

# Lost of data, lots of parameters: a better way to get task zero-shot generalization

Victor Sanh

# Lost of data, lots of parameters: a better way to get task zero-shot generalization



... and many other awesome collaborators

## *Unsupervised pre-training*

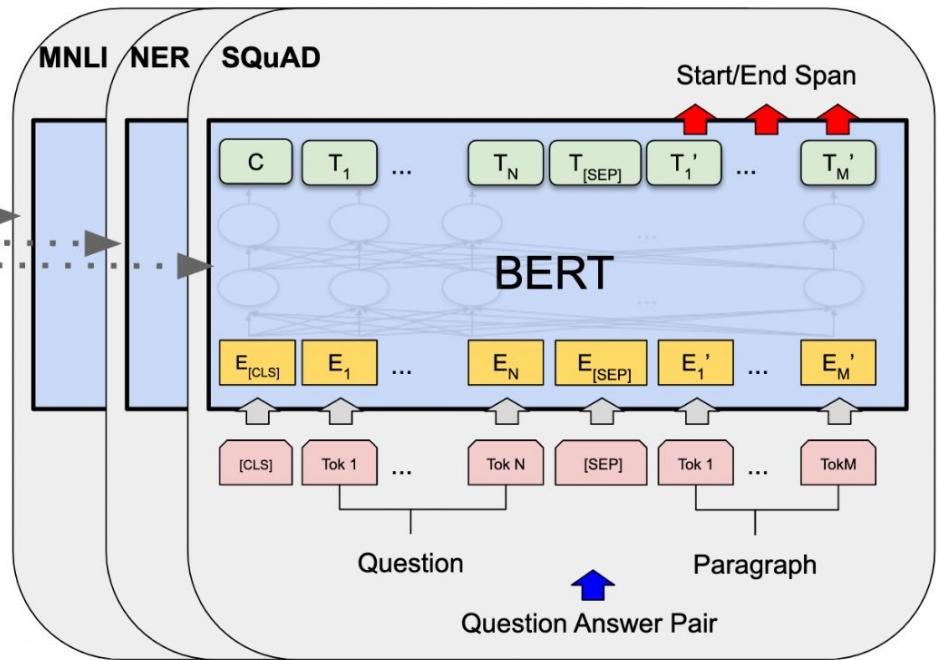
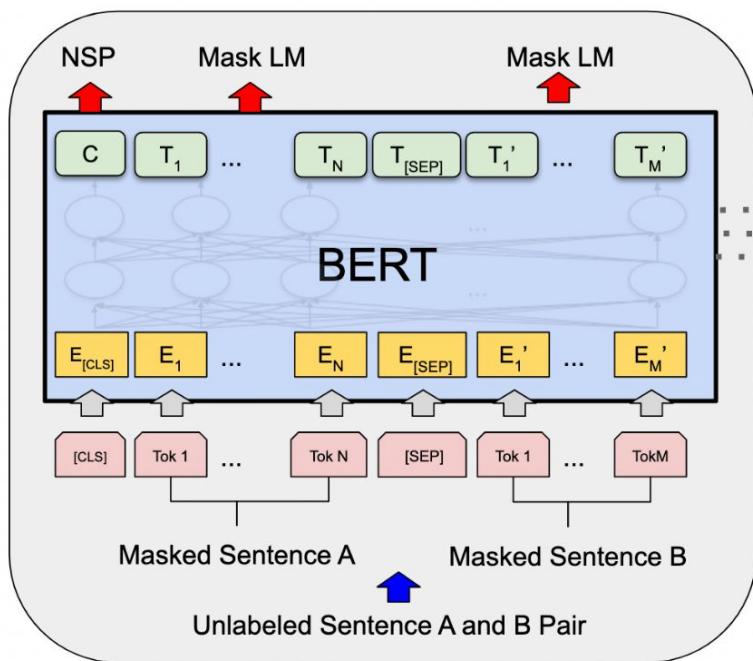
The cabs \_\_\_ the same rates as those \_\_\_ by horse-drawn cabs and were \_\_\_ quite popular, \_\_\_ the Prince of Wales (the \_\_\_ King Edward VII) travelled in \_\_\_. The cabs quickly \_\_\_ known as "hummingbirds" for \_\_\_ noise made by their motors and their distinctive black and \_\_\_ livery. Passengers \_\_\_ \_\_\_ the interior fittings were \_\_\_ when compared to \_\_\_ cabs but there \_\_\_ some complaints \_\_\_ the \_\_\_ lighting made them too \_\_\_ to those outside \_\_\_.

## *Supervised fine-tuning*

This movie is terrible! The acting is bad and I was bored the entire time. There was no plot and nothing interesting happened. I was really surprised since I had very high expectations. I want 103 minutes of my life back!

negative

charged, used, initially, even, future, became, the, yellow, reported, that, luxurious, horse-drawn, were that, internal, conspicuous, cab



"translate English to German: That is good."

"cola sentence: The course is jumping well."

"stsbt sentence1: The rhino grazed on the grass. sentence2: A rhino is grazing in a field."

"summarize: state authorities dispatched emergency crews tuesday to survey the damage after an onslaught of severe weather in mississippi..."

T5

"Das ist gut."

"not acceptable"

"3.8"

"six people hospitalized after a storm in attala county."

## *Unsupervised pre-training*

The cabs charged the same rates as those used by horse-drawn cabs and were initially quite popular; even the Prince of Wales (the future King Edward VII) travelled in one. The cabs quickly became known as "hummingbirds" for the noise made by their motors and their distinctive black and yellow livery. Passengers reported that the interior fittings were luxurious when compared to horse-drawn cabs but there were some complaints that the internal ...



lighting made them too conspicuous to those outside the cab. The fleet peaked at around 75 cabs, all of which needed to return to the single depot at Lambeth to switch batteries.



