

# NLP 201: Large Language Models

Jeffrey Flanigan

University of California Santa Cruz  
[jmflanig@ucsc.edu](mailto:jmflanig@ucsc.edu)

Slides adapted from Mohit Iyyer

Fall 2023

# Plan

---

- Prompting and In-Context Learning
- Retrieval-Augmented Methods
- Instruction Tuning

# The language model “scaling wars”!

ELMo: 93M params, 2-layer biLSTM

BERT-base: 110M params, 12-layer Transformer

BERT-large: 340M params, 24-layer Transformer

Model Name	$n_{\text{params}}$	$n_{\text{layers}}$	$d_{\text{model}}$	$n_{\text{heads}}$	$d_{\text{head}}$	Batch Size	Learning Rate
GPT-3 Small	125M	12	768	12	64	0.5M	$6.0 \times 10^{-4}$
GPT-3 Medium	350M	24	1024	16	64	0.5M	$3.0 \times 10^{-4}$
GPT-3 Large	760M	24	1536	16	96	0.5M	$2.5 \times 10^{-4}$
GPT-3 XL	1.3B	24	2048	24	128	1M	$2.0 \times 10^{-4}$
GPT-3 2.7B	2.7B	32	2560	32	80	1M	$1.6 \times 10^{-4}$
GPT-3 6.7B	6.7B	32	4096	32	128	2M	$1.2 \times 10^{-4}$
GPT-3 13B	13.0B	40	5140	40	128	2M	$1.0 \times 10^{-4}$
GPT-3 175B or “GPT-3”	175.0B	96	12288	96	128	3.2M	$0.6 \times 10^{-4}$

# The language model “scaling wars”!

ELMo: 93M params, 2-layer biLSTM

BERT-base: 110M params, 12-layer Transformer

BERT-large: 340M params, 24-layer Transformer

Model Name	$n_{\text{params}}$	$n_{\text{layers}}$	$d_{\text{model}}$	$n_{\text{heads}}$	$d_{\text{head}}$	Batch Size	Learning Rate
GPT-3 Small	125M	12	768	12	64	0.5M	$6.0 \times 10^{-4}$
GPT-3 Medium	350M	24	1024	16	64	0.5M	$3.0 \times 10^{-4}$
GPT-3 Large	760M	24	1536	16	96	0.5M	$2.5 \times 10^{-4}$
GPT-3 XL	1.3B	24	2048	24	128	1M	$2.0 \times 10^{-4}$
GPT-3 2.7B	2.7B	32	2560	32	80	1M	$1.6 \times 10^{-4}$
GPT-3 6.7B	6.7B	32	4096	32	128	2M	$1.2 \times 10^{-4}$
GPT-3 13B	13.0B	40	5140	40	128	2M	$1.0 \times 10^{-4}$
GPT-3 175B or “GPT-3”	175.0B	96	12288	96	128	3.2M	$0.6 \times 10^{-4}$

# The language model “scaling wars”!

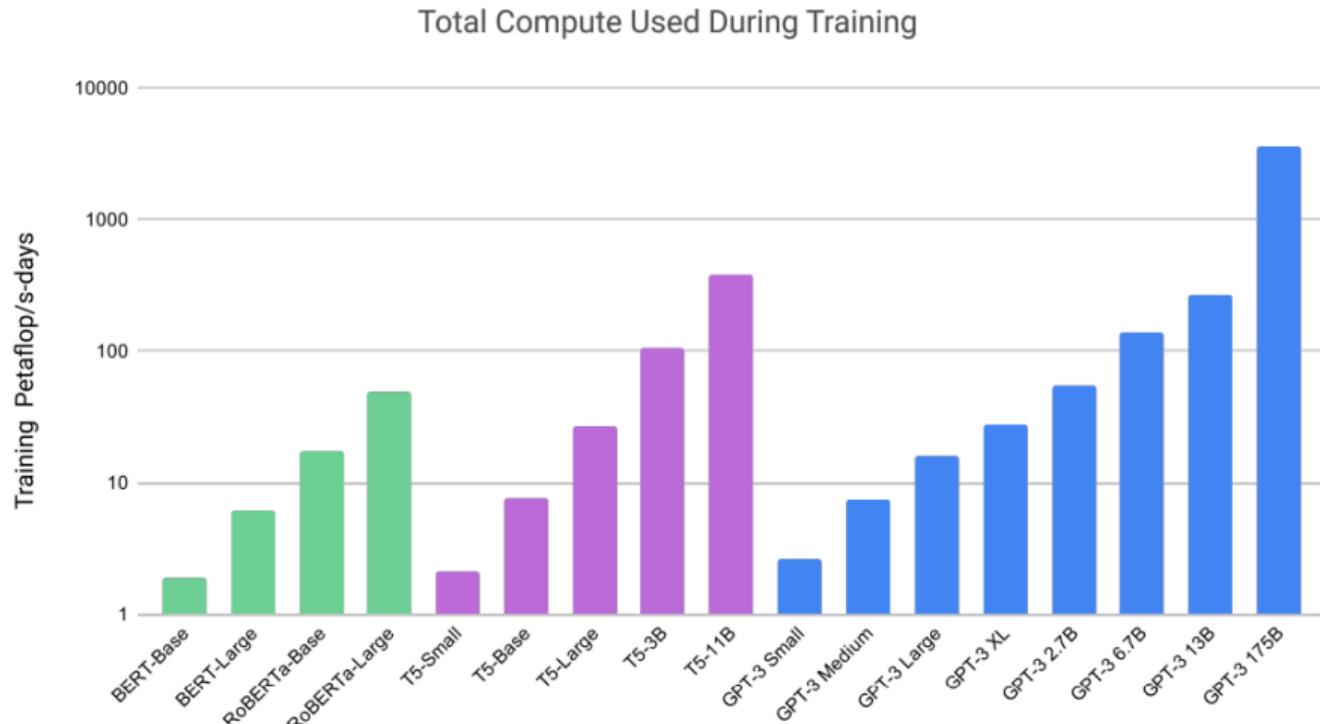
ELMo: 1B training tokens

BERT: 3.3B training tokens

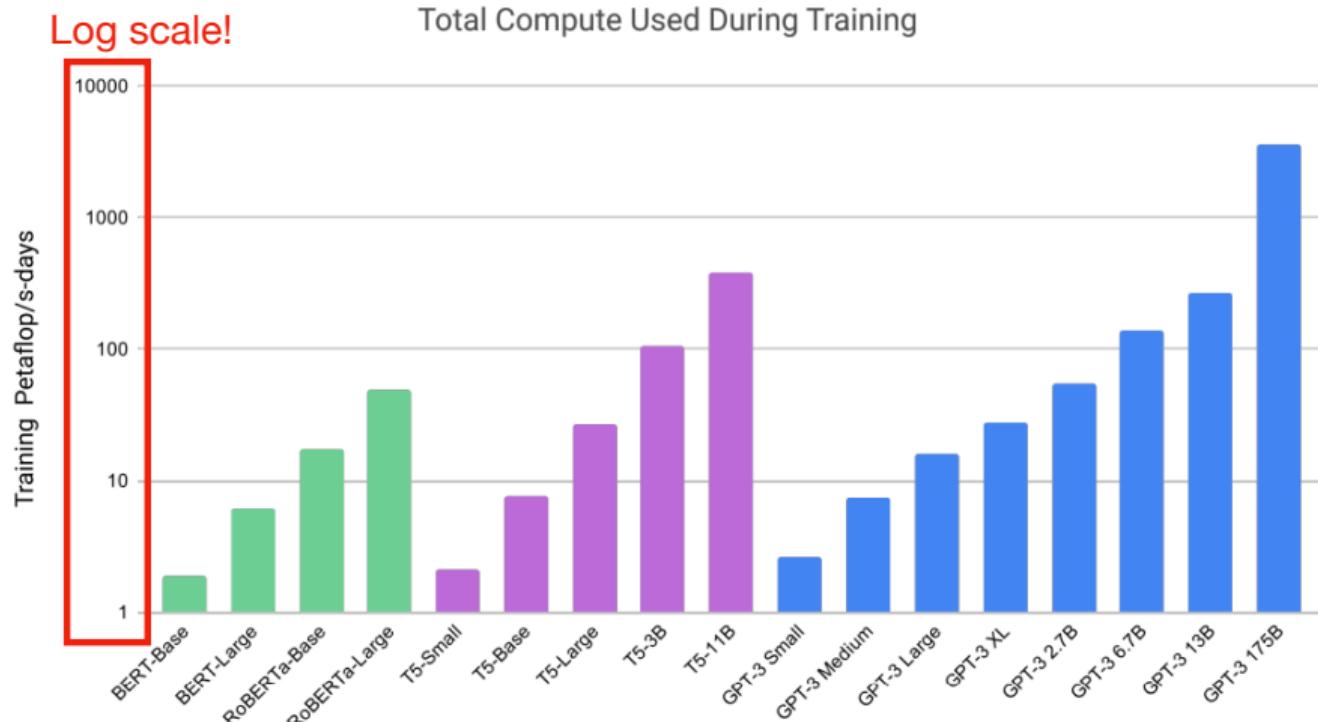
RoBERTa: ~30B training tokens

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

# The language model “scaling wars”!



# The language model “scaling wars”!



# 2022-2023:

**PaLM (Google)**: 540B params, 118 layers, 18432 d\_model, 780 billion training tokens Model not available

**ChatGPT (OpenAI)**: Params, layers, dimensionality, training data size unknown Model available only through blackbox API

**LLaMa (Meta)**: 65B params, 80 layers, 8192 d\_model, 1.4 trillion tokens of training data Model parameters publicly available!

