

# CSCSE 638 Natural Language Processing Foundation and Techniques

## Lecture 12: Large Language Models and Alignment

Kuan-Hao Huang

Spring 2025



(Some slides adapted from Graham Neubig, Jesse Mu, and Hung-Yi Lee)

# Quiz 1

- Average: 92.74
- Median: 95
- Standard deviation: 7.52

TA



Rahul Baid

Email: [rahulbaid@tamu.edu](mailto:rahulbaid@tamu.edu)

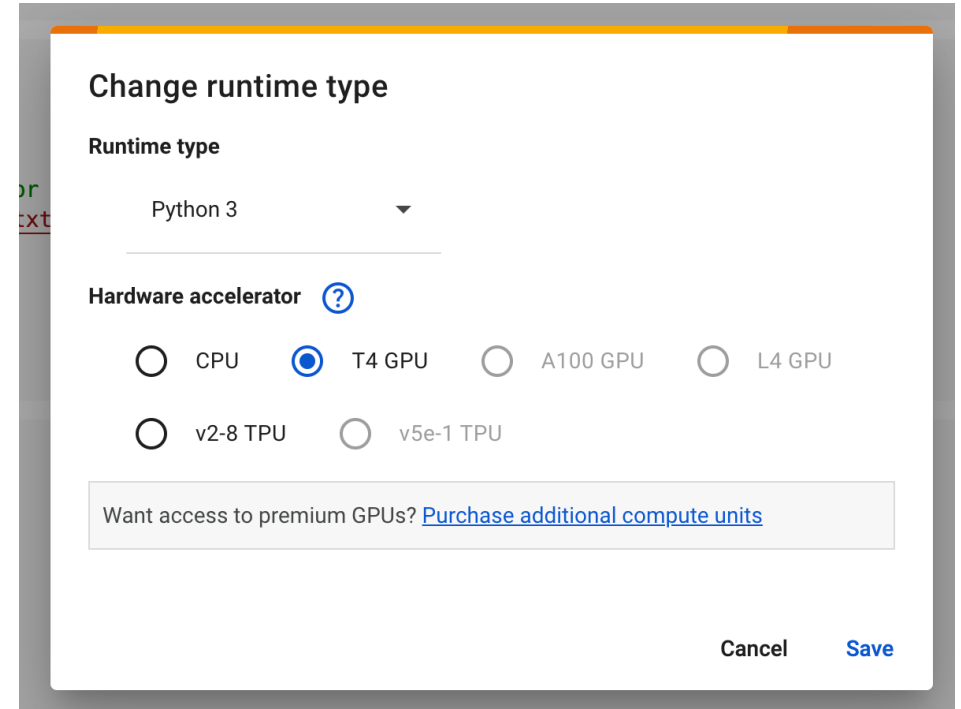
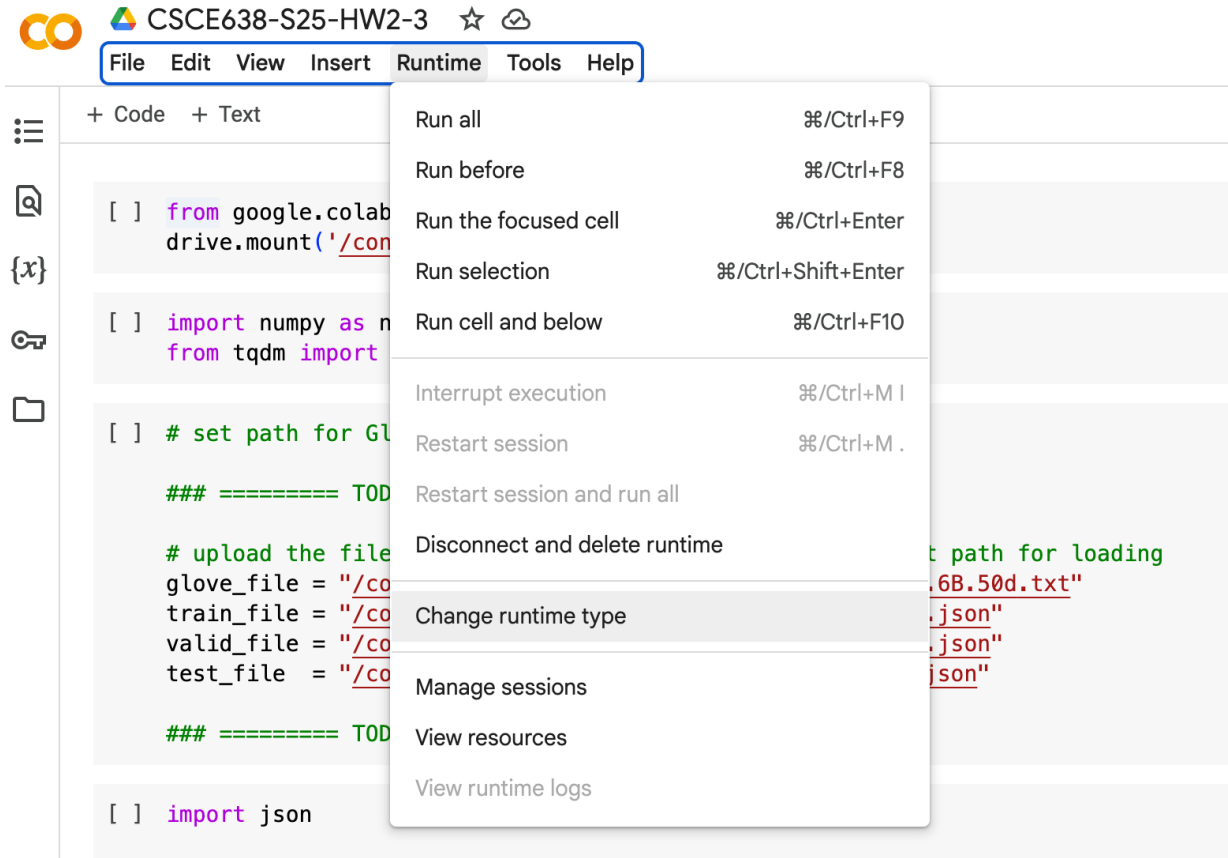
Office Hour: Wed. 12pm – 1pm