## *"Zero-shot" prompting*

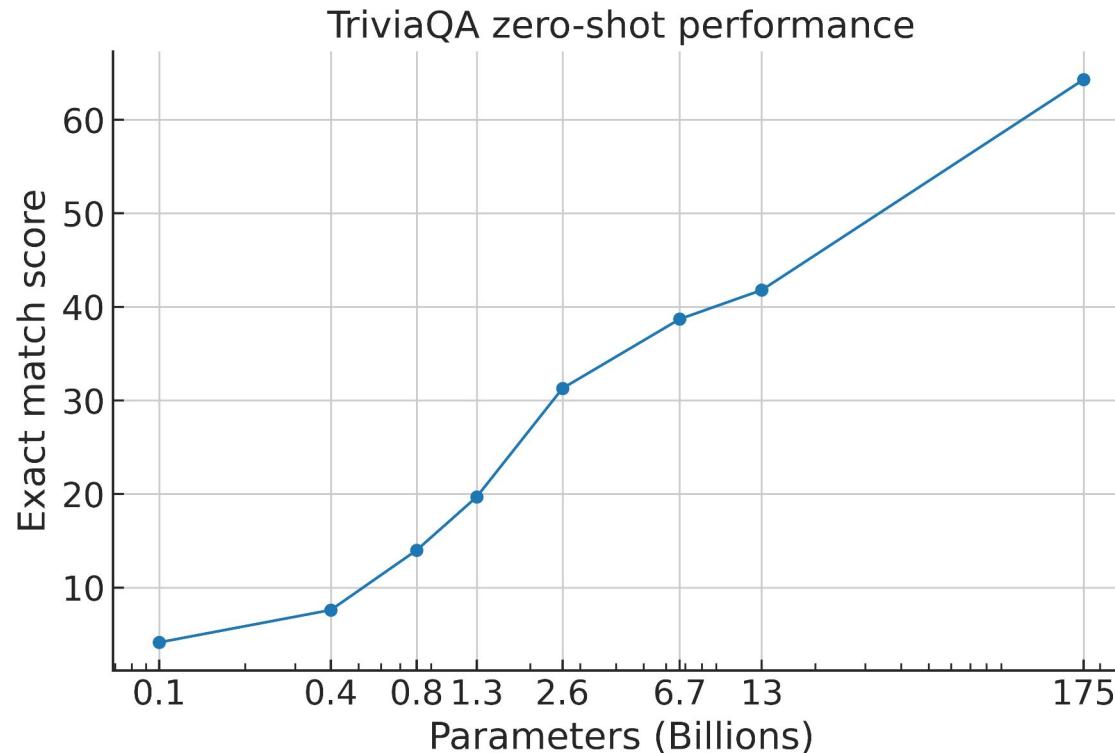
Suppose "The banker contacted the professors and the athlete". Can we infer that "The banker contacted the professors"?

yes

---

# Language Models are Few-Shot Learners

---



## Closed-book question answering

<http://www.autos weblog.com/cat/trivia-questions-from-the-50s>

who was frank sinatra? a: an american singer, actor, and producer.

## Paraphrase identification

<https://www.usingenglish.com/forum/threads/60200-Do-these-sentences-mean-the-same>

Do these sentences mean the same? No other boy in this class is as smart as the boy. No other boy is as smart as the boy in this class.

## Natural Language Inference

<https://ell.stackexchange.com/questions/121446/what-does-this-sentence-imply>

If I say: He has worked there for 3 years. does this imply that he is still working at the moment of speaking?

## Summarization

<https://blog.nytsoi.net/tag/reddit>

... Lately I've been seeing a pattern regarding videos stolen from other YouTube channels, reuploaded and monetized with ads. These videos are then mass posted on Reddit by bots masquerading as real users. tl;dr: Spambots are posting links to stolen videos on Reddit, copying comments from others to masquerade as legitimate users.

## Pronoun resolution

<https://nursecheung.com/ati-teas-guide-to-english-language-usage-understanding-pronouns/>

Jennifer is a vegetarian, so she will order a nonmeat entrée. In this example, the pronoun she is used to refer to Jennifer.

# UNIFIEDQA: Crossing Format Boundaries with a Single QA System

Daniel Khashabi<sup>1</sup>   Sewon Min<sup>2</sup>   Tushar Khot<sup>1</sup>   Ashish Sabharwal<sup>1</sup>  
Oyvind Tafjord<sup>1</sup>   Peter Clark<sup>1</sup>   Hannaneh Hajishirzi<sup>1,2</sup>

<sup>1</sup>Allen Institute for AI, Seattle, U.S.A.

<sup>2</sup>University of Washington, Seattle, U.S.A.

## Extractive [SQuAD]

**Question:** At what speed did the turbine operate?

**Context:** (Nikola\_Tesla) On his 50th birthday in 1906, Tesla demonstrated his 200 horsepower (150 kilowatts) 16,000 rpm bladeless turbine. ...

**Gold answer:** 16,000 rpm

## Multiple-Choice [ARC-challenge]

**Question:** What does photosynthesis produce that helps plants grow?

**Candidate Answers:** (A) water (B) oxygen (C) protein (D) sugar

**Gold answer:** sugar

## Abstractive [NarrativeQA]

**Question:** What does a drink from narcissus's spring cause the drinker to do?

**Context:** Mercury has awakened Echo, who weeps for Narcissus, and states that a drink from Narcissus's spring causes the drinkers to "Grow dotingly enamored of themselves." ...

**Gold answer:** fall in love with themselves

## Yes/No [BoolQ]

**Question:** Was America the first country to have a president?

**Context:** (President) The first usage of the word president to denote the highest official in a government was during the Commonwealth of England ...

**Gold answer:** no

# The Natural Language Decathlon: Multitask Learning as Question Answering

Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, Richard Socher

Salesforce Research

{bmccann,nkeskar,cxiong,rsocher}@salesforce.com

## Examples

Question	Context	Answer	Question	Context	Answer
What is a major importance of Southern California in relation to California and the US?	...Southern California is a <b>major economic center</b> for the state of California and the US....	major economic center	What has something experienced?	Areas of the Baltic that have experienced <b>eutrophication</b> .	eutrophication
What is the translation from English to German?	Most of the planet is ocean water.	Der Großteil der Erde ist Meerwasser	Who is the illustrator of Cycle of the Werewolf?	Cycle of the Werewolf is a short novel by Stephen King, featuring illustrations by comic book artist <b>Bernie Wrightson</b> .	Bernie Wrightson
What is the summary?	Harry Potter star Daniel Radcliffe gains access to a reported £320 million fortune...	Harry Potter star Daniel Radcliffe gets £320M fortune...	What is the change in dialogue state?	Are there any Eritrean restaurants in town?	food: Eritrean
Hypothesis: Product and geography are what make cream skimming work. <b>Entailment</b> , neutral, or contradiction?	Premise: Conceptually cream skimming has two basic dimensions – product and geography.	Entailment	What is the translation from English to SQL?	The <b>table</b> has column names... Tell me what the <b>notes</b> are for <b>South Australia</b>	SELECT notes from table WHERE 'Current Slogan' = 'South Australia'
Is this sentence positive or negative?	A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s horror film.	positive	Who had given help? Susan or Joan?	Joan made sure to thank Susan for all the help she had given.	Susan