**GPT4 (OpenAI)**: Params, layers, dimensionality, training data size unknown Model available only through blackbox API

**Bard (Google)**: Params, layers, dimensionality, training data size unknown Model available only through blackbox API

so... what does all of this scaling buy us?

## Traditional fine-tuning (not used for GPT-3)

### Fine-tuning

The model is trained via repeated gradient updates using a large corpus of example tasks.

Downstream  
training data

1 sea otter => loutre de mer ← example #1



gradient update



1 peppermint => menthe poivrée ← example #2



gradient update



• • •



1 plush giraffe => girafe peluche ← example #N

gradient update

1 cheese => ..... ← prompt

Downstream  
test data

## Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.



### Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.



**No fine-tuning!!! Literally just take a pretrained LM and give it the following prefix:**

**“Translate English to French: cheese =>”**

### One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.

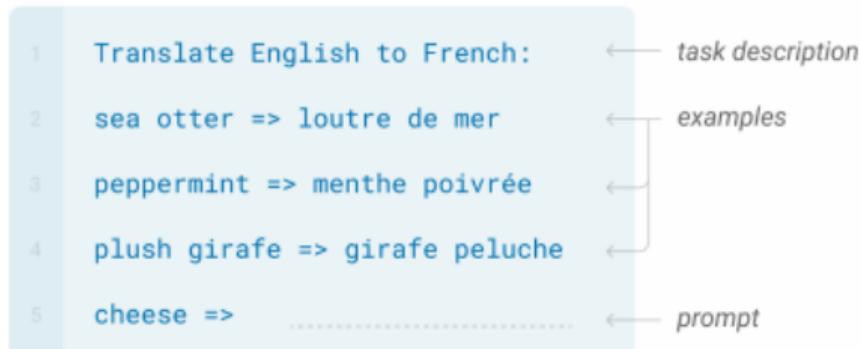


**No fine-tuning!!! Literally just take a pretrained LM and give it the following prefix:**

**“Translate English to French: sea otter => loutre de mer, cheese =>”**

## Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



**No fine-tuning!!! Literally just take a pretrained LM and give it the following prefix:**

**“Translate English to French: sea otter => loutre de mer, peppermint => ... (few more examples), cheese =>”**

**Max of 100 examples fed into the prefix in this way**

How does this new paradigm  
compare to “pretrain + finetune”?

# TriviaQA

---

## Question

---

Miami Beach in Florida borders which ocean?

---

What was the occupation of Lovely Rita according to the song by the Beatles

---

Who was Poopdeck Pappys most famous son?

---

The Nazi regime was Germany's Third Reich; which was the first Reich?

---

At which English racecourse did two horses collapse and die in the parade ring due to electrocution, in February 2011?

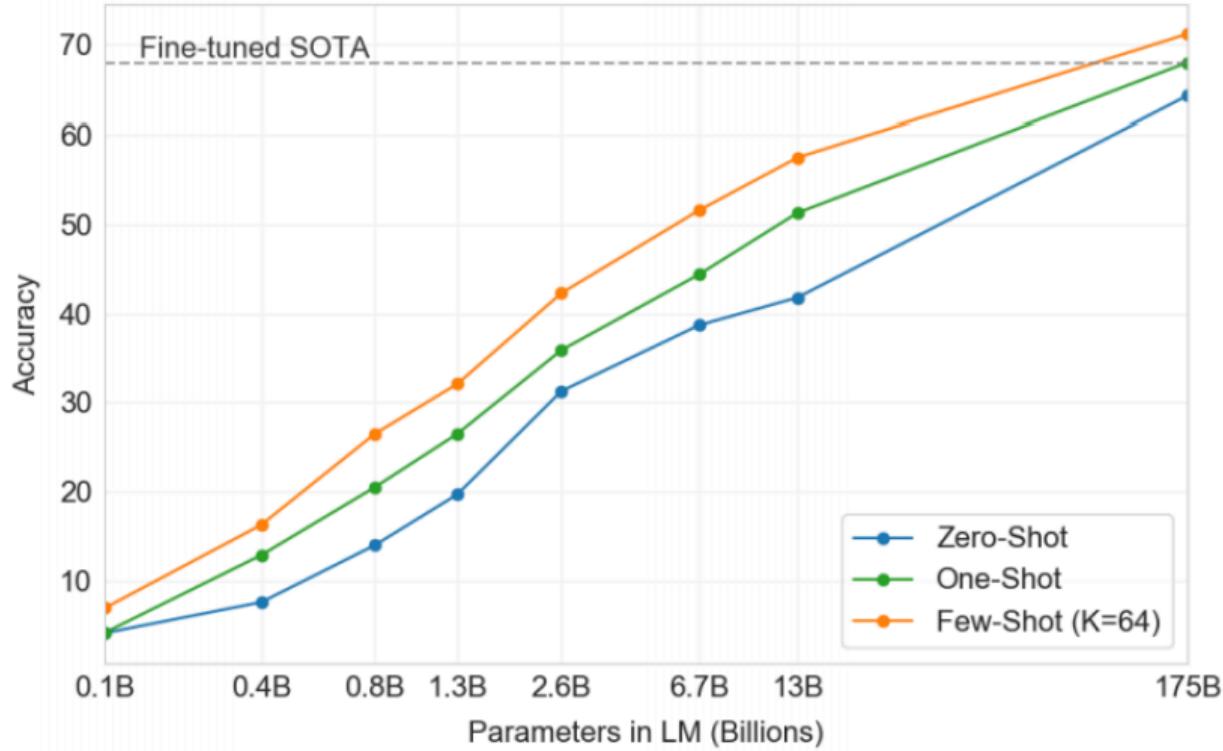
---

Which type of hat takes its name from an 1894 novel by George Du Maurier where the title character has the surname O'Ferrall ?

---

What was the Elephant Man's real name?

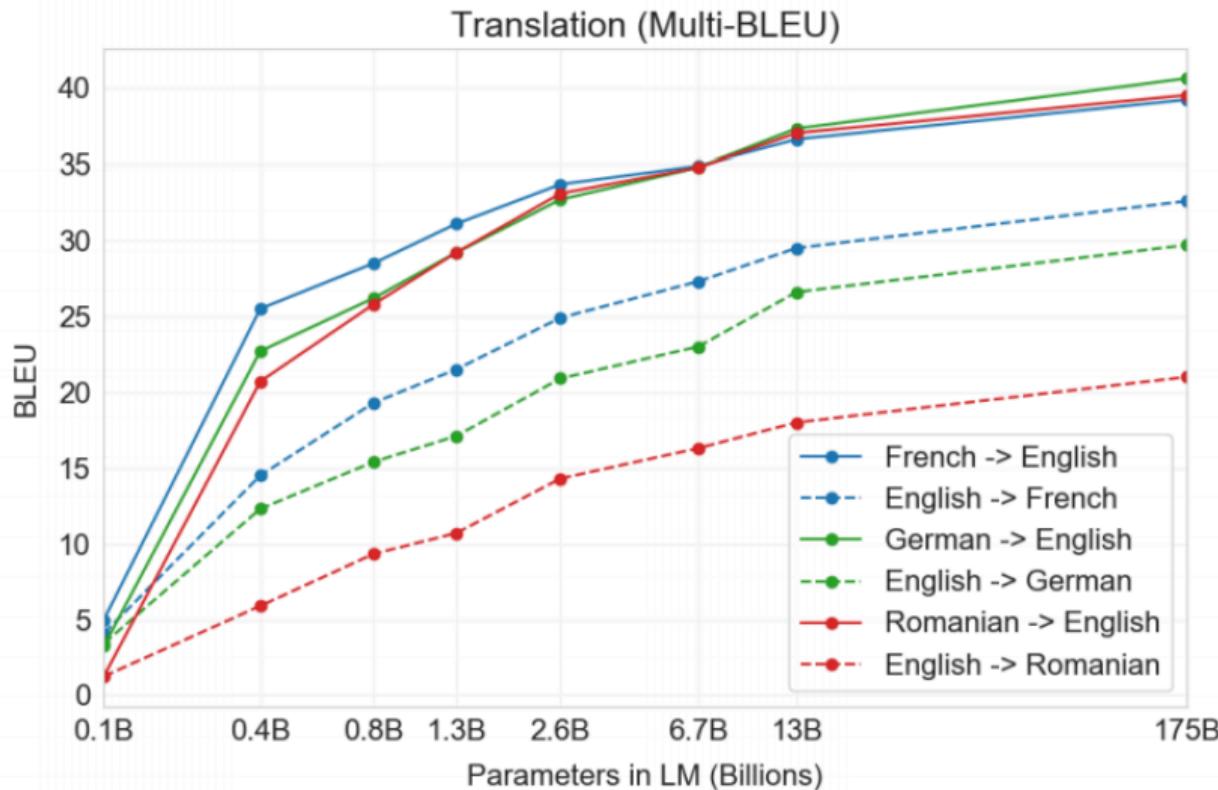
## TriviaQA



What about translation? (7% of  
GPT3's training data is in  
languages other than English)

Setting	En→Fr	Fr→En	En→De	De→En	En→Ro	Ro→En
SOTA (Supervised)	<b>45.6<sup>a</sup></b>	35.0 <sup>b</sup>	<b>41.2<sup>c</sup></b>	40.2 <sup>d</sup>	<b>38.5<sup>e</sup></b>	<b>39.9<sup>e</sup></b>
XLM [LC19]	33.4	33.3	26.4	34.3	33.3	31.8
MASS [STQ <sup>+</sup> 19]	<u>37.5</u>	34.9	28.3	35.2	<u>35.2</u>	33.1
mBART [LGG <sup>+</sup> 20]	-	-	<u>29.8</u>	34.0	35.0	30.5
GPT-3 Zero-Shot	25.2	21.2	24.6	27.2	14.1	19.9
GPT-3 One-Shot	28.3	33.7	26.2	30.4	20.6	38.6
GPT-3 Few-Shot	32.6	<u>39.2</u>	29.7	<u>40.6</u>	21.0	<u>39.5</u>

Improvements haven't plateaued!