Office: PETR 359

# Assignment 2

- [https://khhuang.me/CSCE638-S25/assignments/assignment2\\_0224.pdf](https://khhuang.me/CSCE638-S25/assignments/assignment2_0224.pdf)
- Due: 3/17 11:59pm
- Submit a .zip file to Canvas
  - [submission.pdf](#) for the writing section
  - [submission.py](#) and [submission.ipynb](#) for the coding section
- For questions
  - Discuss on Canvas
  - Send an email to [csce638-ta-25s@list.tamu.edu](mailto:csce638-ta-25s@list.tamu.edu), don't need to CC TA or me

# Assignment 2



Colab has daily quota limit on GPU, so start earlier!

# Course Project – Proposal

- Due: 3/3 11:59pm
- Page limit: 2 pages (excluding references)
- Format: [ACL style](#)
- The proposal should include
  - Introduction to the topic you choose
  - Related literature
  - Novelty and challenges
  - The dataset, models, and approaches you plan to use
  - Evaluation plan

# Course Project: Sign-Up

- <https://docs.google.com/spreadsheets/d/15Rj4AovtHtIzxlLbX1ydrw7lEylamXuV7Dtg7cBD2EU/edit?usp=sharing>
- Please sign up by 2/26
- 3~4 each team

3~4 members per team									
Team	Project Topic	Member 1 (Name)	Member 1 (E-mail)	Member 2 (Name)	Member 2 (E-mail)	Member 3 (Name)	Member 3 (E-mail)	Member 4 (Name)	Member 4 (E-mail)
Team 1									
Team 2									
Team 3	TBD	Wahib Kapdi	wahibkapdi@tamu.edu	Agastya Todi	agastyatodi@tamu.edu	Yash Honrao	yash.honrao@tamu.edu	Shivam Singhal	shivamsinghal@tamu.edu
Team 4	TBD	Sonjoy Paul	skpaul@tamu.edu	Mabon Ninan	ninanmm@tamu.edu	Ashwini Ravindran	ashwinir@tamu.edu		
Team 5	TBD	Raj Purohith Arjun	raj2001@tamu.edu	Venkateswarlu Nagineni	venkates2002@tamu.edu	Shuvam Chowdhury	schowdhury@tamu.edu	Jeffrey Kevin Joseph	jeffrey98@tamu.edu
Team 6	TBD	Prakhar Suryavansh	ps41@tamu.edu	Rusali Saha	rs0921@tamu.edu	Priyal Khapra	priyalkhapra@tamu.edu		
Team 7	TBD	Chuan-Hsin Wang	chuanhsin0110@tamu.edu	Wei-Chien Cheng	wcheng@tamu.edu	Chi-Ming Lee	chiminglee831@tamu.edu		
Team 8	TBD	Afreen Ahmed	afreen04@tamu.edu	Rhea Sudheer	rheasudheer19@tamu.edu	Hitha Magadi Vijayanand	hoshi_1996@tamu.edu	Sai Aakarsh Padma	saiakarsh@tamu.edu
Team 9	Multilingual Video Grounding: Cross-Language Temporal Localization	Ramana Heggadal Math	ramana_hm@tamu.edu	Ruthvik Kanumuri	kruthvik007@tamu.edu	Jnana Preeti Parlapalli	pj.preeti@tamu.edu	Shravan Conjeevaram	shravan10@tamu.edu
