Supervised Learning
(Neural Network)
Architecture Engineering
Conv, RNN, Self-Atten

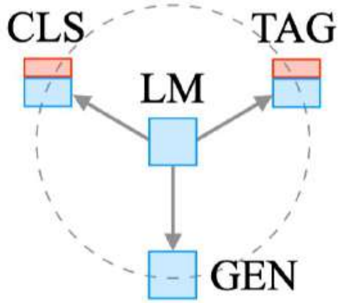




Fully Supervised Learning
(Non-Neural Network)
Features Engineering
TF-IDF, POS, Length

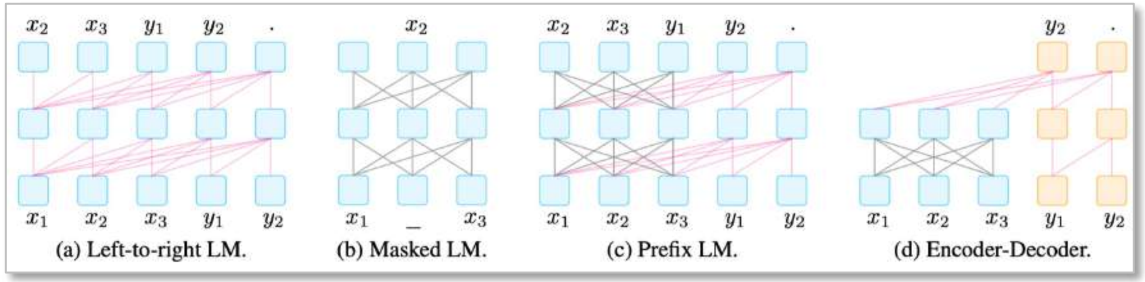


Pre-train, Fine-tune
Objective Engineering
MLM, NSP

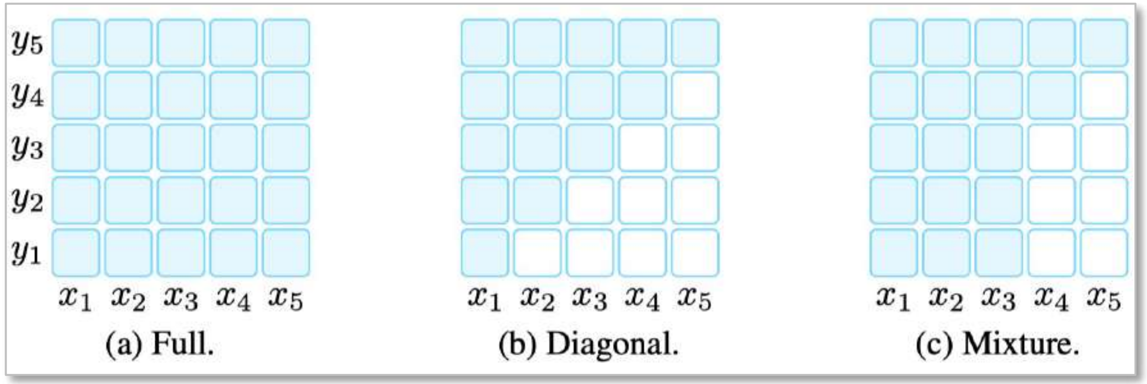


PLMs

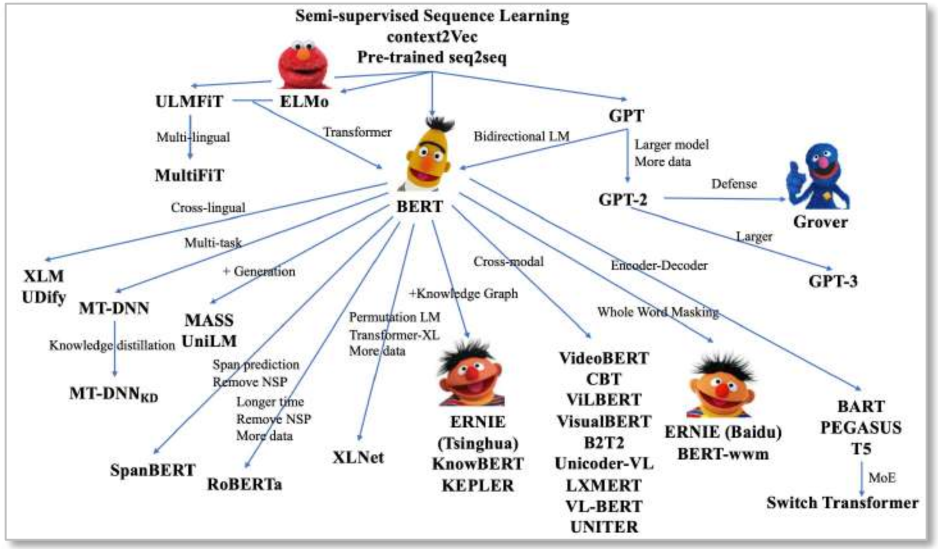
- Typical paradigms of pre-trained LMs

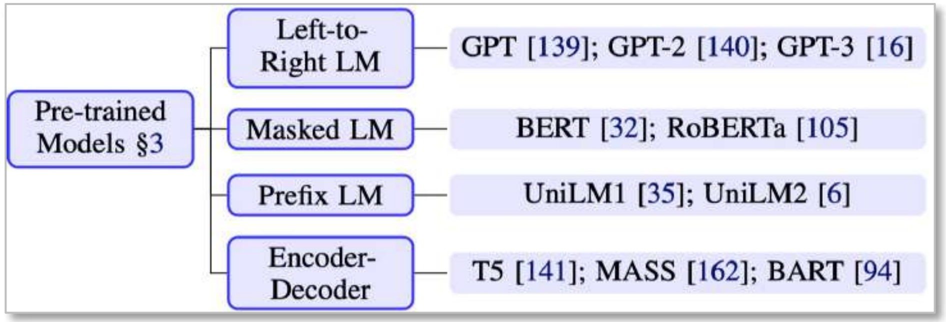


- Attention mask patterns



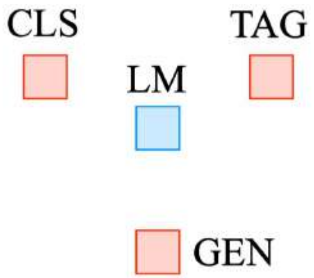
- PLMs



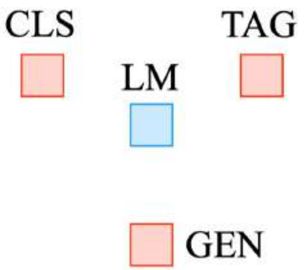


Four Paradigms in NLP

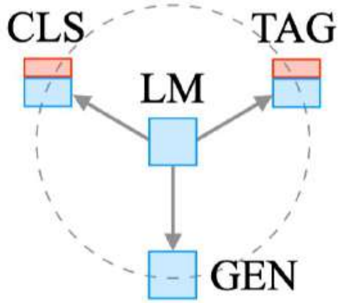
Fully Supervised Learning
(Non-Neural Network)
Features Engineering
TF-IDF, POS, Length



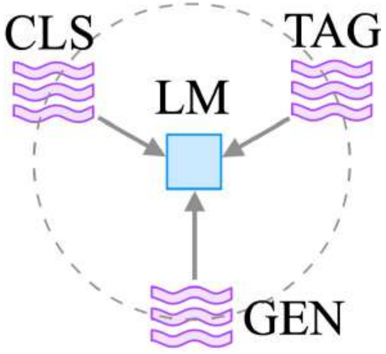
Fully Supervised Learning
(Neural Network)
Architecture Engineering
Conv, RNN, Self-Atten