# MEASURING MASSIVE MULTITASK LANGUAGE UNDERSTANDING

**Dan Hendrycks**  
UC Berkeley

**Collin Burns**  
Columbia University

**Steven Basart**  
UChicago

**Andy Zou**  
UC Berkeley

**Mantas Mazeika**  
UIUC

**Dawn Song**  
UC Berkeley

**Jacob Steinhardt**  
UC Berkeley

---

## Few Shot Prompt and Predicted Answer

The following are multiple choice questions about high school mathematics.

How many numbers are in the list 25, 26, ..., 100?  
(A) 75 (B) 76 (C) 22 (D) 23

Answer: B

Compute  $i + i^2 + i^3 + \dots + i^{258} + i^{259}$ .  
(A) -1 (B) 1 (C)  $i$  (D) - $i$

Answer: A

If 4 daps = 7 yaps, and 5 yaps = 3 baps,  
how many daps equal 42 baps?  
(A) 28 (B) 21 (C) 40 (D) 30

Answer: C

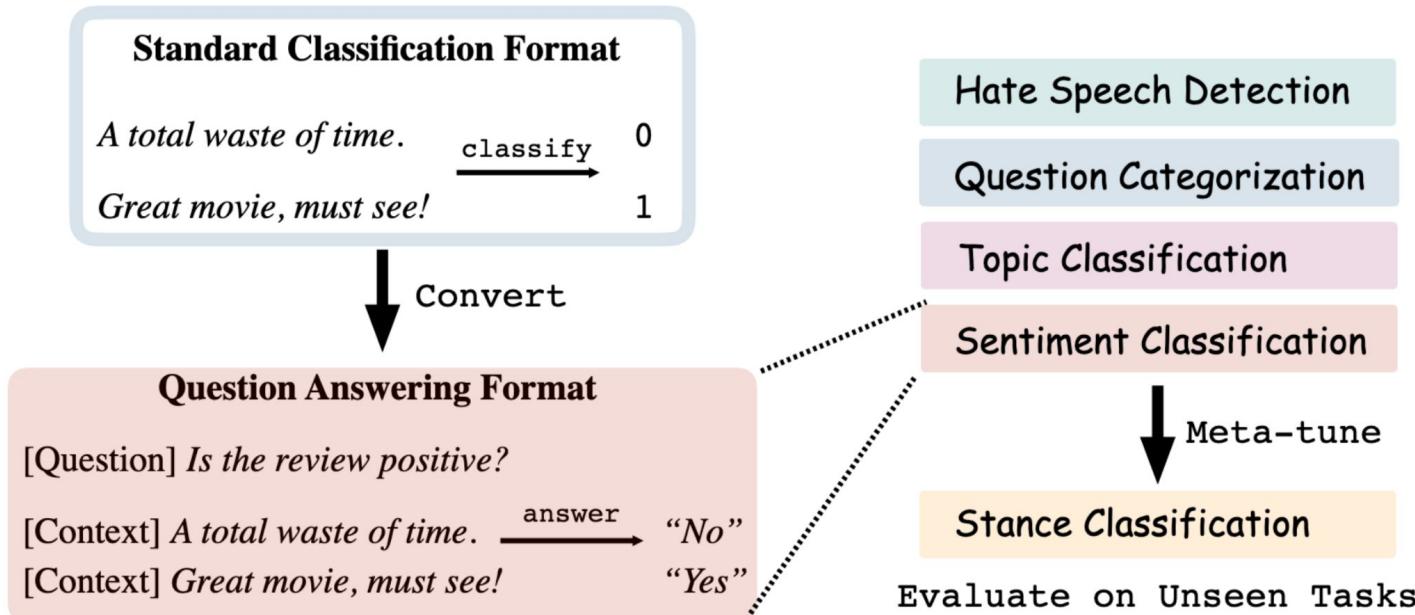
Model	Humanities	Social Science	STEM	Other	Average
Random Baseline	25.0	25.0	25.0	25.0	25.0
RoBERTa	27.9	28.8	27.0	27.7	27.9
ALBERT	27.2	25.7	27.7	27.9	27.1
GPT-2	32.8	33.3	30.2	33.1	32.4
UnifiedQA	45.6	56.6	40.2	54.6	48.9
GPT-3 Small (few-shot)	24.4	30.9	26.0	24.1	25.9
GPT-3 Medium (few-shot)	26.1	21.6	25.6	25.5	24.9
GPT-3 Large (few-shot)	27.1	25.6	24.3	26.5	26.0
GPT-3 X-Large (few-shot)	40.8	50.4	36.7	48.8	43.9