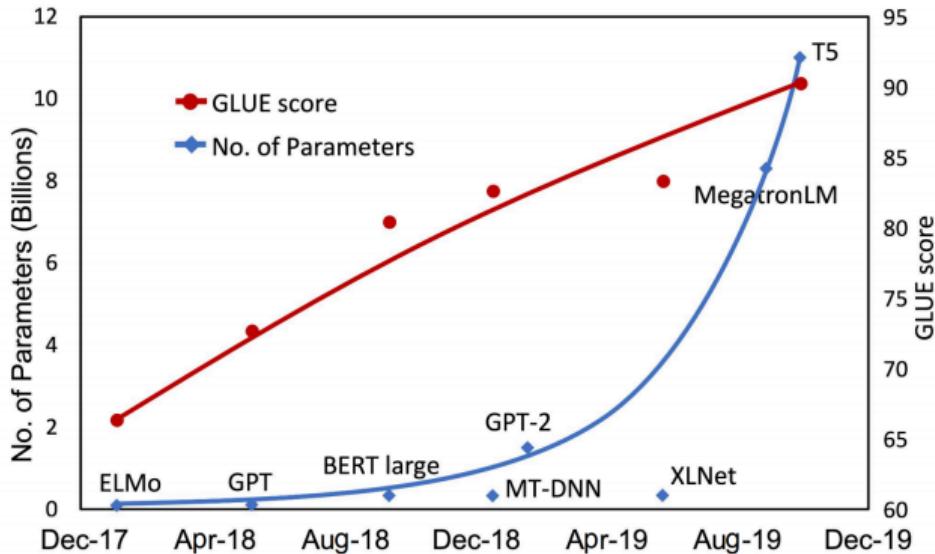


What about reading  
comprehension QA?

Setting	CoQA	DROP	QuAC	SQuADv2	RACE-h	RACE-m
Fine-tuned SOTA	<b>90.7<sup>a</sup></b>	<b>89.1<sup>b</sup></b>	<b>74.4<sup>c</sup></b>	<b>93.0<sup>d</sup></b>	<b>90.0<sup>e</sup></b>	<b>93.1<sup>e</sup></b>
GPT-3 Zero-Shot	81.5	23.6	41.5	59.5	45.5	58.4
GPT-3 One-Shot	84.0	34.3	43.3	65.4	45.9	57.4
GPT-3 Few-Shot	85.0	36.5	44.3	69.8	46.8	58.1

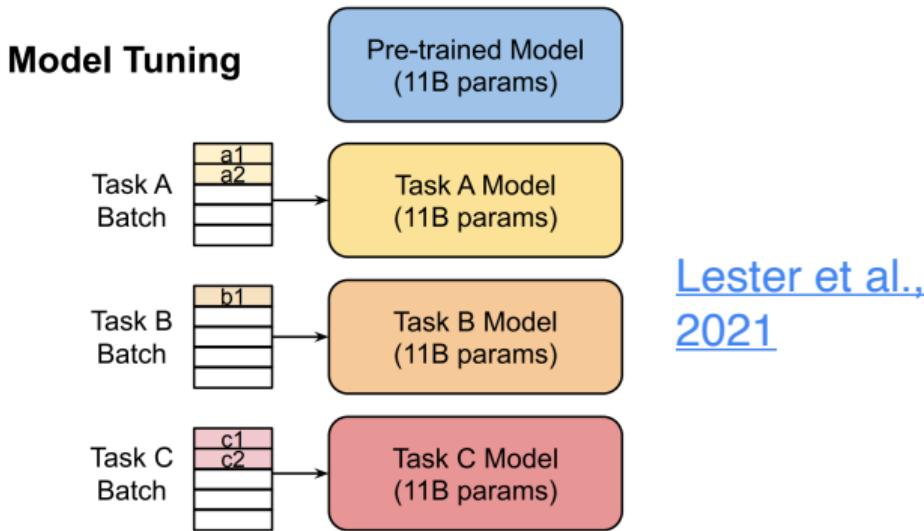
Struggles on “harder” datasets

# Scaling up the model size is one of the most important ingredients for achieving the best performance



[Ahmet and Abdullah., 2021](#)

# Practical challenges: large-scale models are costly to share and serve



## Language model prompting to the rescue!

GPT-3 ([Brown et al., 2020](#)): In-context learning

- **natural language instruction** and/or **a few task demonstrations** → output

“Translate English to German:” That is good → Das  
is gut

- *no* gradient updates or fine-tuning

# Sub-optimal and sensitive discrete/hard prompts

## Discrete/hard prompts

- natural language instructions/task descriptions

## Problems

- requiring domain expertise/understanding of the model's inner workings
- performance still lags far behind SotA model tuning results
- sub-optimal and sensitive
  - prompts that humans consider reasonable is not necessarily effective for language models ([Liu et al., 2021](#))
  - pre-trained language models are sensitive to the choice of prompts ([Zhao et al., 2021](#))

## Sub-optimal and sensitive discrete/hard prompts (cont.)

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

Table 1. Case study on LAMA-TREx P17 with bert-base-cased. A single-word change in prompts could yield a drastic difference.

[Liu et al., 2021](#)

# Shifting from discrete/hard to continuous/soft prompts

## Progress in prompt-based learning

- manual prompt design ([Brown et al., 2020; Schick and Schutze, 2021a,b](#))
- mining and paraphrasing based methods to automatically augment the prompt sets ([Jiang et al., 2020](#))
- gradient-based search for improved discrete/hard prompts ([Shin et al., 2020](#))
- automatic prompt generation using a separate generative language model (i.e., T5) ([Gao et al., 2020](#))
- learning continuous/soft prompts ([Liu et al., 2021; Li and Liang., 2021; Qin and Eisner., 2021; Lester et al., 2021](#))

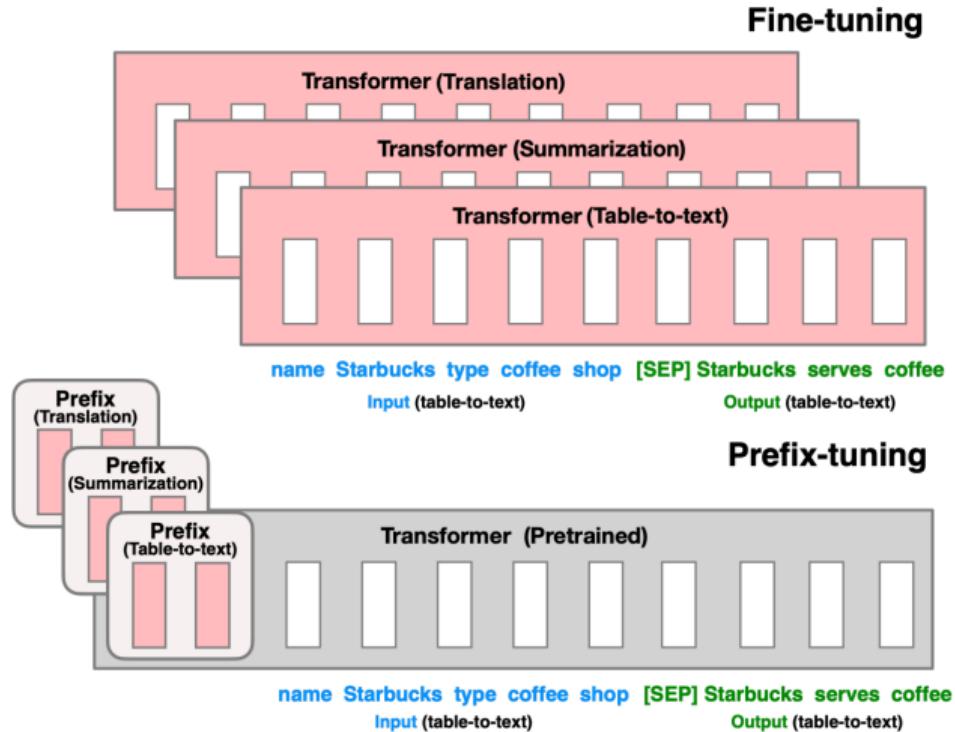
## Continuous/soft prompts

- additional learnable parameters injected into the model

## It remains unclear how to learn continuous/ soft prompts effectively?

- **P-tuning** ([Liu et al., 2021](#)): encode dependencies between prompt tokens using a BiLSTM network
- **P-tuning** ([Liu et al., 2021](#)), **Prefix Tuning** ([Li and Liang., 2021](#)): inject prompts at different positions of the input / model
- **P-tuning** ([Liu et al., 2021](#)): use mixed prompt initialization strategies
- **Soft Prompts** ([Qin and Eisner., 2021](#)): use ensemble methods, e.g., mixture-of-experts