Pre-train, Fine-tune
Objective Engineering
MLM, NSP

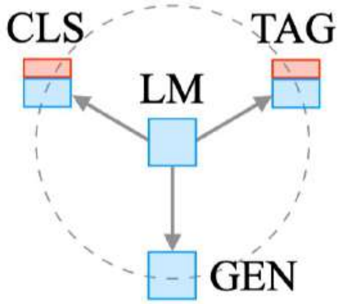


Pre-train, Prompt, Predict
Prompt Engineering

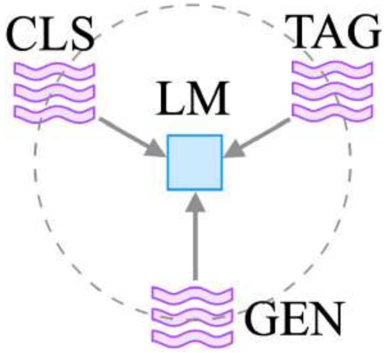


Four Paradigms in NLP

Pre-train, Fine-tune
Objective Engineering
MLM, NSP

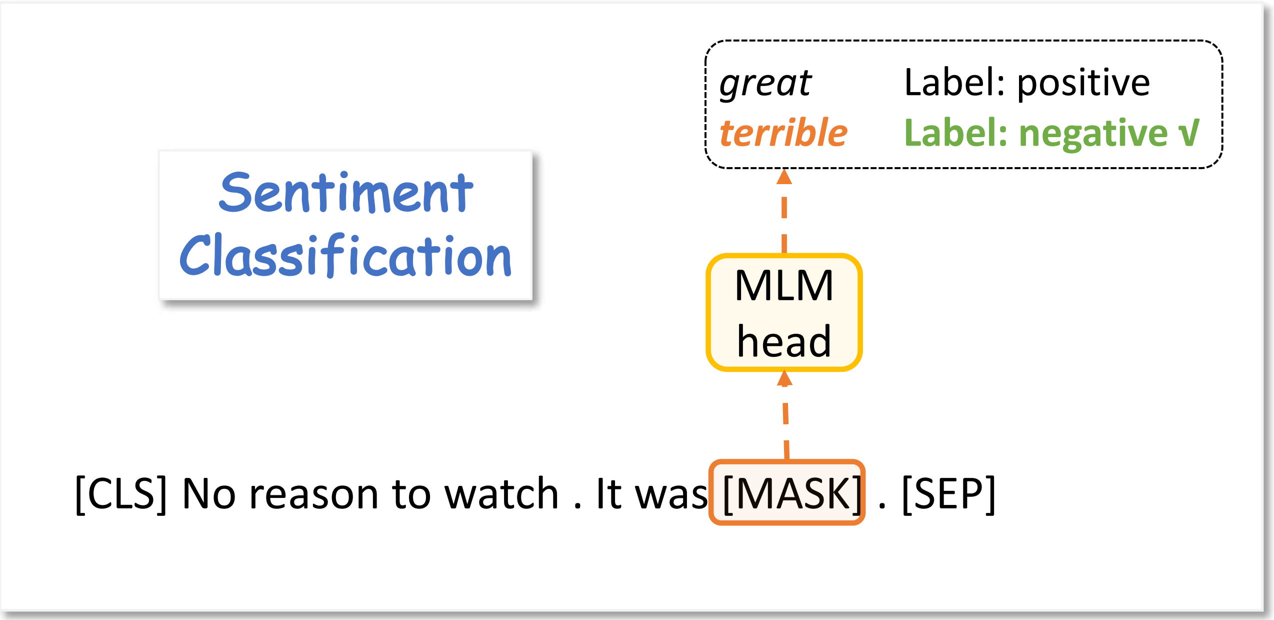


Pre-train, Prompt, Predict
Prompt Engineering



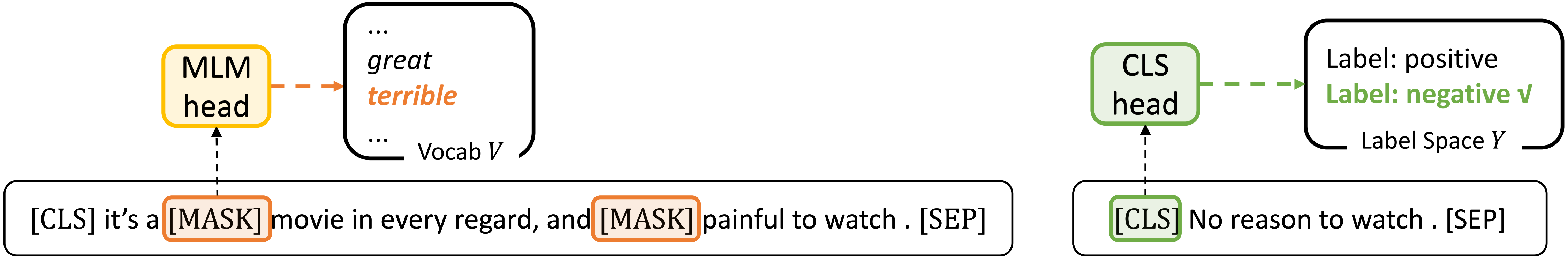
Prompt Learning

Naïve Prompt Learning



Two Gaps

Pre-train and Fine-tune Paradigm

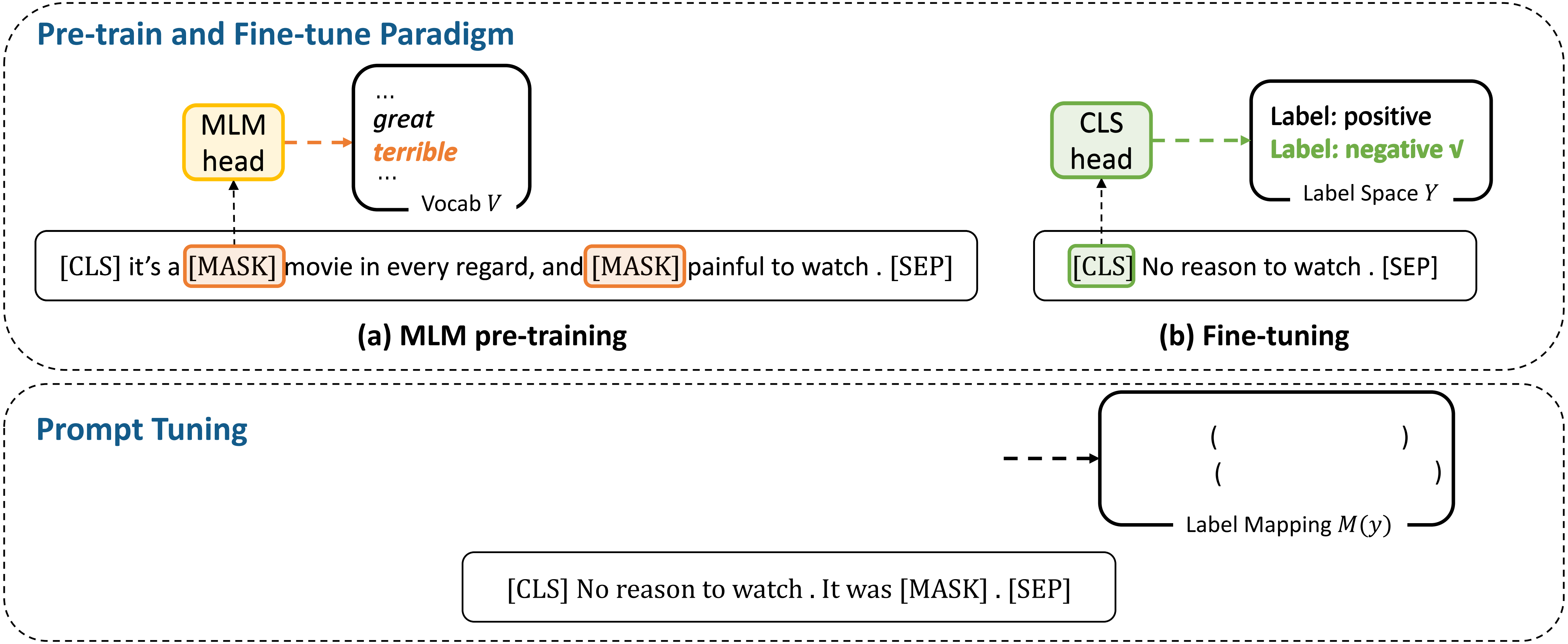


(a) MLM pre-training

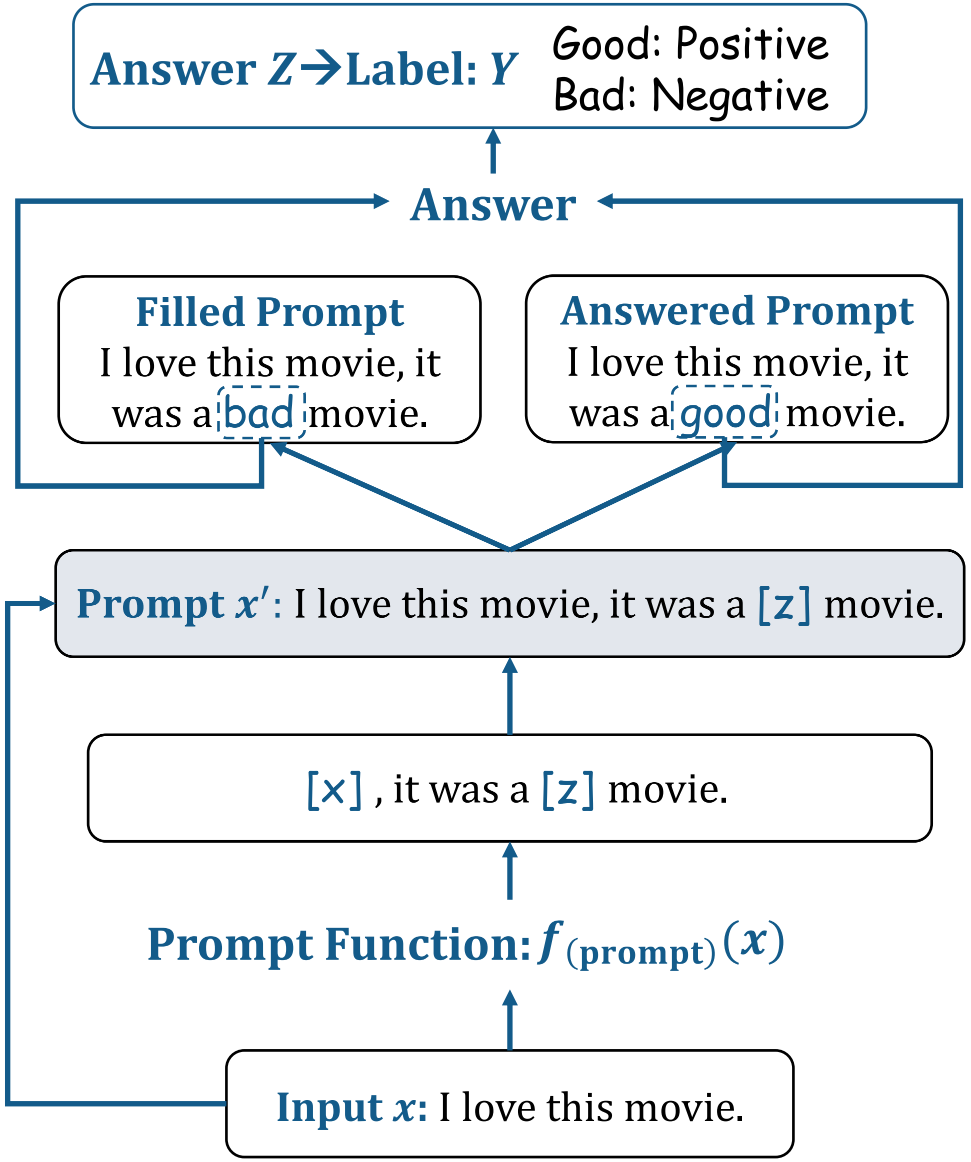
(b) Fine-tuning

Two Gaps

Why Prompt Learning?