# Meta-tuning Language Models to Answer Prompts Better

Ruiqi Zhong Kristy Lee\* Zheng Zhang\* Dan Klein

Computer Science Division, University of California, Berkeley

{ruiqi-zhong, kristylee, zhengzhang1216, klein}@berkeley.edu

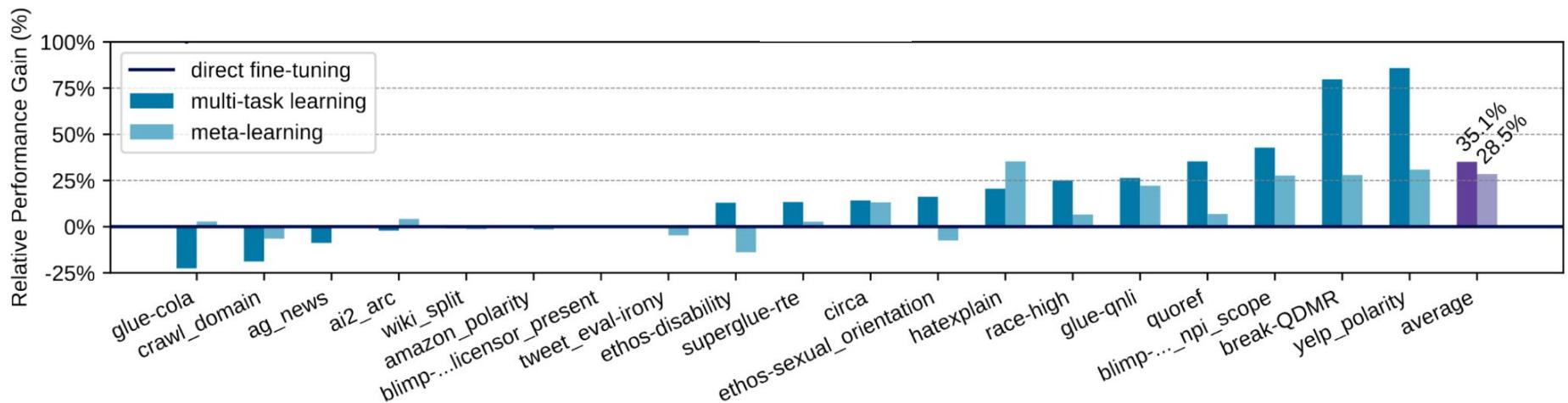


# CROSSFIT : A Few-shot Learning Challenge for Cross-task Generalization in NLP

**Qinyuan Ye Bill Yuchen Lin Xiang Ren**

University of Southern California

{qinyuany, yuchen.lin, xiangren}@usc.edu



Can we obtain good **zero-shot task generalization** by training a large language model on a **massively multitask mixture of diversely prompted datasets?**

## Summarization

The picture appeared on the wall of a Poundland store on Whymark Avenue [...] How would you rephrase that in a few words?

## Paraphrase identification

"How is air traffic controlled?" "How do you become an air traffic controller?"  
Pick one: these questions are duplicates or not duplicates.

## Question answering

I know that the answer to "What team did the Panthers defeat?" is in "The Panthers finished the regular season [...]" . Can you tell me what it is?

## Multi-task training

## Zero-shot generalization

## Natural language inference

Suppose "The banker contacted the professors and the athlete". Can we infer that "The banker contacted the professors"?

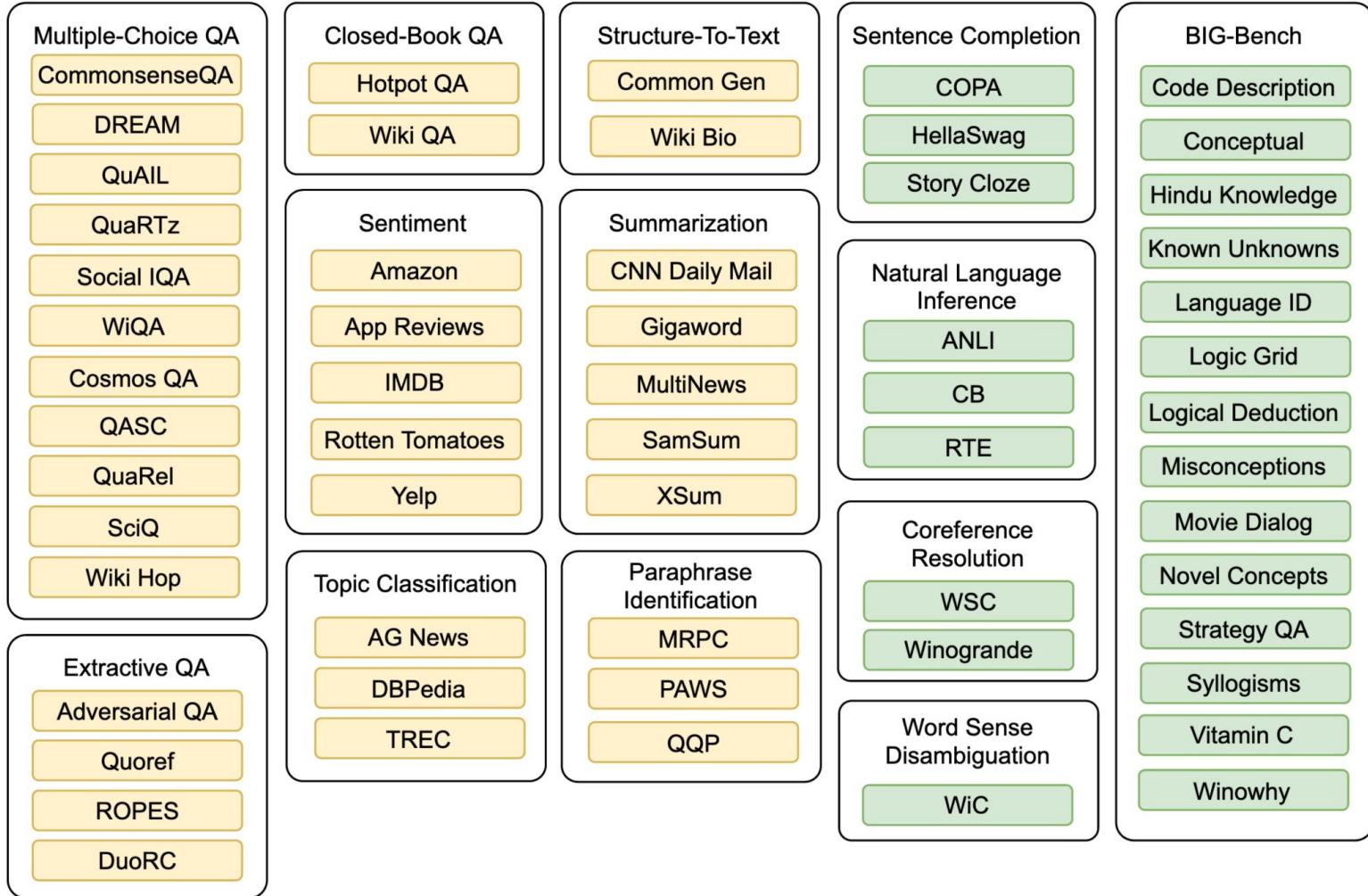
T0

Graffiti artist Banksy is believed to be behind [...]