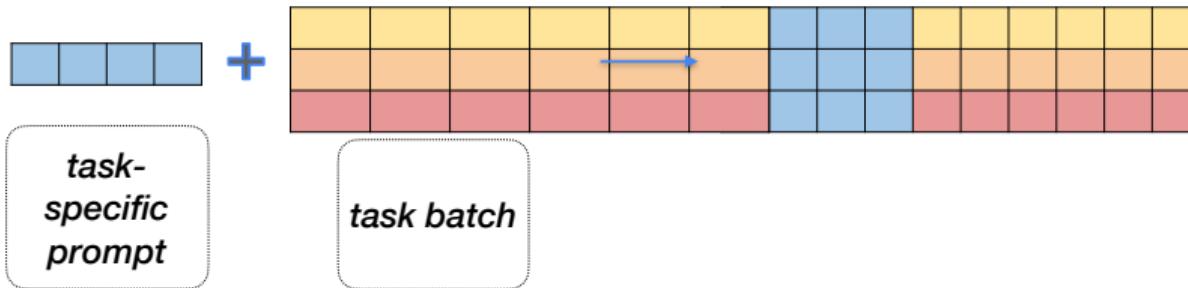
# Prefix tuning (Li & Liang, ACL 2021)



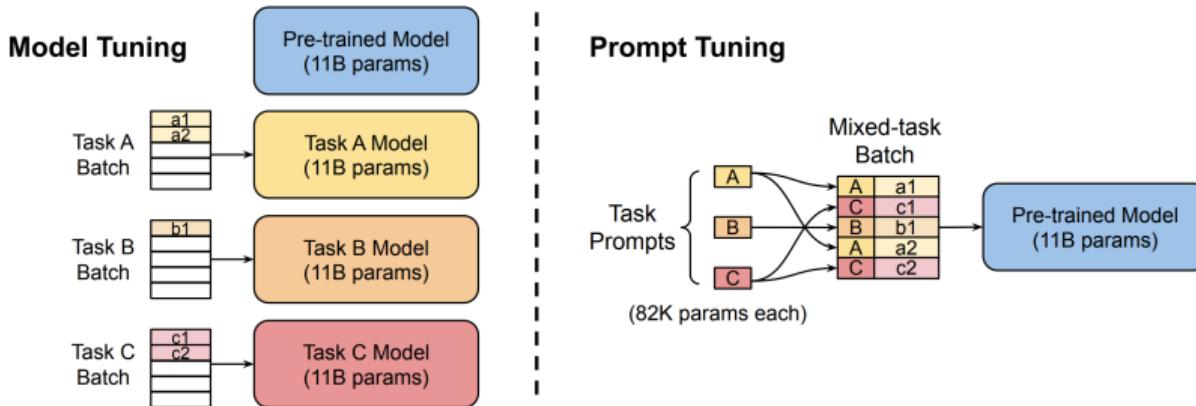
# Prompt Tuning idea ([Lester et al., 2021](#))

## What is a prompt in Prompt Tuning?

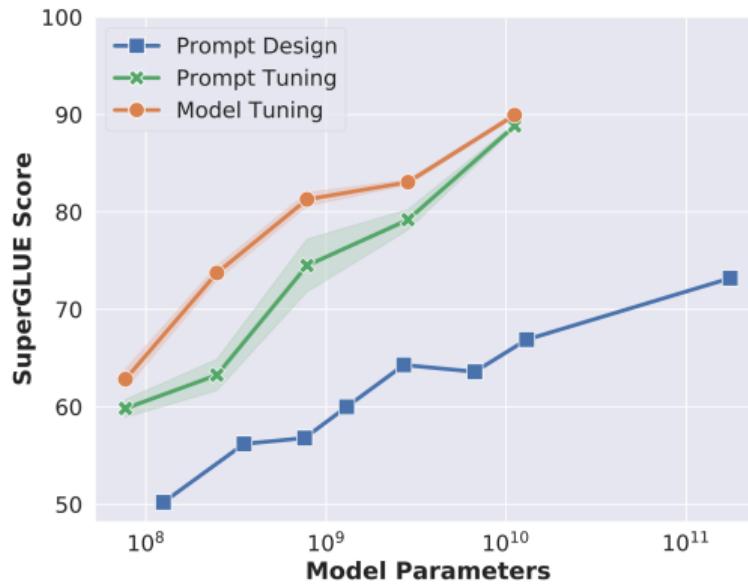
- a sequence of additional task-specific tunable tokens prepended to the input text



# Parameter-efficient Prompt Tuning



## Prompt Tuning becomes more competitive with scale

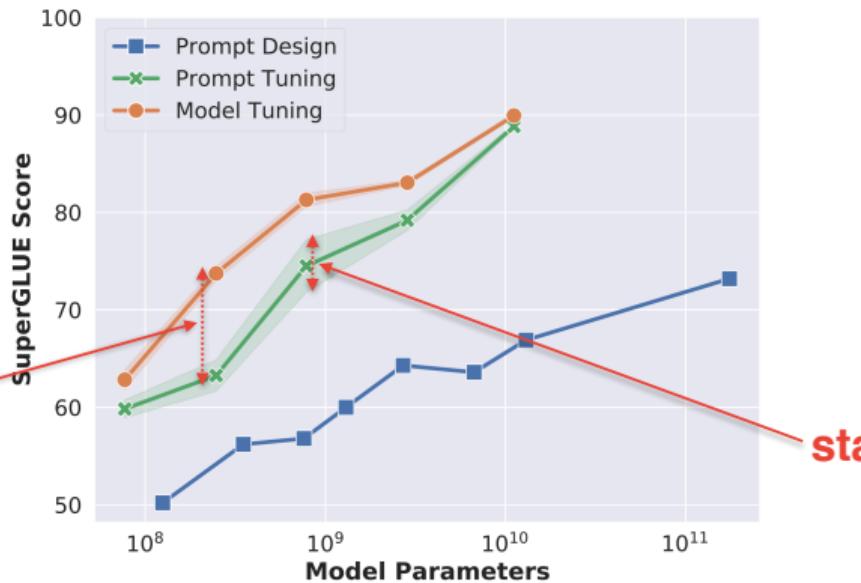


# Room for improving Prompt Tuning

Lester et al.,  
2021

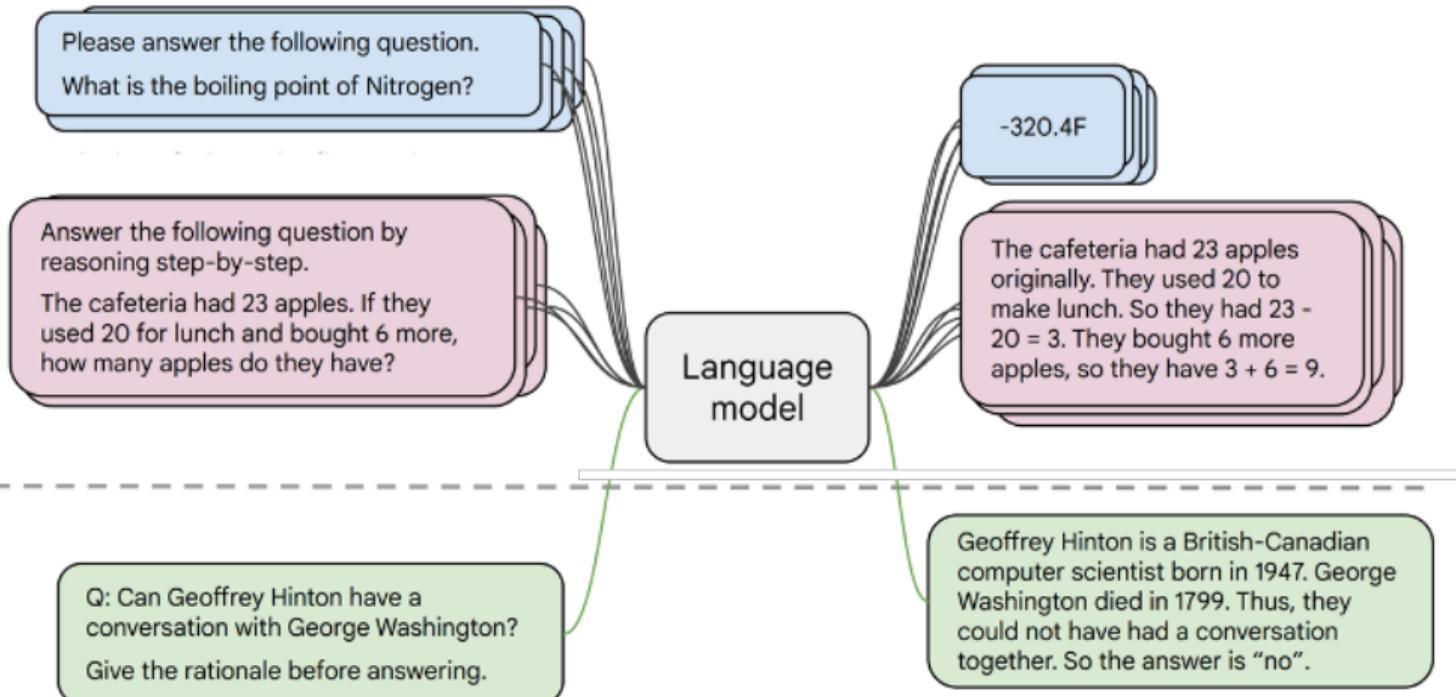
performance

stability



# Instruction Finetuning

Collect examples of (instruction, output) pairs across many tasks and finetune an LM