Notation of Prompt Learning



Answer Mapping

Answer: *A token, phrase, or sentence that fills [Z]*

Filled Prompt: *A prompt where slot [Z] is filled with any answer.*
Answered Prompt: *A prompt where slot [Z] is filled with a true answer.*

Prompt: *A text where [X] is instantiated by input x but answer slot [Z] is not.*

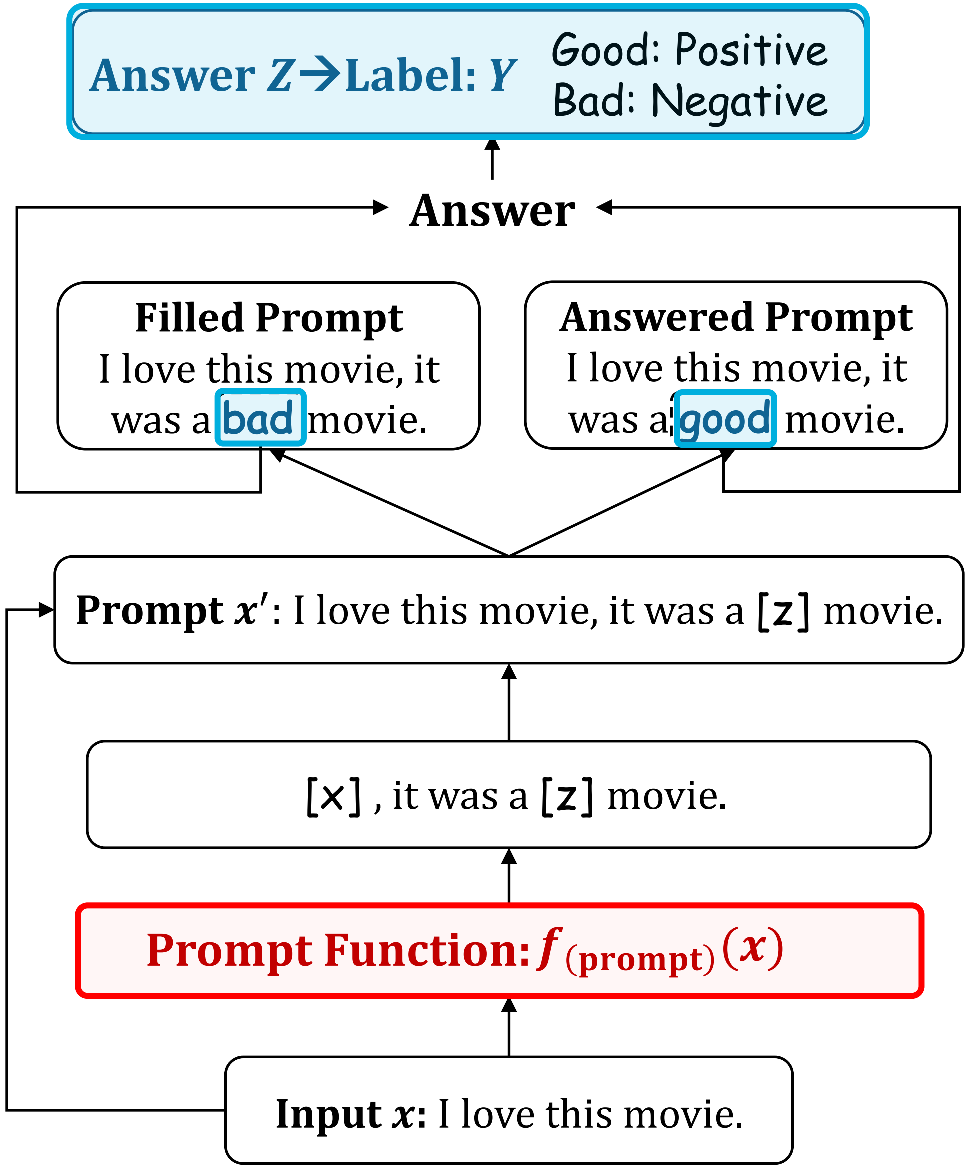
Prompt Function f : *A function that converts the input into a specific form by inserting the input x and adding a slot [Z] where answer z may be filled later.*

Input x: *One or multiple texts*

Examples

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. ...

Two Key Components

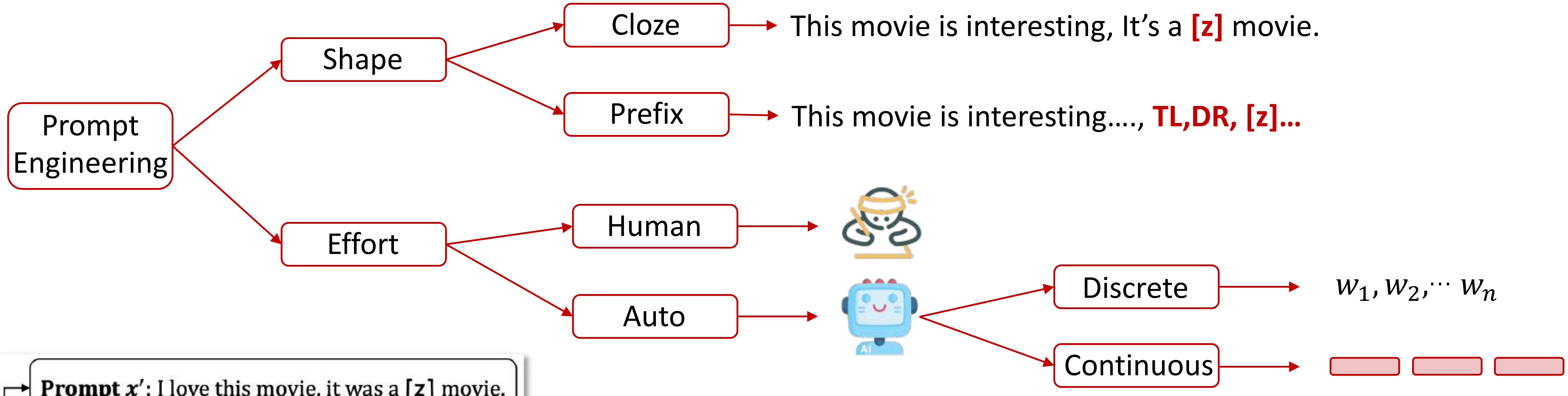


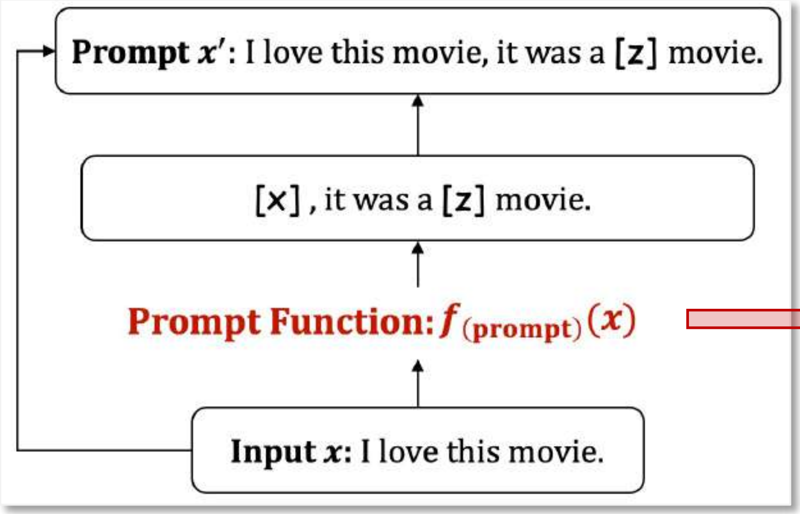
Answer Engineering

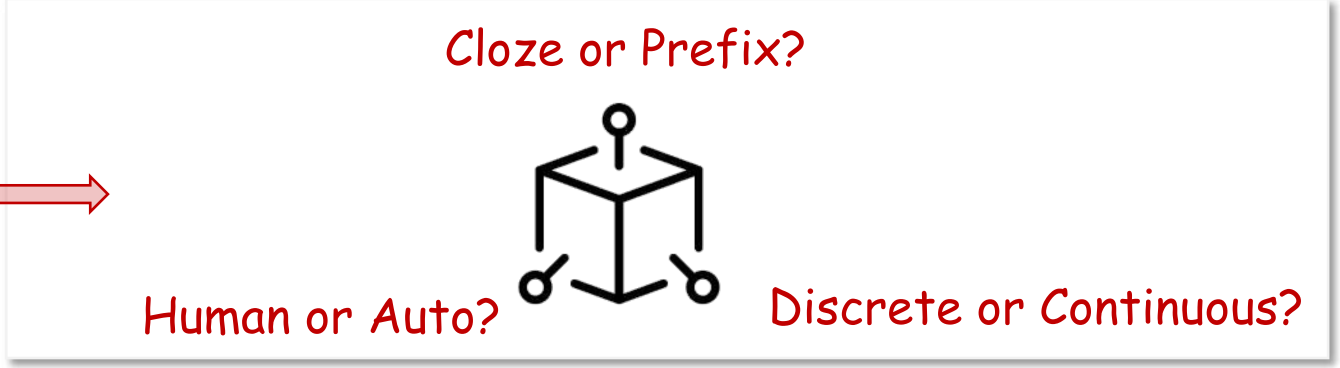
Prompt Engineering

Prompt Engineering

Prompt Engineering







Prompt Engineering

Manually created prompts

Prompt-tuning
Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference

In-context learning
Language Models are Few-Shot Learners (GPT-3)

Automatically created prompts

Mining-based and Paraphrasing-based
How Can We Know What Language Models Know?