Not duplicates

Arizona Cardinals

Yes



Name  
i\_am\_hesitating

Template Reference

Original Task? [?](#)

Choices in Prompt? [?](#)

Metrics  
Accuracy [X](#)

Answer Choices  
{{choice1}} ||| {{choice2}}

Template  
{{ premise }}

I am hesitating between two options. Help me choose the more likely % if question == "cause" %} cause: {%- else %} effect: {%- endif %}  
- {{choice1}}  
- {{choice2}} ||| {%- if label != -1 %}{% answer\_choices[label] %}{%- endif %}

Save

### Prompt + X

My body cast a shadow over the grass.

I am hesitating between two options. Help me choose the more likely cause:

- The sun was rising.
- The grass was cut.

Y

The sun was rising.

# Number of prompted datasets: 170

# Number of prompts: 2052

Name  
i\_am\_hesitating

Template Reference

Original Task? ?

Choices in Prompt? ?

Accuracy X

Answer Choices  
{{choice1}} ||| {{choice2}}

Template

```
{{ premise }}  
  
I am hesitating between two options. Help me choose the more likely {  
% if question ==  
"cause" %} cause: {  
% else %} effect: {  
% endif %}  
- {{choice1}}  
- {{choice2}} ||| {  
% if label != -1 %}{  
% answer_choices[label] %}{  
% endif %}
```

Save

Prompt + X

My body cast a shadow over  
the grass.

I am hesitating between two  
options. Help me choose the  
more likely cause:

- The sun was rising.
- The grass was cut.

The sun was rising.  
Y

## QQP (Paraphrase)

Question1	How is air traffic controlled?
Question2	How do you become an air traffic controller?
Label	0

{Question1} {Question2}  
Pick one: These questions  
are duplicates or not  
duplicates.

{Choices[label]}

I received the questions  
"{Question1}" and  
"{Question2}". Are they  
duplicates?

{Choices[label]}

## XSum (Summary)

Document	The picture appeared on the wall of a Poundland store on Whymark Avenue...
Summary	Graffiti artist Banksy is believed to be behind...

{Document}  
How would you  
rephrase that in  
a few words?

{Summary}

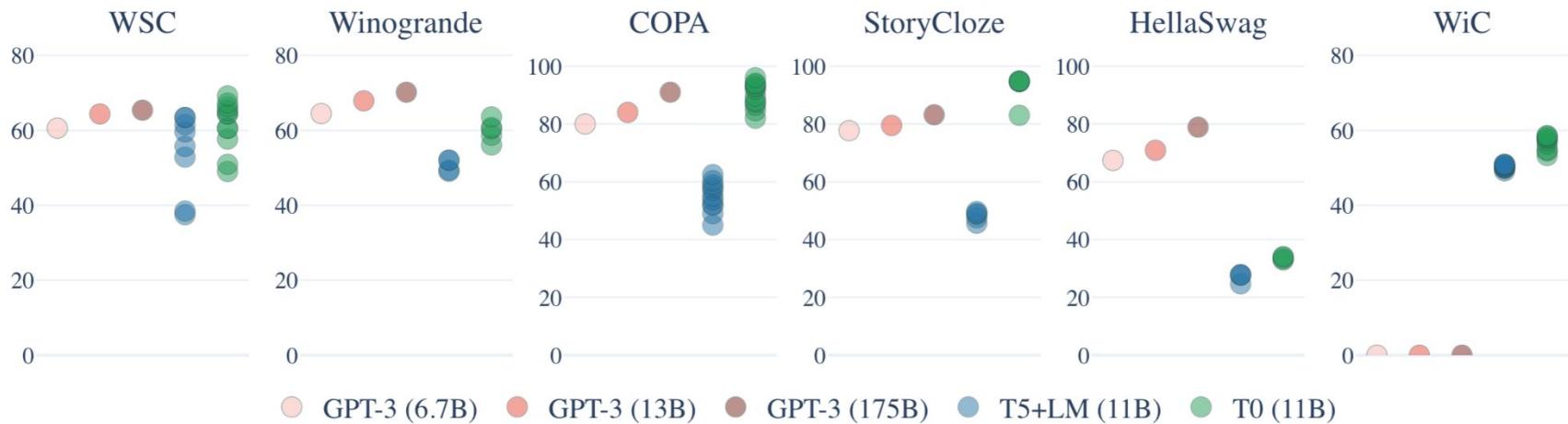
First, please read the article:  
{Document}  
Now, can you write me an  
extremely short abstract for it?