# Limitations of Instruction Finetuning

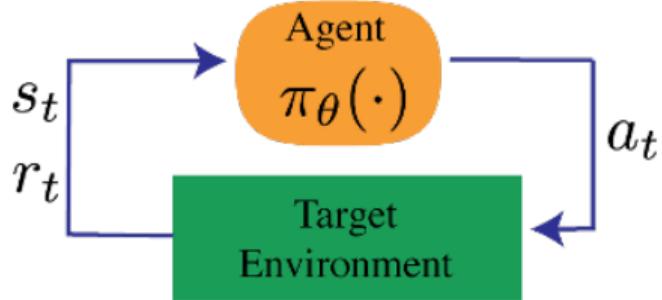
One limitation of instruction finetuning is obvious: it's expensive to collect ground-truth data for tasks

But there are other, subtler limitations too. Can you think of any?

**Problem 1:** tasks like open-ended creative generation have no right answer.

**Problem 2:** language modeling penalizes all token-level mistakes equally, but some errors are worse than others. Even with instruction finetuning, there a mismatch between the LM objective and the objective of "satisfy human preferences"!

# Review: Reinforcement Learning Basics



Some notation:

$s_t$ : state

$r_t$ : reward

$a_t$ : action

$a_t \sim \pi_\theta(s_t)$ : policy

# Why reinforcement learning from human feedback

How do you create /code a loss function for:

- ▶ What is safe?
- ▶ What is ethical?
- ▶ What is socially acceptable?

# Early OpenAI Experiments with RLHF

Human preference versus reference summaries

100%

80%

60%

40%

20%

0%

2e+9

3e+9

4e+9 5e+9

1e+10

Reference summaries

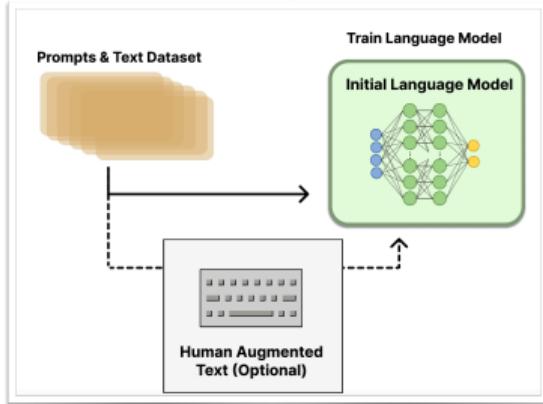
Supervised  
Pre-trained

Human feedback

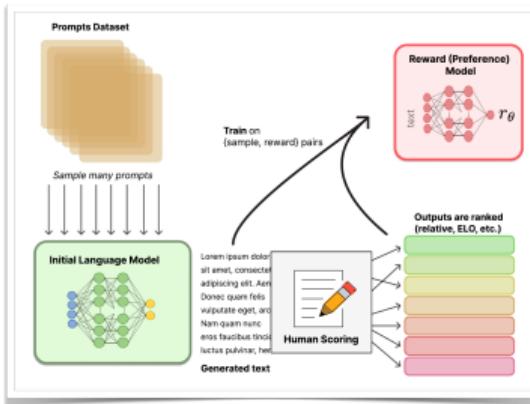
The performance of various training procedures for different model sizes. Model performance is measured by how often summaries from that model are preferred to the human-written reference summaries.

# Modern RLHF Overview

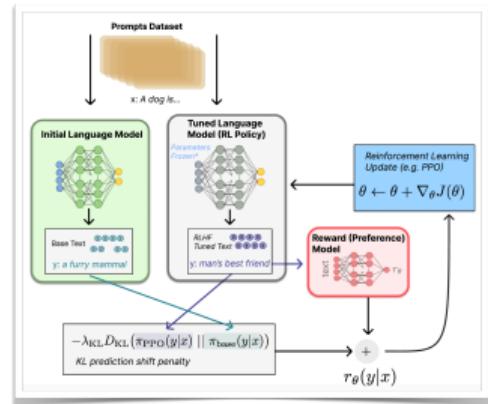
1. Language model pretraining



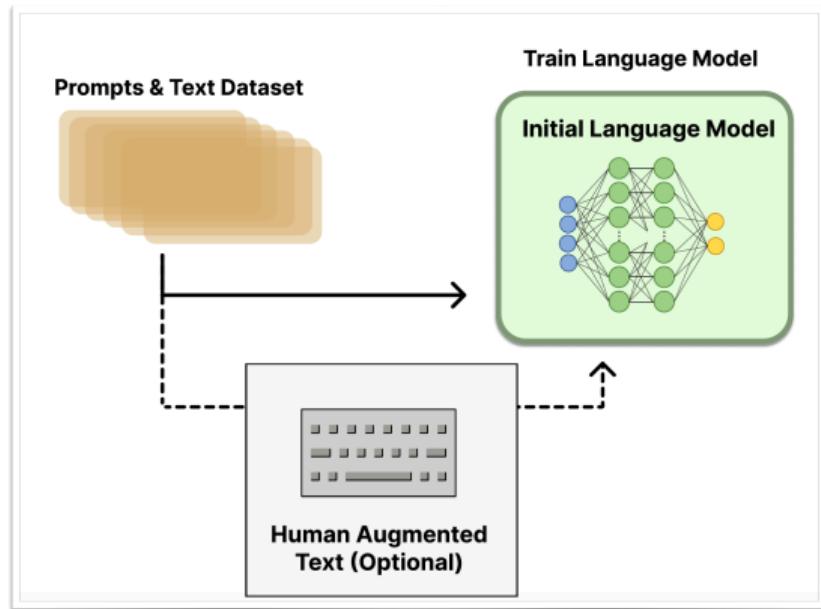
2. Reward model training



3. Fine-tuning with RL



# 1. Language model pertaining



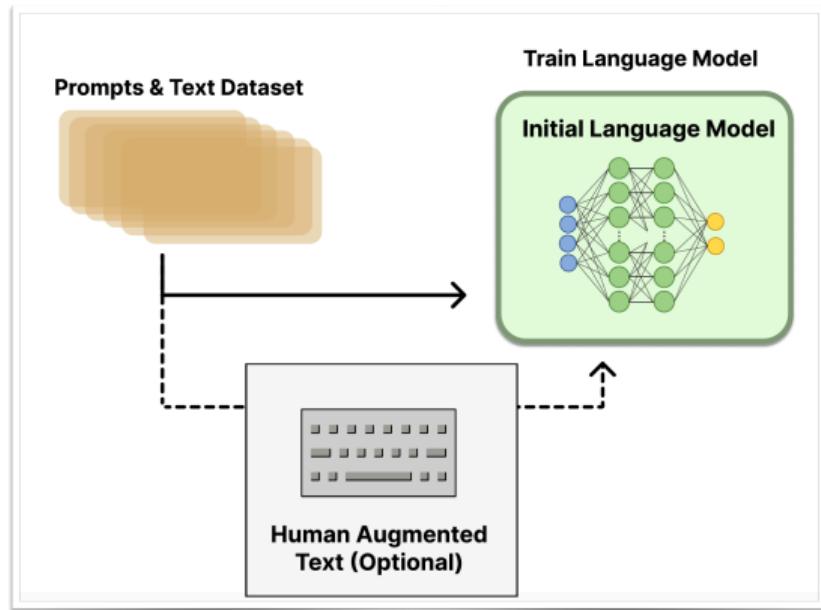
## Common training techniques in NLP:

Unsupervised sequence prediction

Data scrapped from the web

No single answer on “best” model size (examples in industry range 10B-280B parameters)