Gradient Searching
AUTOPROMPT: Eliciting Knowledge from Language Models with Automatically Generated Prompts

Using Separate Model
Making pre-trained language models better few-shot learners.

GPT Understands, Too

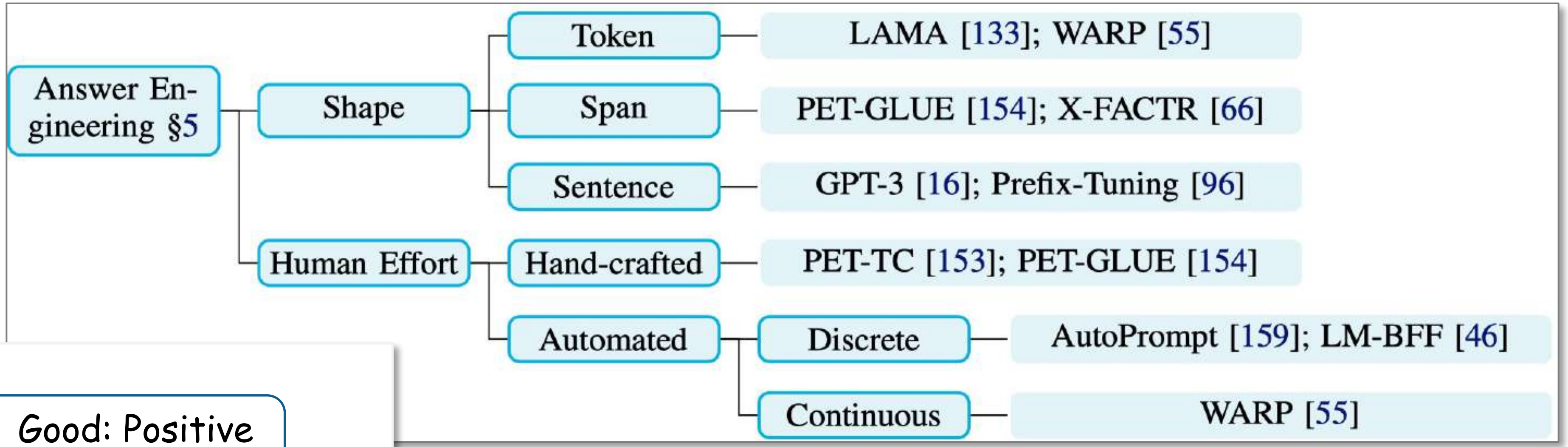
Learning How to Ask: Querying LMs with Mixtures of Soft Prompts

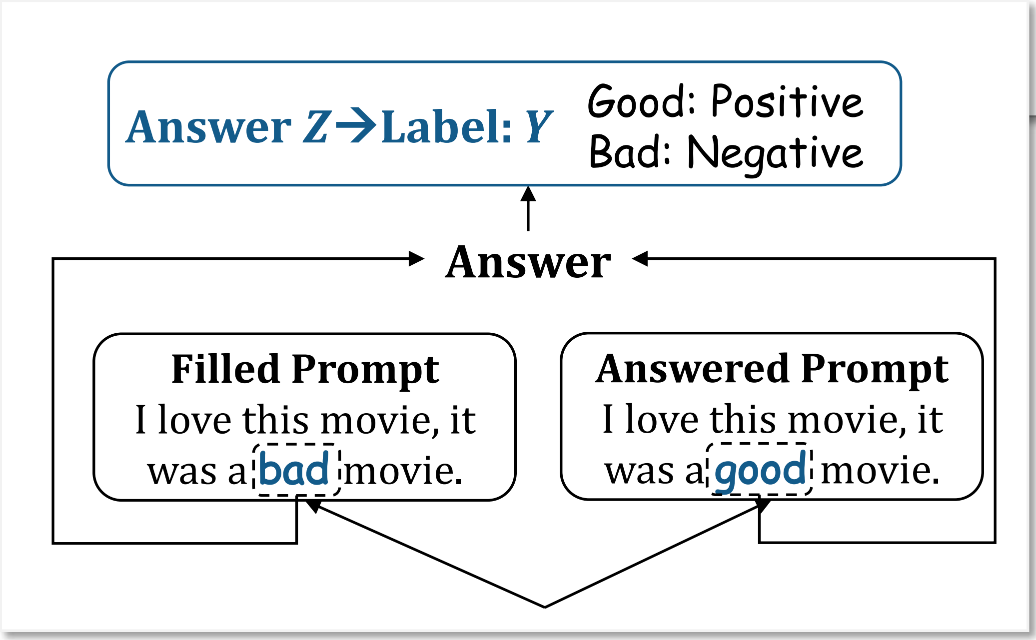
Factual Probing Is [MASK]: Learning vs. Learning to Recall

Differentially Optimized

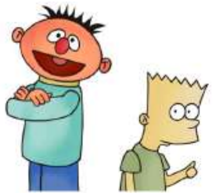
Answer Engineering

Answer Engineering

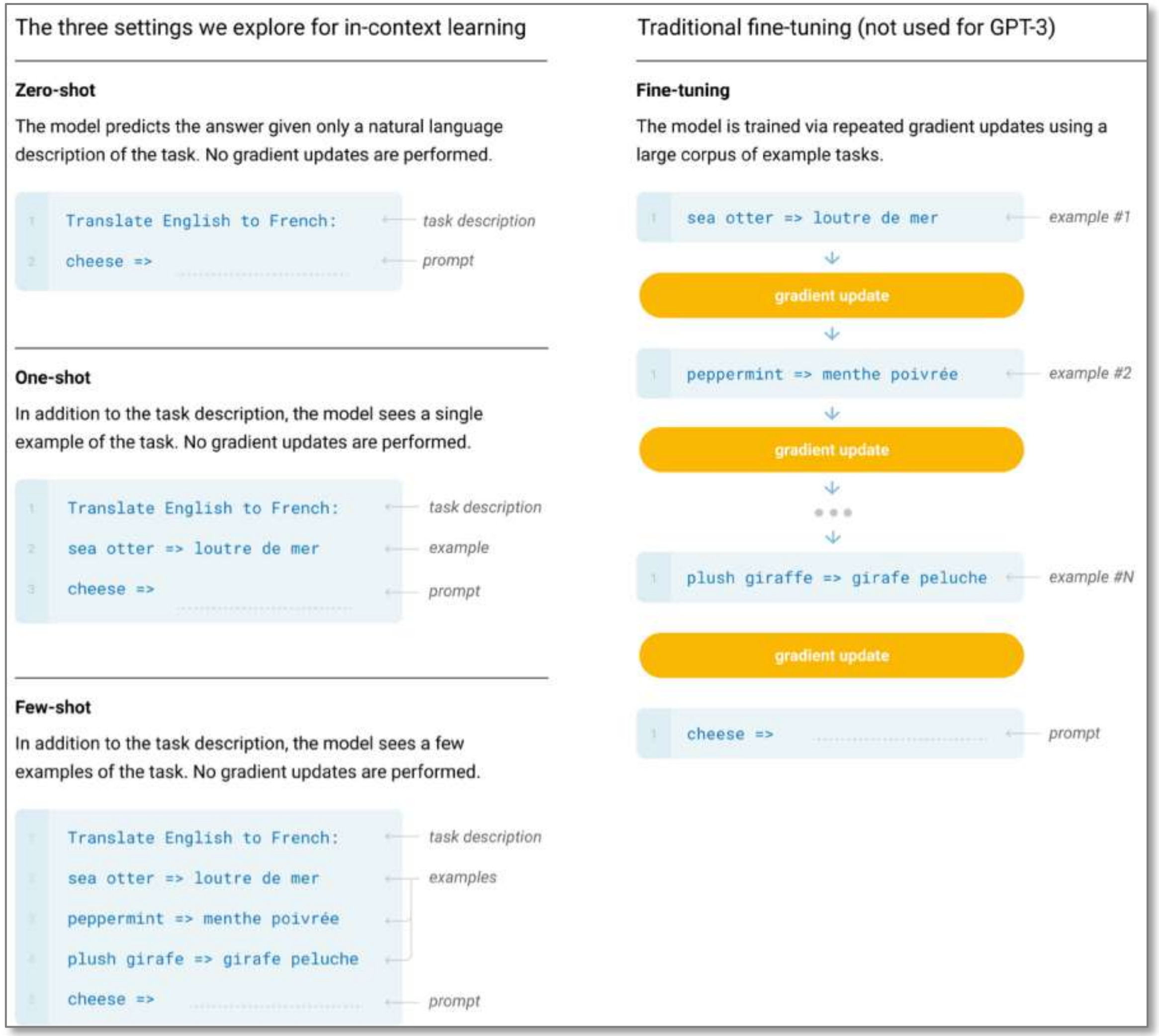




Papers



GPT-3



Prompt Engineering

Human + Discrete

Answer Engineering

Token, Span, Sent

Text Classification

Sentiment Classification

MLM head

great Label: positive

terrible Label: negative √

[CLS] No reason to watch . It was [MASK] . [SEP]