{Summary}

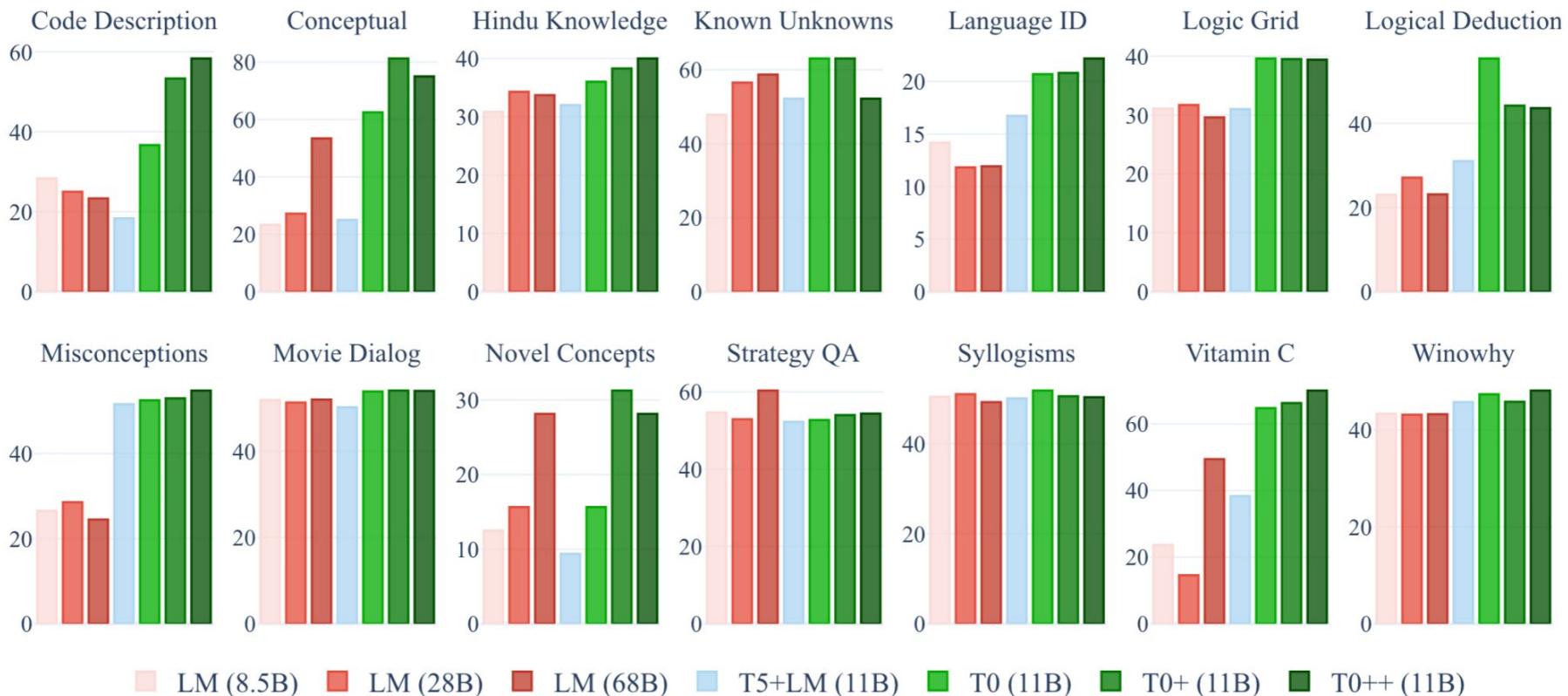
## Natural Language Inference



## Coreference Resolution



# Big Bench



# *More prompts are better than one*

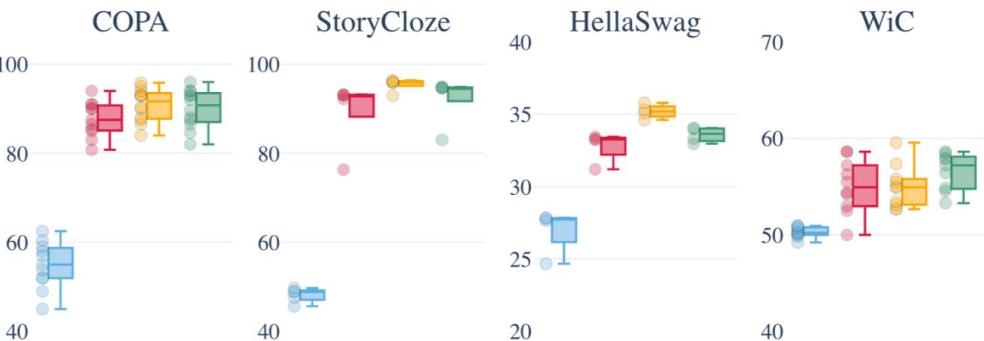
Natural Language Inference



Coreference Resolution

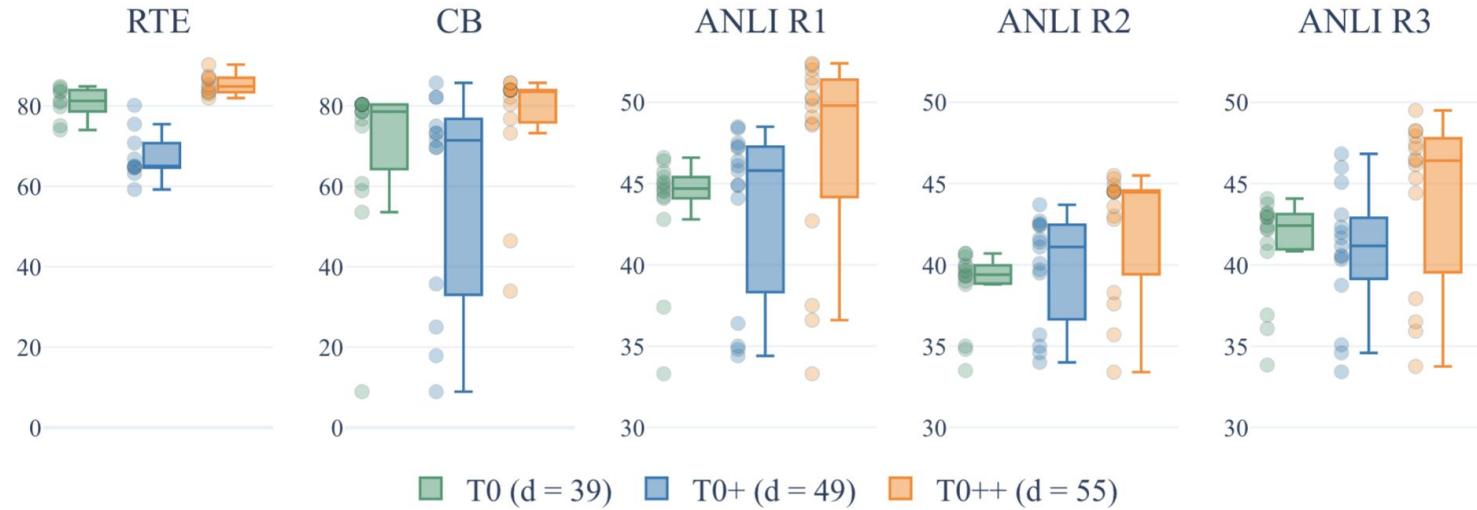


Sentence Completion



■  $p = 0$  (T5+LM) ■  $p = 1$  ■  $p = 5.7$  ■  $p = 8.03$  (T0)

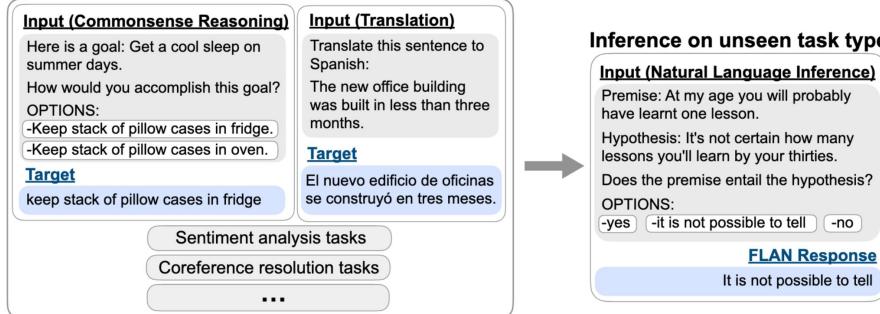
# *Adding datasets (usually) helps*



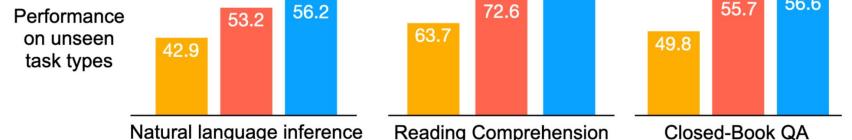
# FINETUNED LANGUAGE MODELS ARE ZERO-SHOT LEARNERS