Team 10	TBD	Harshavardhana	asharsha30@tamu.edu	Rucha Ravindra Gole	ruchagole16@tamu.edu	Shashank Santosh Jagtap	shashankjagtap@tamu.edu		
Team 11	TBD	Logan Bibb	logan.bibb@tamu.edu	Daniel Ortiz-Chaves	dortizchaves@tamu.edu	Sicong Liang	lsc206573@tamu.edu		
Team 12	TBD	Yifan Ren	yfren@tamu.edu	Qinyao Hou	yaoya2618@tamu.edu	Caroline Li	zhiheng@tamu.edu		
Team 13	TBD	Tien-Hung Hsiao	th.hsiao@tamu.edu	Barry Liu	barry89130663@gmail.com	Hsueh-chien Chao	alanchao8669@tamu.edu		
Team 14	TBD	Esben Egholm	esbenegholm@tamu.edu	Michael Norman	michael.norman@tamu.edu	Davran Damkhan	davrandamkhan@tamu.edu		
Team 15	In-Context Learning with LLMs	Dheeraj Mudireddy	dheeraj.reddy@tamu.edu	Ninad Deo	ninzo_05@tamu.edu	Dhruvraj Singh Rathore	dhruvraj_16@tamu.edu	Atharva Phand	ahphand@tamu.edu
Team 16	TBD	Tejashri K	tkelhe@tamu.edu	Sukanya Sahoo	sukanya.sahoo@tamu.edu	Ramneek Kaur	ramneek983@tamu.edu	Saksham Mehta	saksham19@tamu.edu
Team 17	TBD	Arnav Jain	arnavkj11@tamu.edu	Parangjothi	parangjothi.c@tamu.edu	Medha Majumdar	medhamajumdar1@tamu.edu		
Team 18	TBD	Adarsh Kumar	adarsh0801@tamu.edu	Neil Roy	neilroy@tamu.edu	Hwiyeon Kim	hwiyeonkim@tamu.edu	Jawahar Sai Nathani	jawaharsainathani@tamu.edu
Team 19	TBD	Satvik Praveen	satvikpraveen_164@tamu.edu	Jonathan Tong	tongjo@tamu.edu	Vinay Chandra Bandi	vinaychandra@tamu.edu	Yamini Preethi Kamisetty	yamini_preethi_k@tamu.edu
Team 20	TBD	Piyush Sharan	pisharan@tamu.edu	Manisha Panda	mpanda27@tamu.edu	Abhishek Singh	abhi_singh@tamu.edu	Jaydeep Radadiya	jd@tamu.edu
Team 21	TBD	Yamini Preethi Kamisetty	yamini_preethi_k@tamu.edu	Vinay Chandra Bandi					
Team 22	Jaillbreaking LLMs using Graph of Thought	Aayush Upadhyay	aaupadhy@tamu.edu	Anant Mehta	anant_mehta@tamu.edu	Ajay Jagannath	ajayjagan2511@tamu.edu		
Team 23	TBD	Dishant Parag Zaveri	dishant.zaveri@tamu.edu	Saransh Agrawal	saransh.agrawal@tamu.edu	Faizan Ali Khaji	khajifaizanali@tamu.edu	Pavan Santosh	pavan_santosh@tamu.edu
Team 24	TBD	Bitia Malekianboroujeni	Bitia.malekian@tamu.edu	Kimia Mirhosseini	kimia1379@tamu.edu	Maddhurima Mondal	mmkpa2012@tamu.edu		
Team 25									

# Course Project: Project Highlight

- Date: 3/5 in person
- Each team has 3 minutes to introduce the project
  - Introduction to the topic you choose
  - Short related literature overview
  - Novelty and challenges
  - The dataset, models, and approaches you plan to use
  - Evaluation plan