# 1. Language model pertaining

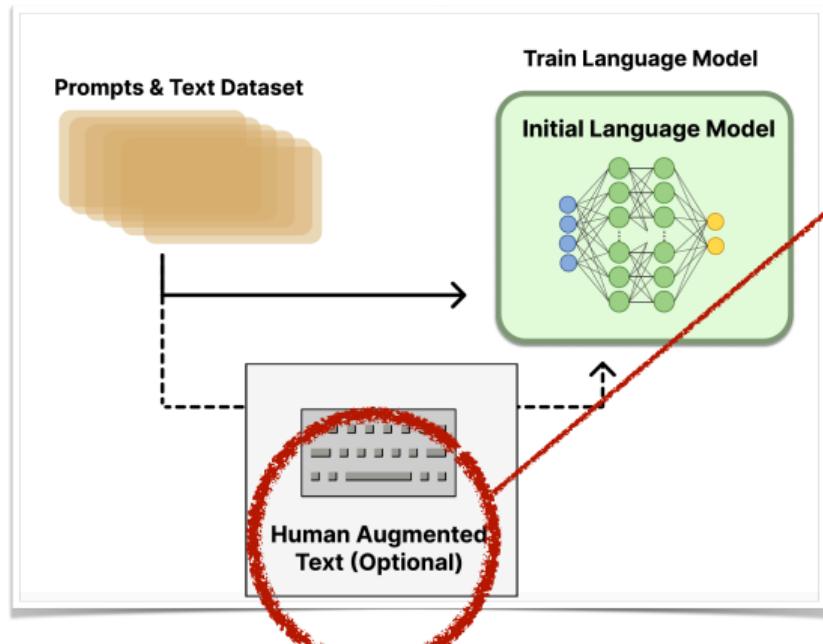


## Dataset:

Reddit, other forums, news, books

Optionally include human-written text from predefined prompts

# 1. Language model pertaining

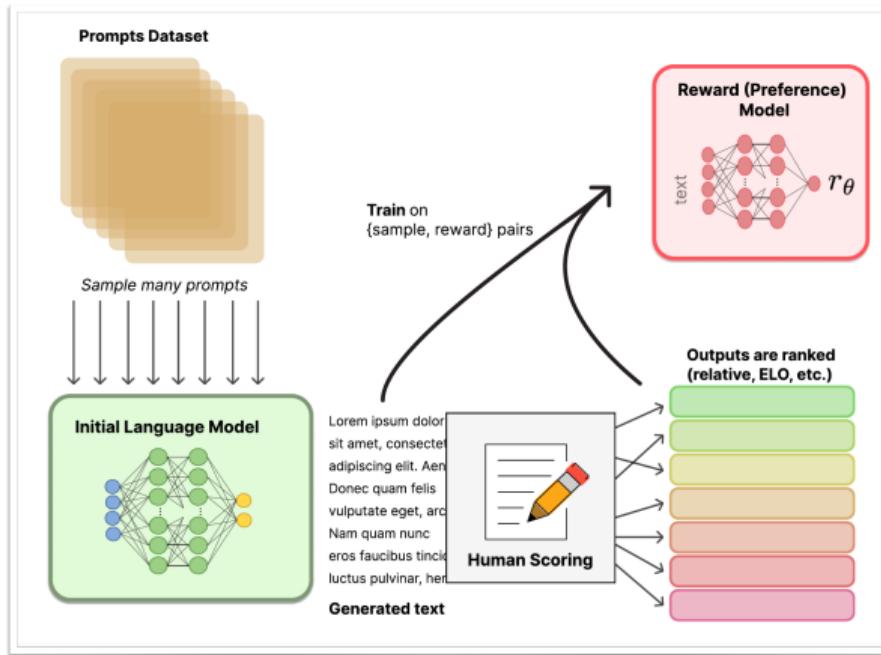


## Optional step:

Pay humans to write responses (\$\$\$), often viewed as high-quality initialization for RLHF

Supervised fine-tuning

## 2. Reward model training

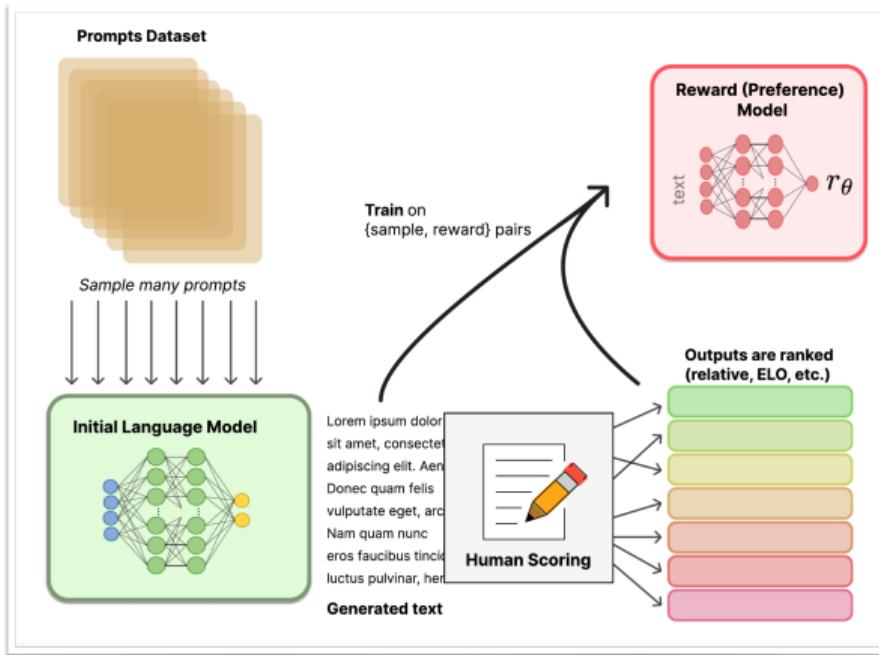


How to calculate human sentiments in samples and curated text?

**Goal:** get a model that maps

Input text → scalar reward

## 2. Reward model training

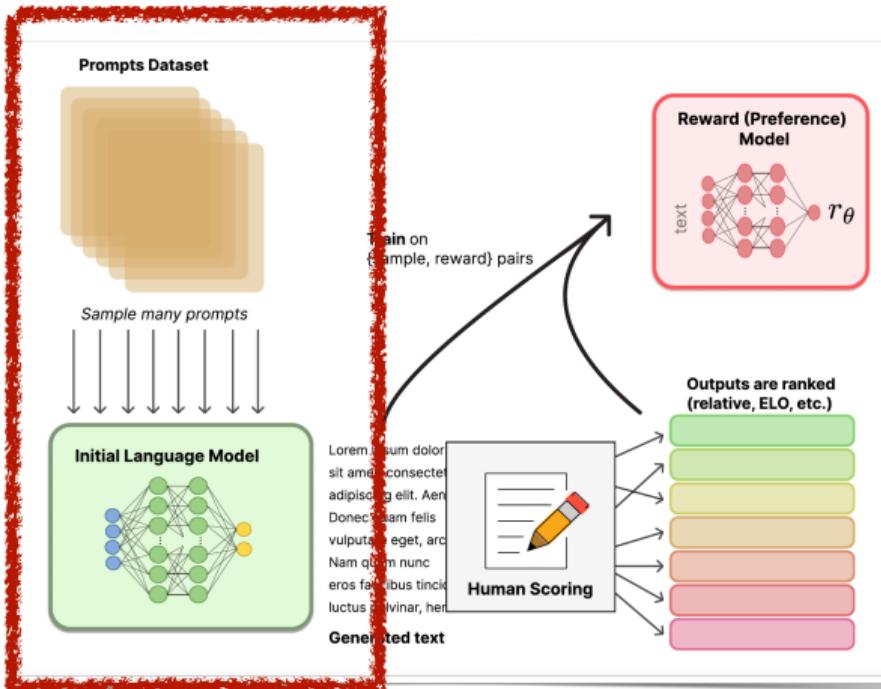


Why not human-in-the-loop?

Human-in-the-loop is expensive!

**Solution:** instead of directly asking humans for preferences, model their preferences as a separate (NLP) problem

## 2. Reward model training



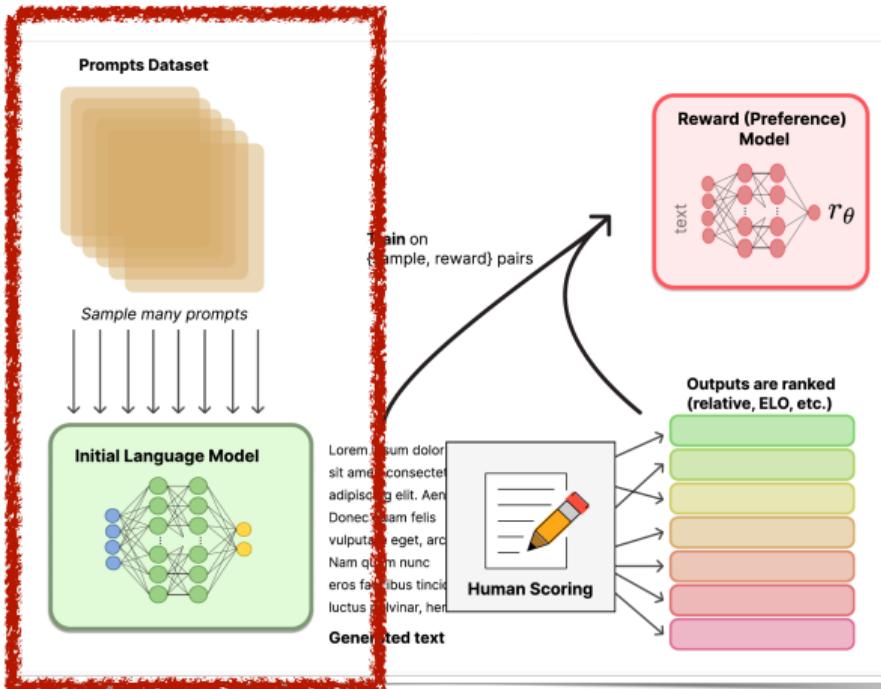
Prompts (input) dataset:

Prompts for specific use-case model will be used for

e.g., chat questions or prompt-based data

Much smaller than original pretraining

## 2. Reward model training



Generating data to rank:

Often can use multiple models to create diverse ranking,

## 2. Reward model training

Why ranking?