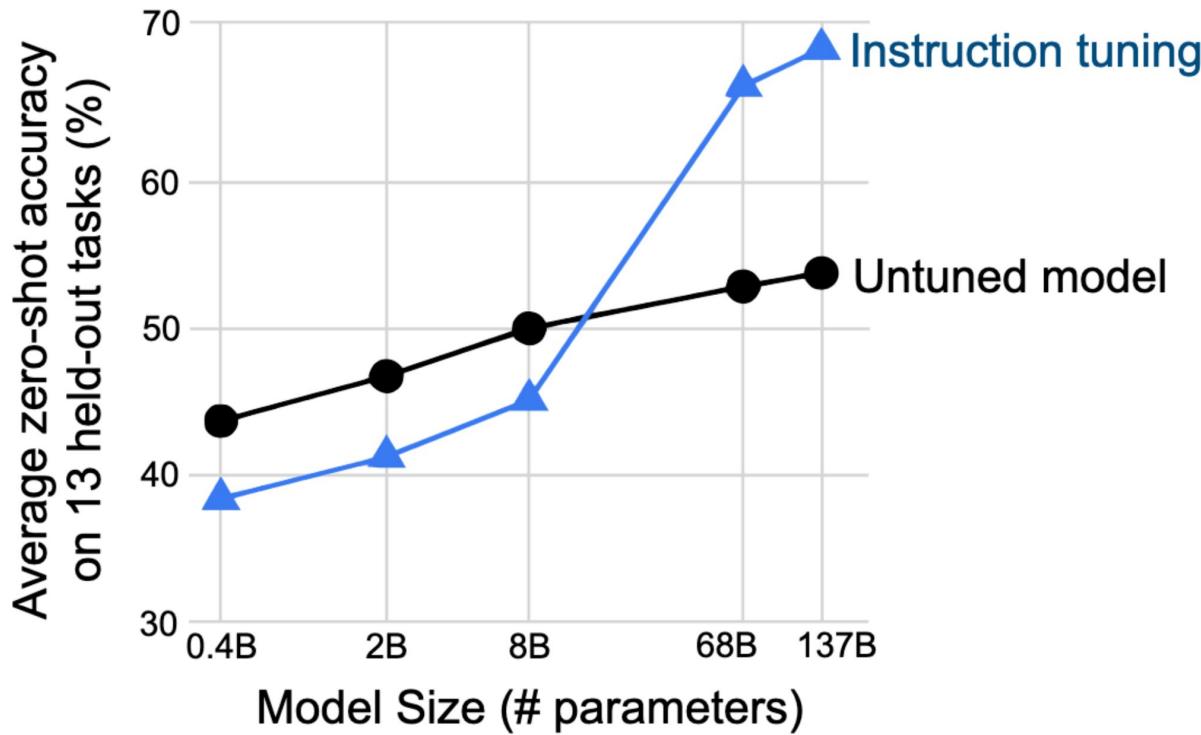
Jason Wei\* Maarten Bosma\* Vincent Y. Zhao\* Kelvin Guu\* Adams Wei Yu  
Brian Lester Nan Du Andrew M. Dai Quoc V. Le  
Google Research

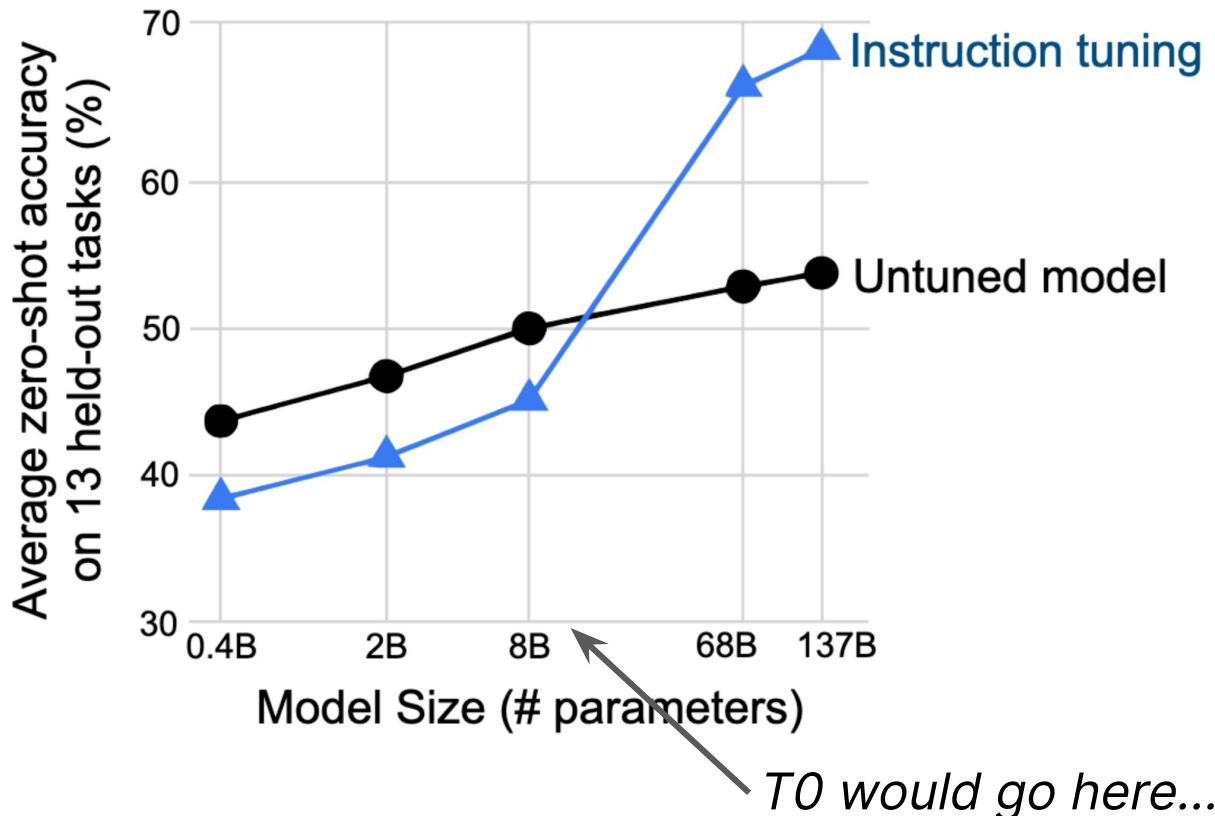
## Finetune on many tasks (“instruction-tuning”)