Prompt Engineering

Cloze + Human + Discrete

Answer Engineering

Token + Human + Discrete

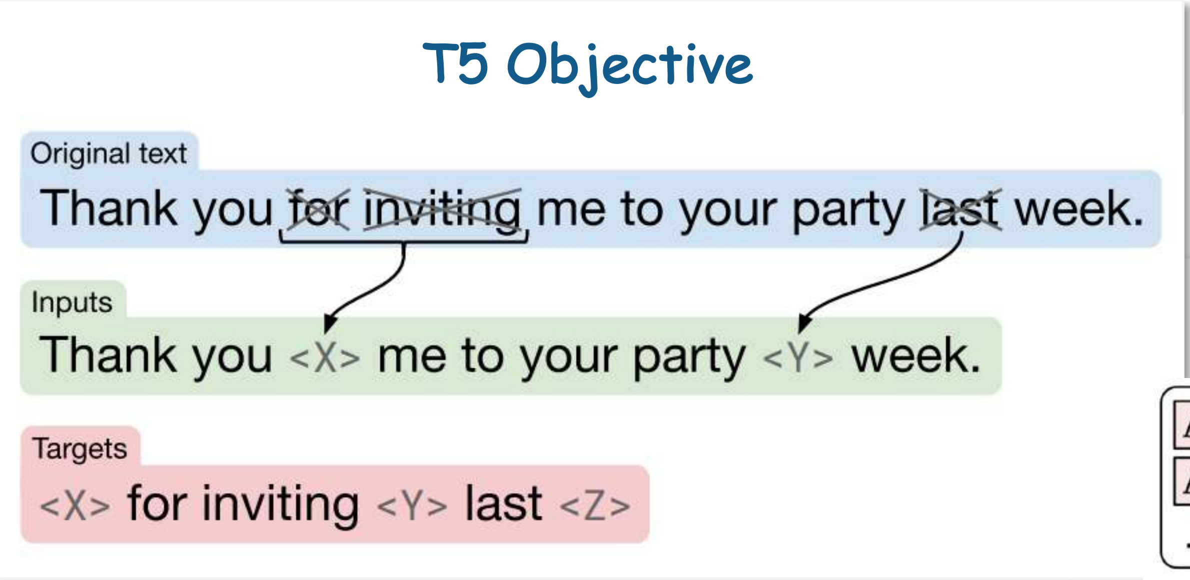
Topic detection — this text is about {"health", "finance", "politics", "sports", etc.}

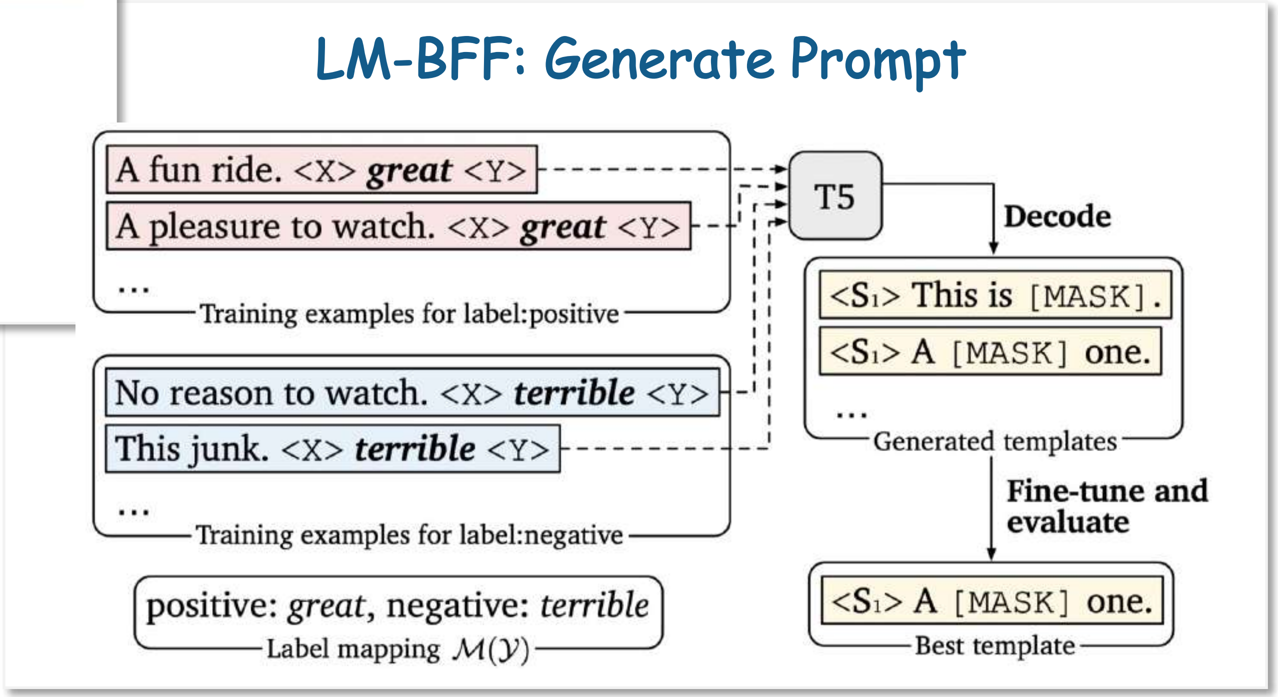
Emotion detection — this text expresses {"anger", "joy", "sadness", "fear", etc.}

Situation detection — the people there need {"shelter", "water", · · · }

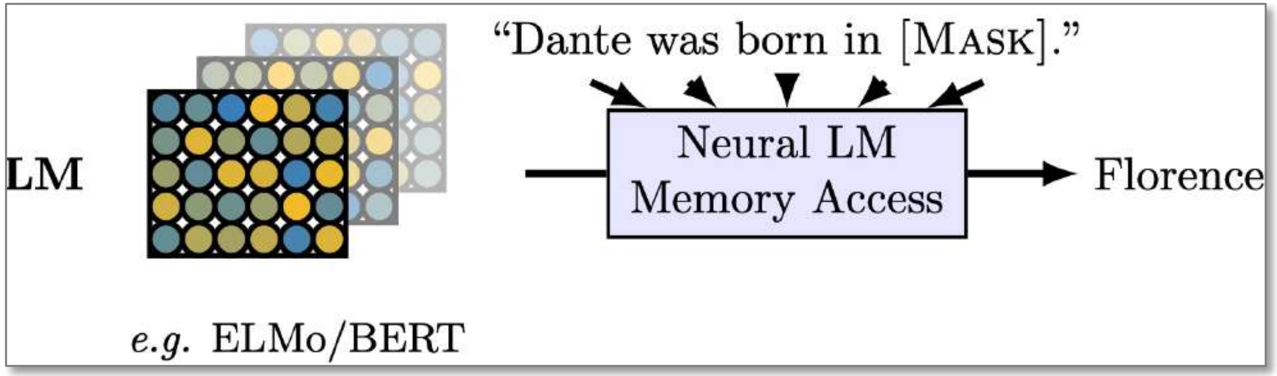
Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach, EMNLP 2019