Human judgments are noisy and miscalibrated!

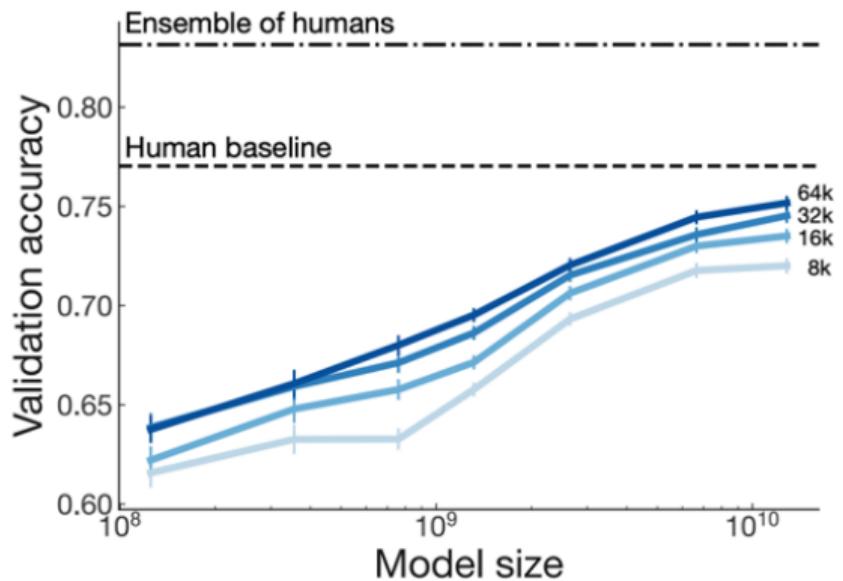
**Solution:** instead asking for direct ratings, ask for pairwise comparisons, which can be more reliable [Phelps et al., 2015; Clark et al., 2018]

$$-\mathbb{E}_{(s^w, s^l) \sim D} [\log \sigma(RM_\phi(s^w) - RM_\phi(s^l))]$$

“winning” sample    “losing” sample     $s^w$  should score higher than  $s^l$

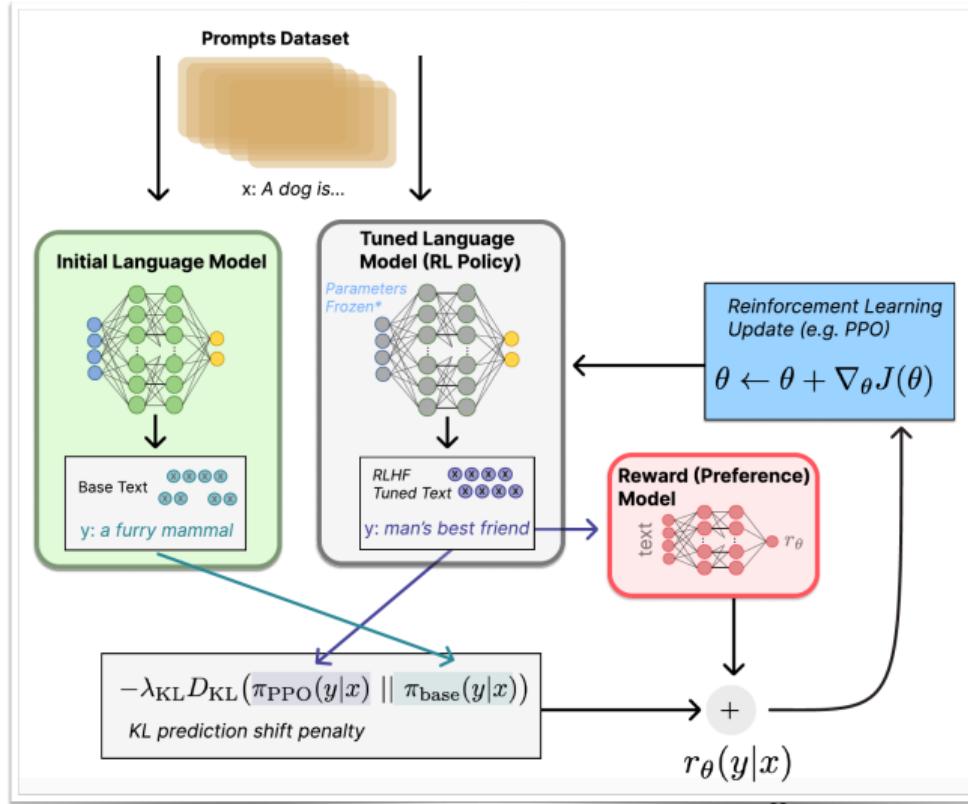
## 2. Reward model training

**Make sure the reward model works!**

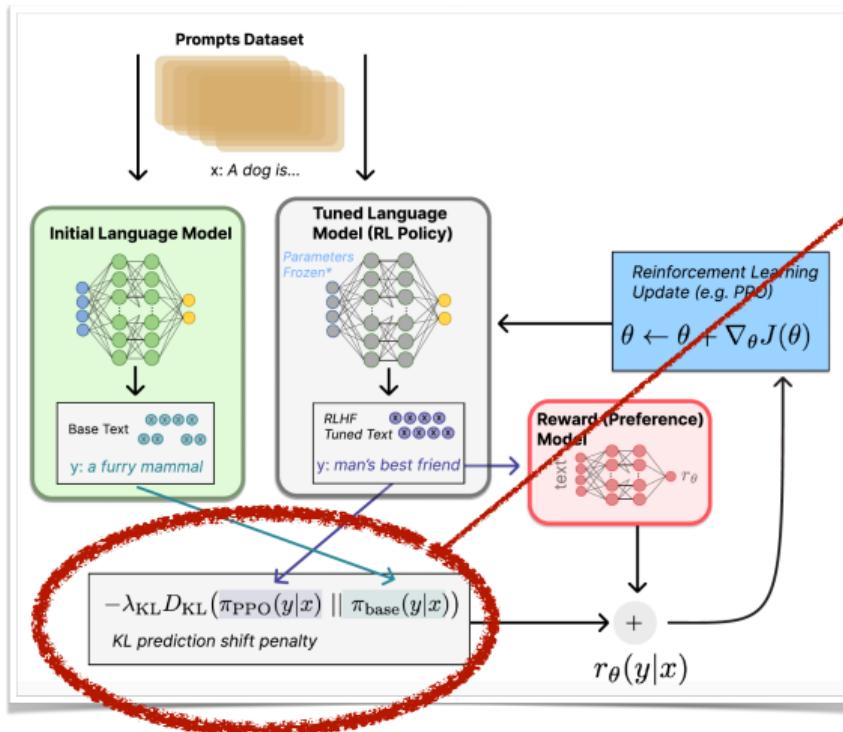


**Large enough RM  
trained on enough  
data approaching  
single human perf**

### 3. Fine tuning with RL: using a reward function



### 3. Fine tuning with RL: KL penalty

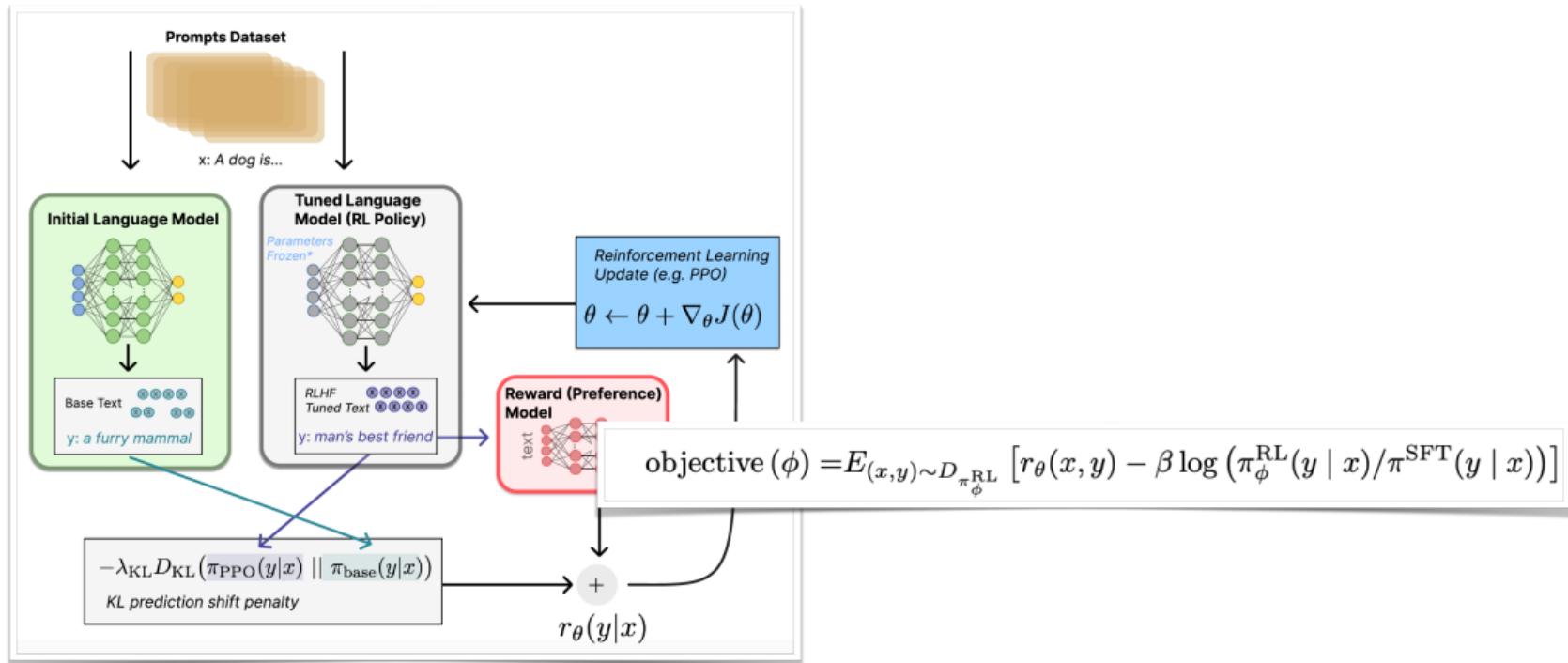


Kullback-Leibler (KL) divergence:

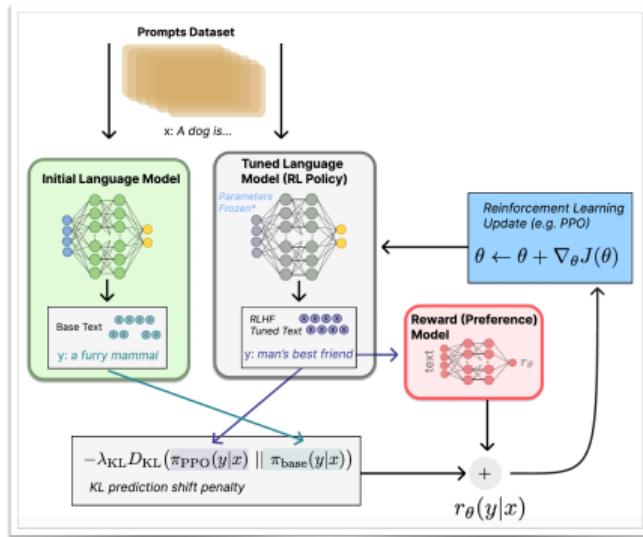
$D_{\text{KL}}(P \parallel Q)$  distance between distributions

Constraints the RL fine-tuning to not result in a LM that output gibberish (to fool the reward model)

### 3. Fine tuning with RL: Combine rewards



### 3. Fine tuning with RL: PPO



Policy gradient updates policy LM directly

Proximal Policy Optimization (PPO):

Optimized for parallelization

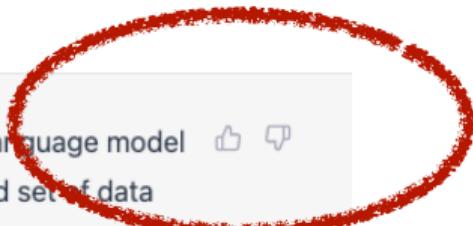
# Reward model training: feedback interfaces

DI

what's the weather today in palo alto



I'm sorry, but I am unable to access current weather information as I am a language model trained to provide general information and assist with tasks based on a fixed set of data that was current as of 2021. My knowledge and ability to browse the internet are limited to the information and resources that were available to me at the time of my training. I can provide you with general information about weather patterns in Palo Alto, California, or answer questions about other topics that you may have. Is there something else I can help you with?



Playground task 

## Talk to the assistant

Task 1 of 5

Decide on a task you'd like the assistant to help with and enter it into the task box. Interact with the AI assistant. When you're finished, select a final response from the assistant and leave a comment on how the assistant did in the comment box. Click to go to the next task.

Comments

I thought the assistant was ...

Rating

<input type="radio"/>				
Bad	Good			

Next Task



Human

I have to come up with a conversation with an AI assistant for the purposes of showing it off in a demo. What sorts of conversation topics would you recommend?



Assistant

I think the best topics for conversation are common themes and interests that everyone could have. Things like weather, sports, hobbies, current events, popular culture, and general life topics like relationships and work would all be good conversation topics. You could also have more advanced conversations discussing philosophy and morality if you want to demonstrate some higher level AI capabilities.



Human

I think I'm more interested in an advanced conversation. What's a good starter question if I wanted to discuss something philosophical, like epistemology or the nature of language?



Assistant

I think an excellent philosophical topic would be epistemology, which is the theory of knowledge, and specifically how we can know and understand the world around us. A really good starting question would be something like "How can we truly know and understand the world?", or "What is the nature of knowledge?". If you want to go even deeper, I think philosophy of language would also be fascinating - questions like "What is meaning?", "How do language and thought relate?", and "How can language mislead and deceive?" would be great discussion topics.



Human

How would you answer a question like: How do language and thought relate?



Choose the most helpful and honest response

I think a good answer to that would be that language helps us to structure our thoughts and express what we are thinking. Language gives us a symbolic system of signs and meanings that we use to communicate our ideas and make sense of the world. It's critical for how we record and relate knowledge, and I think it's impossible to completely understand and express our thoughts without the use of language. Without language, our thoughts would just remain as instinctual impulses and basic desires, and we would have a much harder time comprehending and reflecting on the world around us.



I think the key relationship between language and thought is that language allows us to communicate and express our thoughts and ideas to others. Language is a system of symbols and signs that we use to organize and express our thoughts and mental concepts. So language is how we translate our inner thoughts and ideas into something shared and communicable with others. Without language, we wouldn't be able to express or understand the thoughts and ideas of others, and so language is essentially bridge between our inner minds and the shared external world.

A    A    A    A    B    B    B    B

A is better

B is better

# Reward model training: feedback interfaces

The opportunity for  
text feedback!

NA If you were a storyteller, would you think ChatGPT is good for society?

 As a large language model trained by OpenAI, I do not have personal opinions or beliefs. I am a tool designed to assist with answering questions and providing information on a wide range of topics. I do not have the ability to evaluate the impact of ChatGPT or any other technology on society. My function is to provide accurate and reliable information to the best of my ability.

best of my data

# Limitation of RLHF

## Human preferences are unreliable!

"Reward hacking" is a common problem in RL

Chatbots are rewarded to produce responses that seem authoritative and helpful, regardless of truth

This can result in making up facts + hallucinations

## TECHNOLOGY

### Google shares drop \$100 billion after its new AI chatbot makes a mistake

February 9, 2023 · 10:15 AM ET

## Bing AI hallucinates the Super Bowl

The screenshot shows a Bing search results page. At the top right, there is a blue button with the text "Who won the superbowl?". Below it, the search query "Who won the superbowl?" is displayed. The main content area contains text about the Super Bowl LVI, stating that the most recent Super Bowl was Super Bowl LVI, which was held on February 6, 2023, at SoFi Stadium in Inglewood, California. It mentions that the Philadelphia Eagles defeated the Kansas City Chiefs by 31-24. A small note below says "Eagles, who defeated the Kansas City Chiefs by 31-24". At the bottom of the page, there are links to learn more from en.wikipedia.org, sportingnews.com, and cbssports.com.

<https://news.ycombinator.com/item?id=34776508>

<https://apnews.com/article/kansas-city-chiefs-philadelphia-eagles-technology-science-82bc20f207e3e4cf81abc6a5d9e6b23a>

# Limitation of RLHF

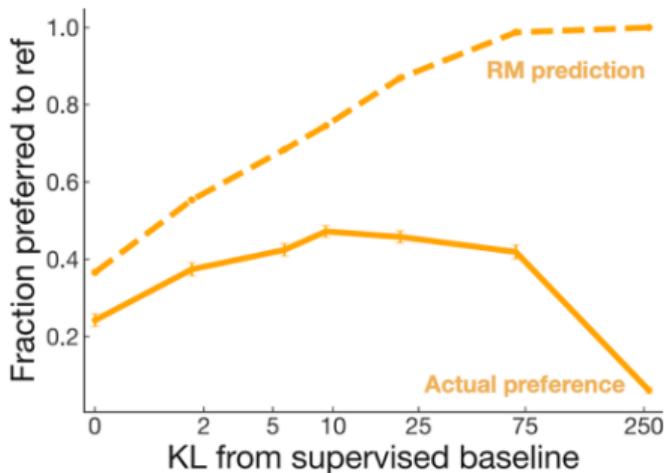
## Human preferences are unreliable!

"Reward hacking" is a common problem in RL

Chatbots are rewarded to produce responses that seem authoritative and helpful, regardless of truth

This can result in making up facts + hallucinations

Reward model over-optimization



$$R(s) = RM_{\phi}(s) - \beta \log \left( \frac{p_{\theta}^{RL}(s)}{p^{PT}(s)} \right)$$

# Limitation of RLHF

## Human preferences are unreliable!

"Reward hacking" is a common problem in RL

Chatbots are rewarded to produce responses that seem authoritative and helpful, regardless of truth

This can result in making up facts + hallucinations

BUSINESS • TECHNOLOGY

Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic



OpenAI ChatGPT Sama