■ GPT-3 175B zero shot ■ GPT-3 175B few-shot ■ FLAN 137B zero-shot



**B**Performance on held-out tasks

**B**Performance on held-out tasks

	GPT-3	FLAN	T0 (ours)
Size	175B	137B	11B
Multitask supervision	Implicit	Explicit	Explicit
Zero-shot performance	Decent	Better	Better
Architecture	Decoder	Decoder	Encoder/decoder
Pre-training	LM	LM	MLM→LM
Multiple prompts?	No	Yes	Yes
Prompt diversity	N/A	Some	Lots
Public	No	No	Yes

[bigscience-workshop / promptsource](https://github.com/bigscience-workshop/promptsource) Public

Code Issues 13 Pull requests 14 Discussions Actions Projects Wiki Security Insights ...

main 43 branches 1 tag

 VictorSanh Some fixes that slipped through (#717) ...

[bigscience / T0pp](https://github.com/bigscience/T0pp) like 159

Text2Text Generation PyTorch Transformers bigscience/P3 en arxiv:2110.08207 apache-2.0 t5 AutoNLP Compatible

Model card Files and versions Settings Edit model card

How do I pronounce the name of the model? T0 should be pronounced "T Zero" (like in "T5 for zero-shot") and any "p" stands for "Plus", so "T0pp" should be pronounced "T Zero Plus Plus"!

**Model Description**

T0\* shows zero-shot task generalization on English natural language prompts, outperforming GPT-3 on many tasks, while being 16x smaller. It is a series of encoder-decoder models trained on a large set of different tasks specified in natural language prompts. We convert numerous English supervised datasets into prompts, each with multiple templates using varying formulations. These prompted datasets allow for benchmarking the ability of a model to perform completely unseen tasks specified in natural language. To obtain T0\*, we fine-tune a pretrained language model on this multitask mixture covering many different NLP tasks.

**Intended uses**

You can use the models to perform inference on tasks by specifying your query in natural language, and the models will generate a prediction. For instance, you can ask *"Is this review positive or negative?* Review: this is the best cast iron skillet you will ever buy, and the model will hopefully generate *"Positive"*.

Reproduce results and replicate training for T0 (Multitask Prompted Training Enables Zero-Shot Task Generalization)

Downloads last month 6,744

Hosted inference API Examples

Text2Text Generation Is this review positive or negative? Review: Best cast iron skillet you will every buy.

Compute Computation time on cpu: cached

Positive JSON Output Maximize

Spaces using bigscience/T0pp

- akhalqil/T0pp bipin/multipurpose-ai Giuliano/T0 micole66/test
- tjj/t0-generation JonathanLehner/al-msgbot-gpt2-M nlpravi/T0demo

[bigscience-workshop / t-zero](https://github.com/bigscience-workshop/t-zero) Public

Code Issues 2 Pull requests Actions Projects Wiki

master 8 branches 0 tags Go to file

 craffel Fix package\_data path for datasets.csv in setup.py (#15) a2bf4f8 4 days ago 27 commits

 Evaluation in comment 144

 <https://github.com/bigscience-workshop/promptsource>

 <https://github.com/bigscience-workshop/t-zero>

 <https://huggingface.co/bigscience/T0pp>



A one-year long  
research workshop  
on large multilingual  
models and datasets

© Kenneth Norrey 2018

<https://bigscience.huggingface.co/>

# MULTITASK PROMPTED TRAINING ENABLES ZERO-SHOT TASK GENERALIZATION

**Victor Sanh\***

Hugging Face

**Albert Webson\***

Brown University

**Colin Raffel\***

Hugging Face

**Stephen H. Bach\***

Brown University

**Lintang Sutawika**

BigScience

**Zaid Alyafeai**

KFUPM

**Antoine Chaffin**

IRISA & IMATAG

**Arnaud Stiegler**

Hyperscience

**Teven Le Scao**

Hugging Face

**Arun Raja**

I<sup>2</sup>R, Singapore

**Manan Dey**

SAP

**M Saiful Bari**

NTU, Singapore

**Canwen Xu**

UCSD & Hugging Face

**Urmish Thakker**

SambaNova Systems

**Shanya Sharma**

Walmart Labs

**Eliza Szczechla**

BigScience

**Taewoon Kim**

VU Amsterdam

**Gunjan Chhablani**

BigScience

**Nihal V. Nayak**

Brown University

**Debajyoti Datta**

University of Virginia

**Jonathan Chang**

ASUS

**Mike Tian-Jian Jiang**

ZEALS, Japan

**Han Wang**

NYU

**Matteo Manica**

IBM Research

**Sheng Shen**

UC Berkeley

**Zheng-Xin Yong**

Brown University

**Harshit Pandey**

BigScience

**Michael McKenna**

Parity

**Rachel Bawden**

Inria, France

**Thomas Wang**

Inria, France

**Trishala Neeraj**

BigScience

**Jos Rozen**

Naver Labs Europe

**Abheesht Sharma**

BITS Pilani, India

**Andrea Santilli**

University of Rome

**Thibault Fevry**

BigScience

**Jason Alan Fries**

Stanford University

**Ryan Teehan**

Charles River Analytics

**Tali Bers**

Brown University

**Stella Biderman**

Booz Allen Hamilton  
EleutherAI

**Leo Gao**

EleutherAI

**Thomas Wolf**

Hugging Face

**Alexander M. Rush**

Hugging Face

Thanks!