Text Classification: LM-BFF

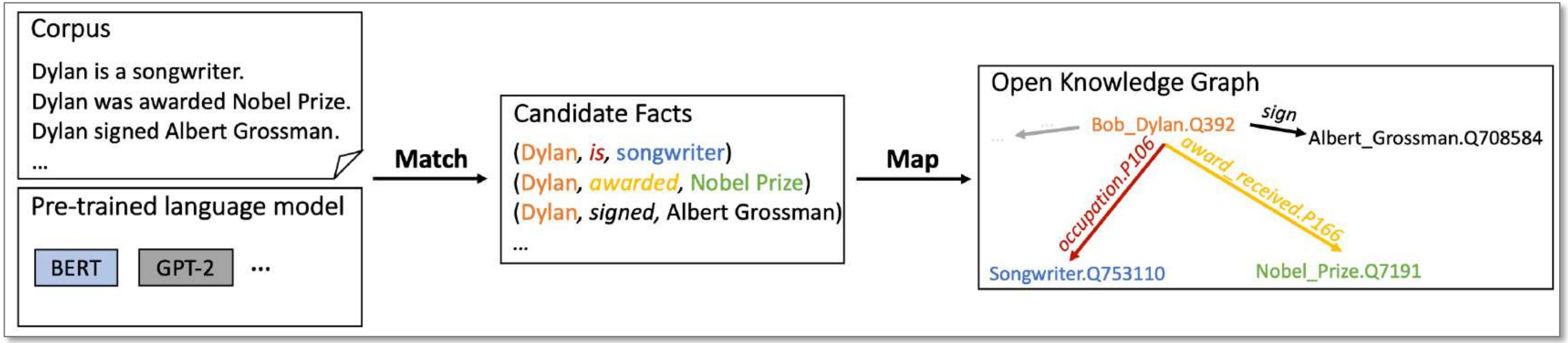




Knowledge Mining

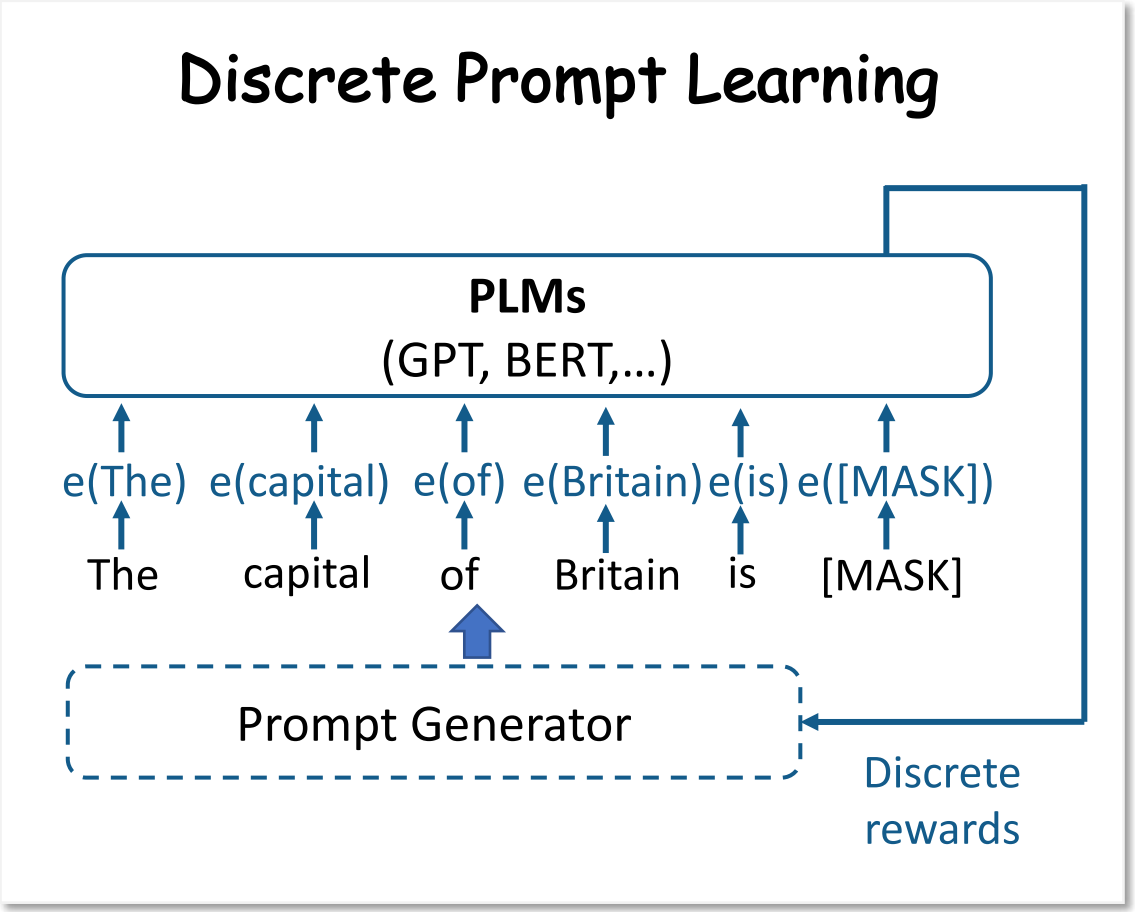


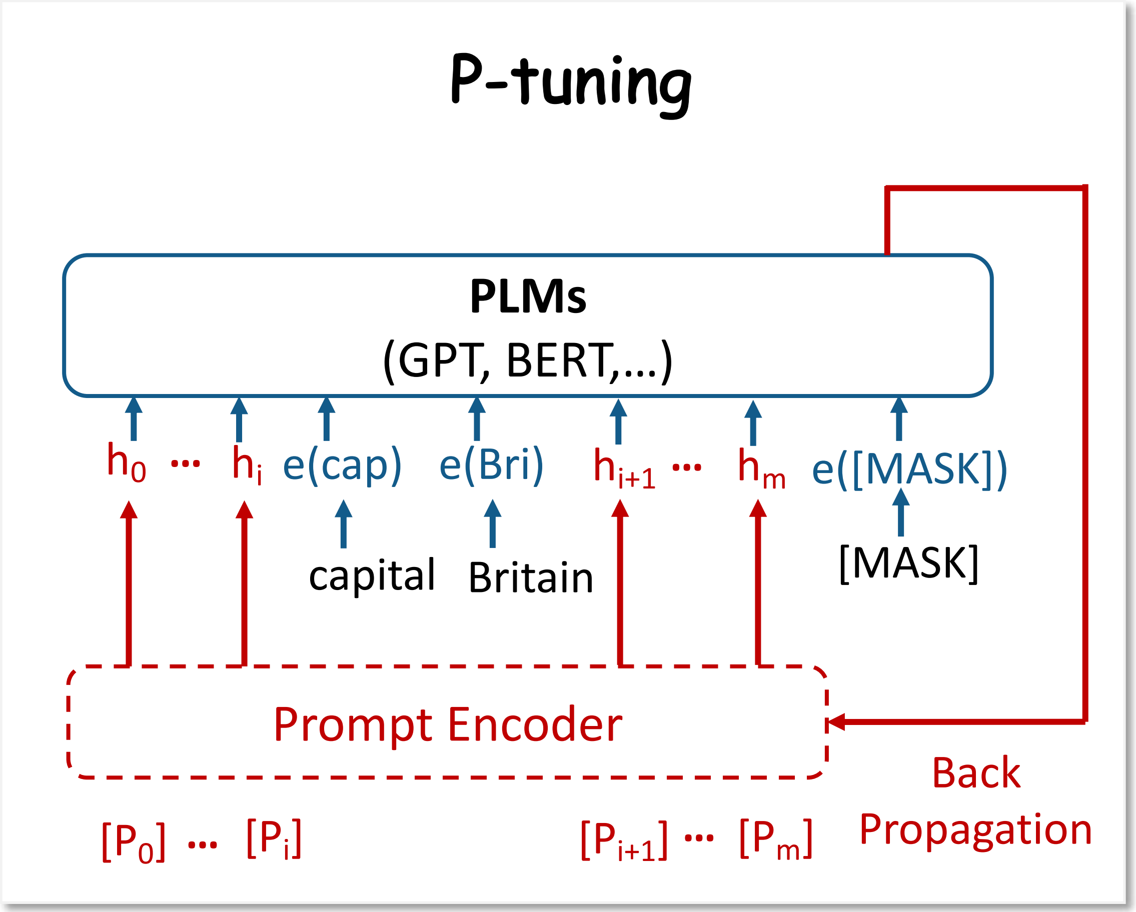
Query	Answer
Francesco Bartolomeo Conti was born in ____.	Florence
Adolphe Adam died in ____.	Paris
English bulldog is a subclass of ____.	dog
The official language of Mauritius is ____.	English
Patrick Oboya plays in ____ position.	midfielder
Hamburg Airport is named after ____.	Hamburg
The original language of Mon oncle Benjamin is ____.	French
Dani Alves plays with ____ .	Barcelona
Paul Toungui is a ____ by profession .	politician
Sodium sulfide consists of ____ .	sodium



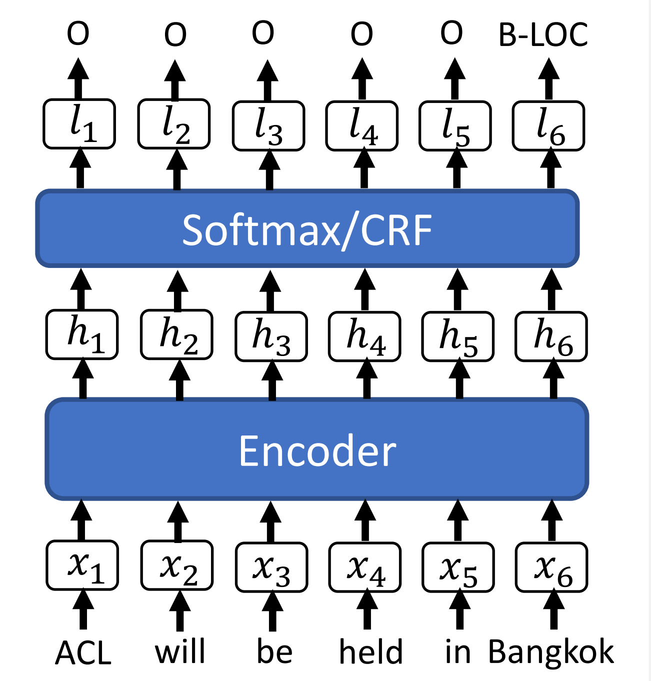
Language Models as Knowledge Bases?
Language Models are Open Knowledge Graphs

Knowledge Mining: P-Tuning





NER



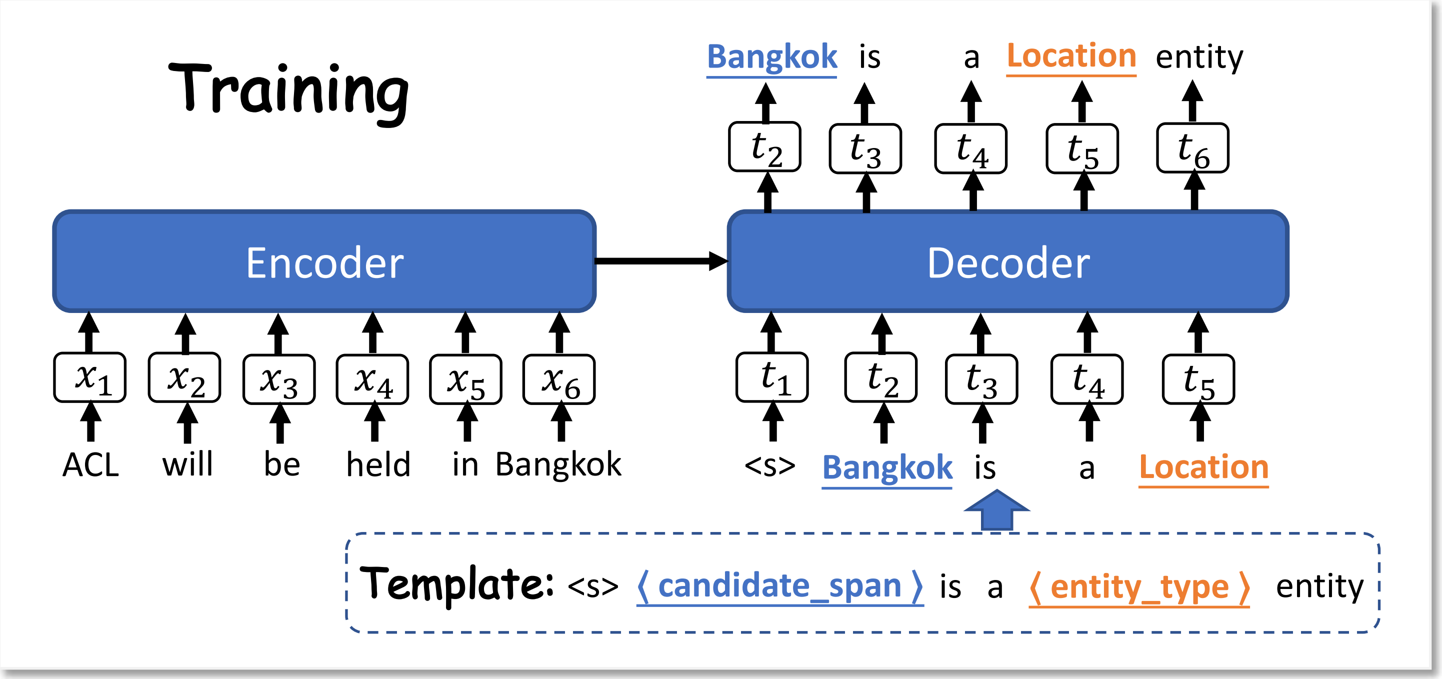
Sequence Labeling

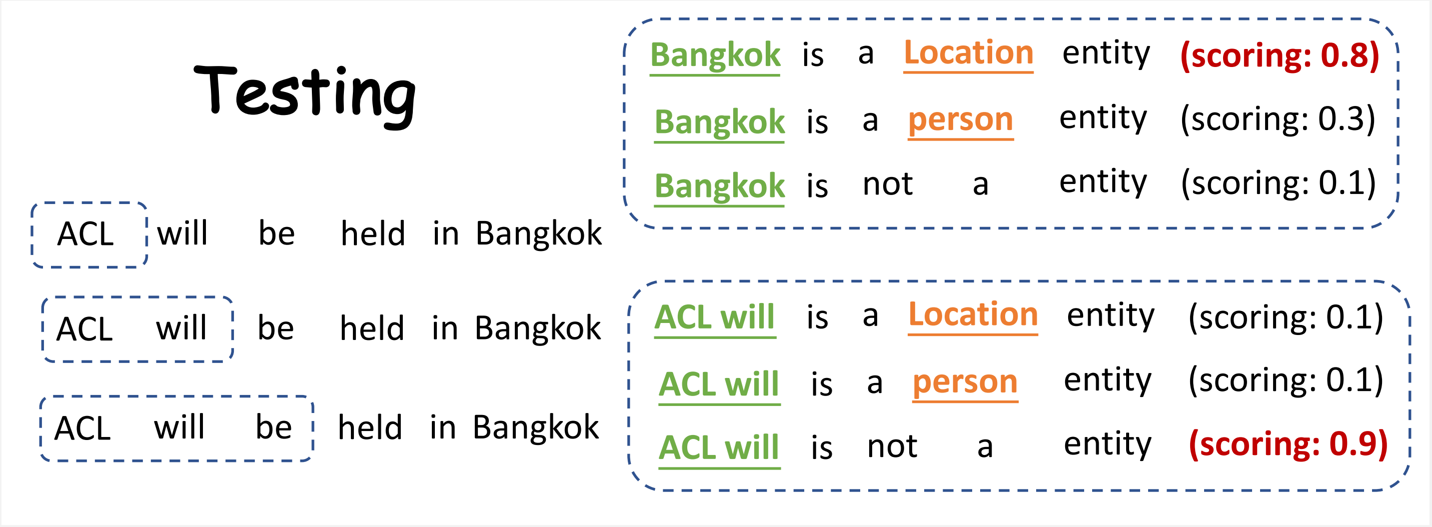
Prompt Engineering

Cloze + Human + Discrete

Answer Engineering

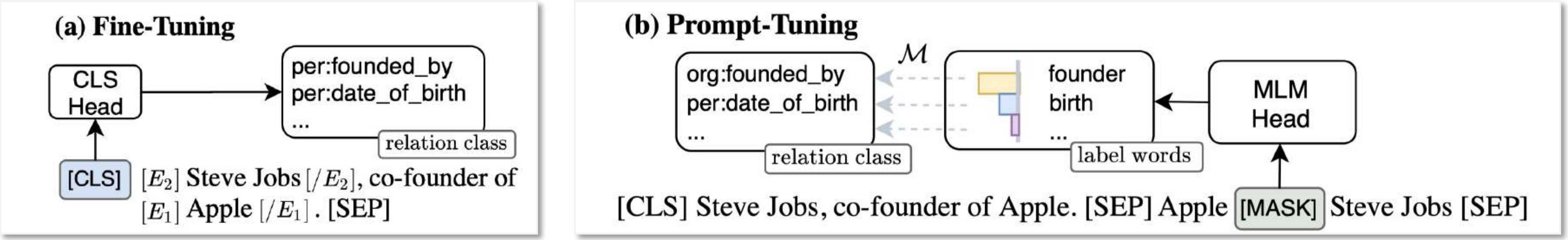
Token + Human + Discrete

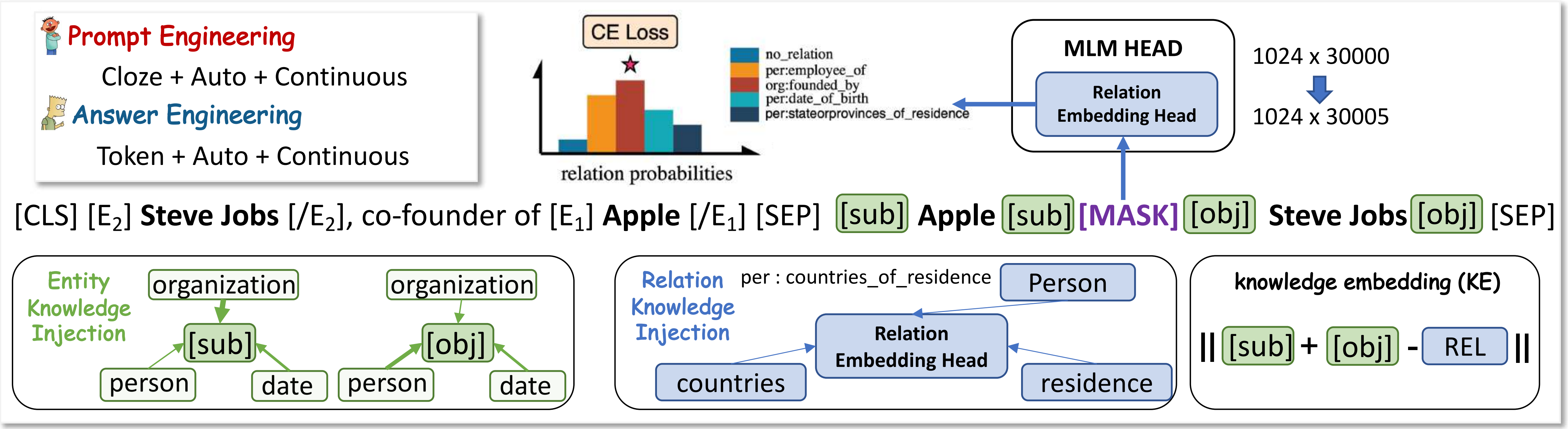




Template-Based Named Entity Recognition Using BART, Findings of ACL 2021

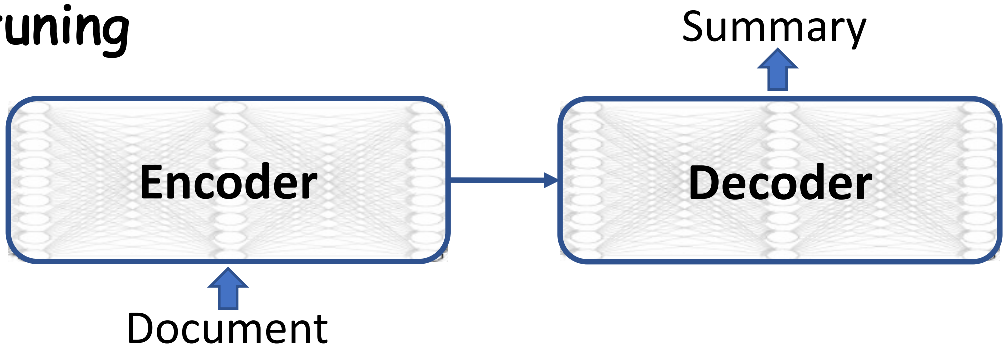
Relation Extraction





Text Generation

Fine-tuning



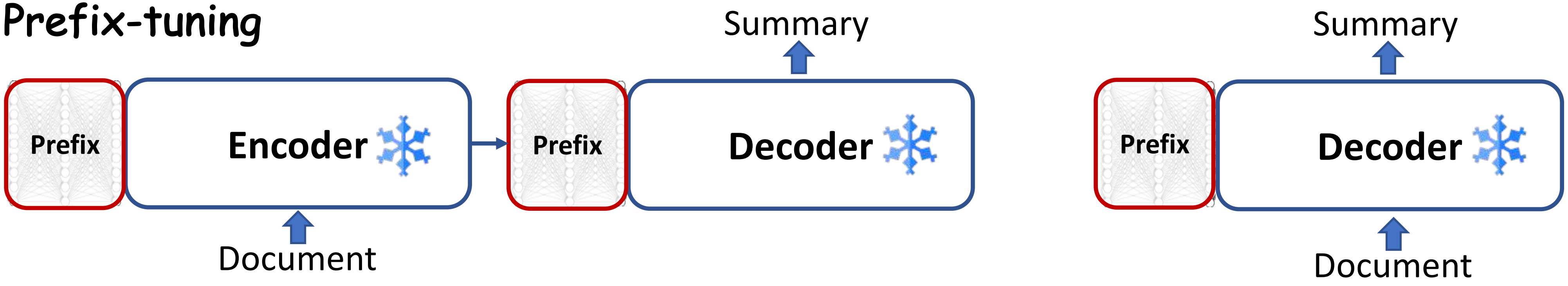
Prompt Engineering

Prefix+ Auto + Continuous

Answer Engineering

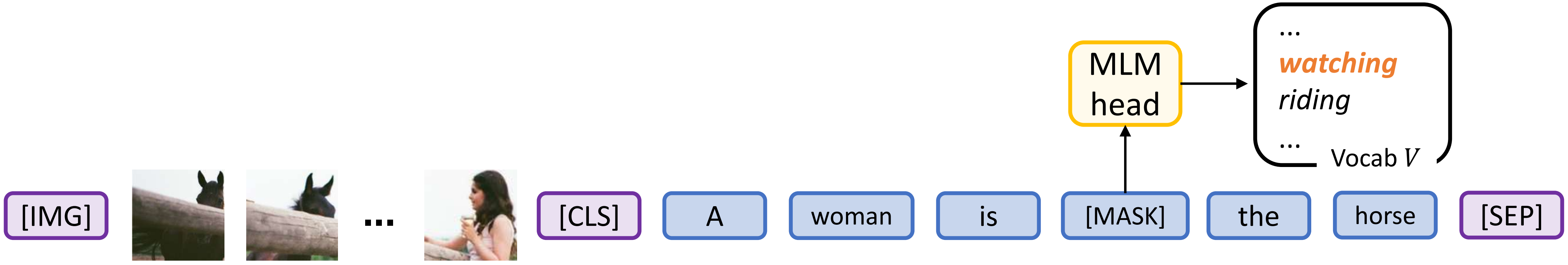
Sent + Discrete

Prefix-tuning

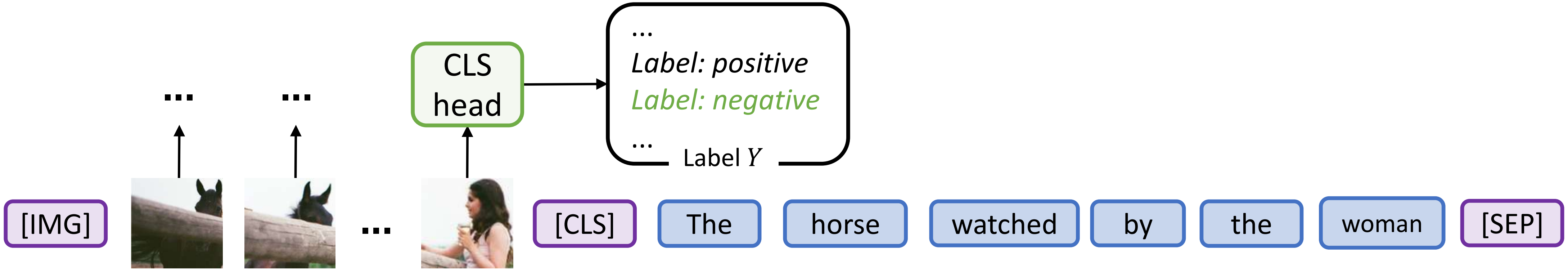


Prefix-Tuning: Optimizing Continuous Prompts for Generation, ACL 2021

Cross-modal

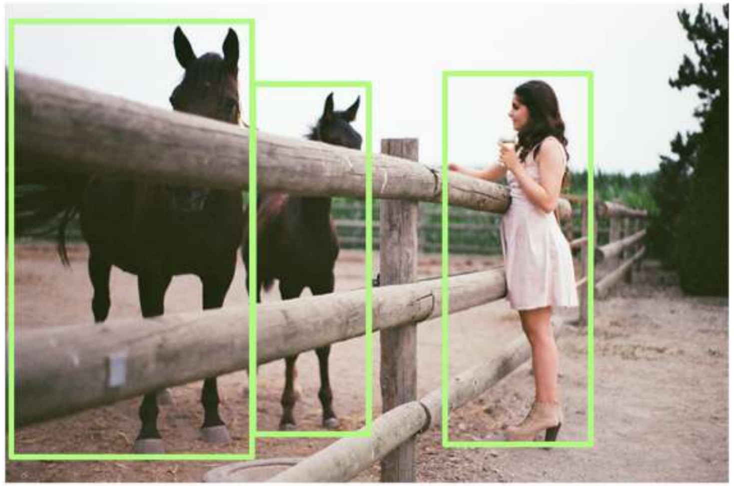


(a) pre-training



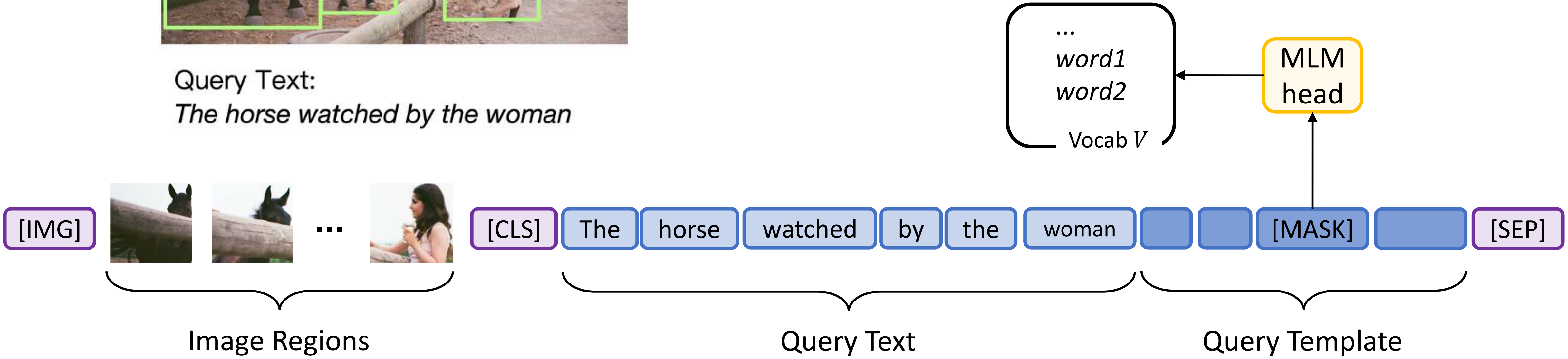
(b) fine-tuning

Cross-modal



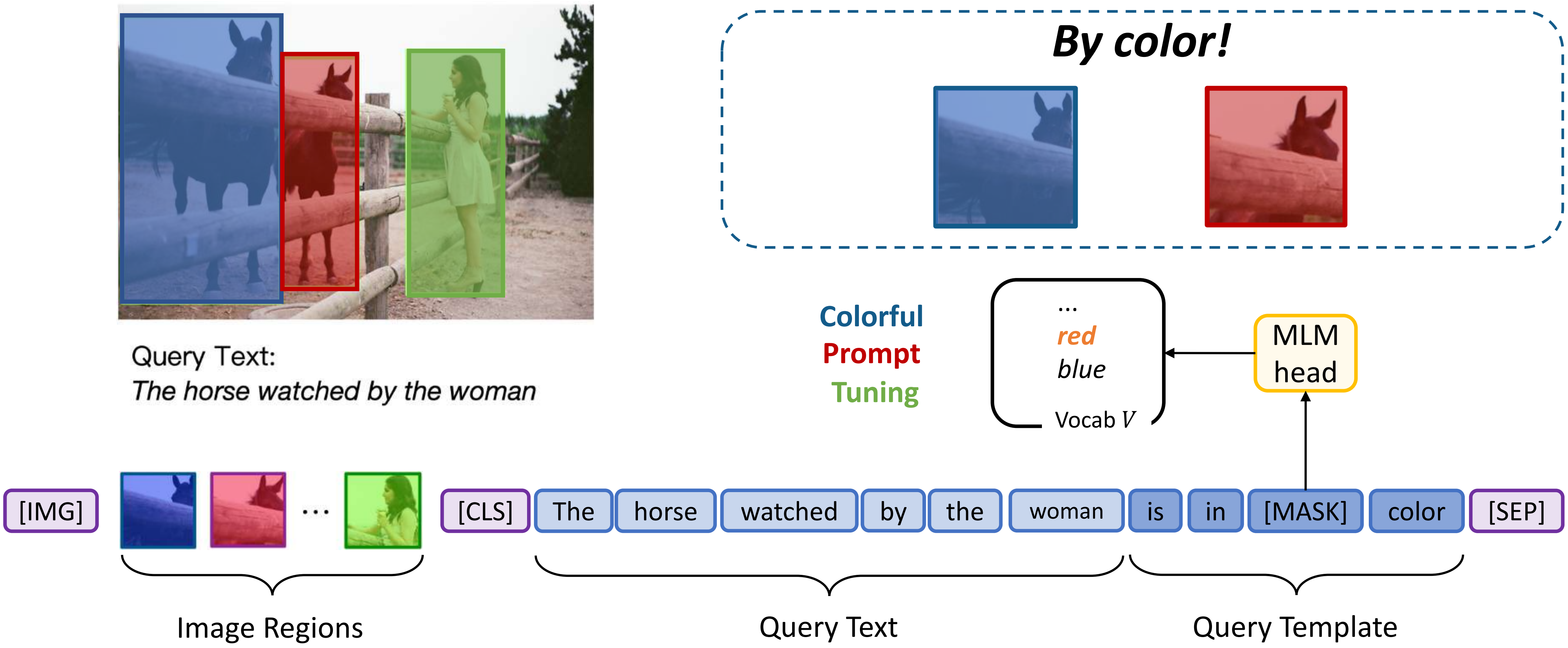
Query Text:
The horse watched by the woman

How to distinguish these two horses in the vocab?



(c) Cross-modal Prompt Tuning

Cross-modal



By color!
Query Text:
The horse watched by the woman
Colorful
Prompt
Tuning
...
red
blue
Vocab V
MLM head
[IMG]
...
[CLS]
The
horse
watched
by
the
woman
is
in
[MASK]
color
[SEP]
Image Regions
Query Text
Query Template

Part Conclusion

Prompt

PLMs

Downstream Tasks

Thanks~