# Lecture Plan

- Large Language Models
  - Prompting
  - In-Context Learning
  - Chain-of-Thought Prompting
- Evaluation of Large Language Models
- Alignment
  - Instruction Tuning
  - Human Preference Optimization





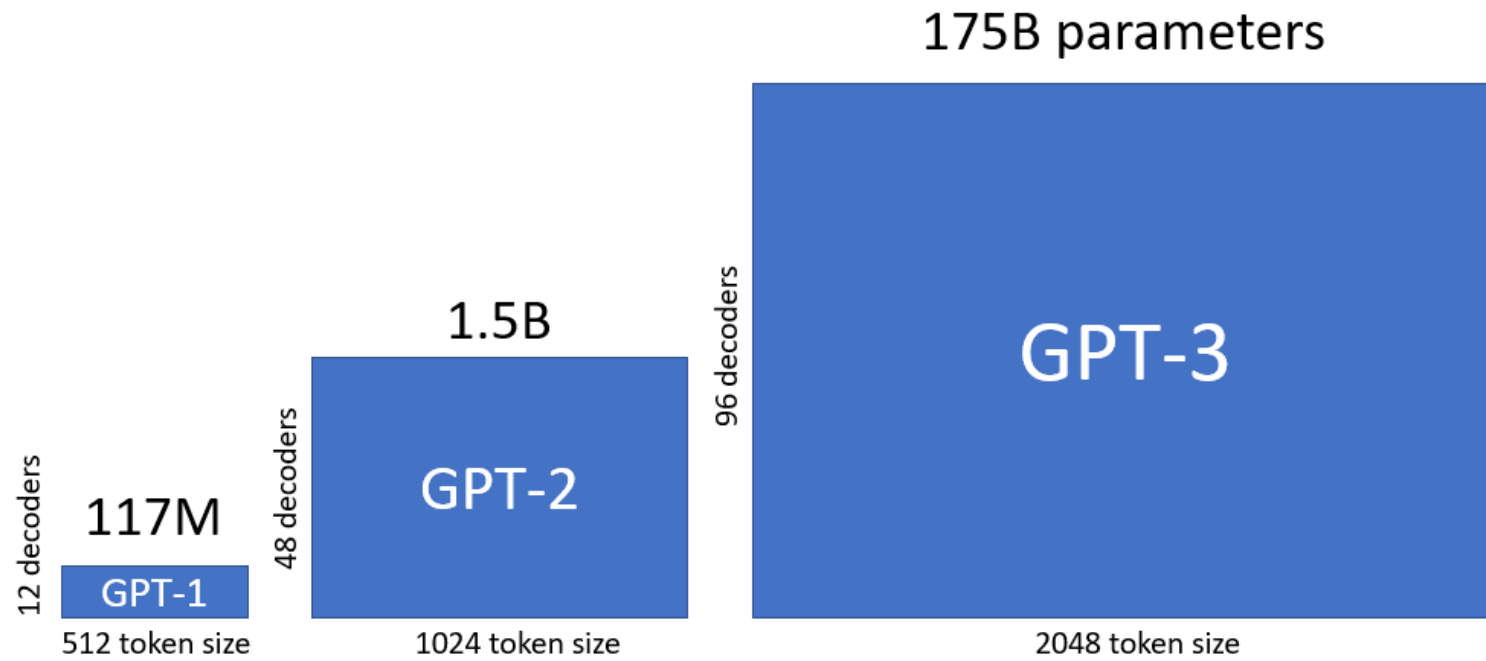
# Large Language Models (LLMs)



**ANTHROPIC**



# Scaling Is The Key



# Zero-Shot Prompting

- Prompt → Completion
  - Continue writing

Prompt

This place is incredible! The lobster is the best I've ever had. The sentiment of the above sentence is

positive.

Completion

# Zero-Shot Prompting

- Prompt → Completion
  - Continue writing

Prompt

Stephen Curry's clutch barrage seals another Olympic gold for USA. The topic of the above sentence is

sport.

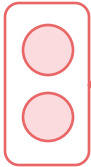
Completion

# Classification with [MASK] Embedding and Prompt

## Topic Classification

The Houston Rockets won an intense overtime game	Sports
Bitcoin hit a new all-time high this week	Finance
Tesla launched a new self-driving software update	Technology
Flu cases are rising in several major cities	Health

Classification with [MASK] embedding

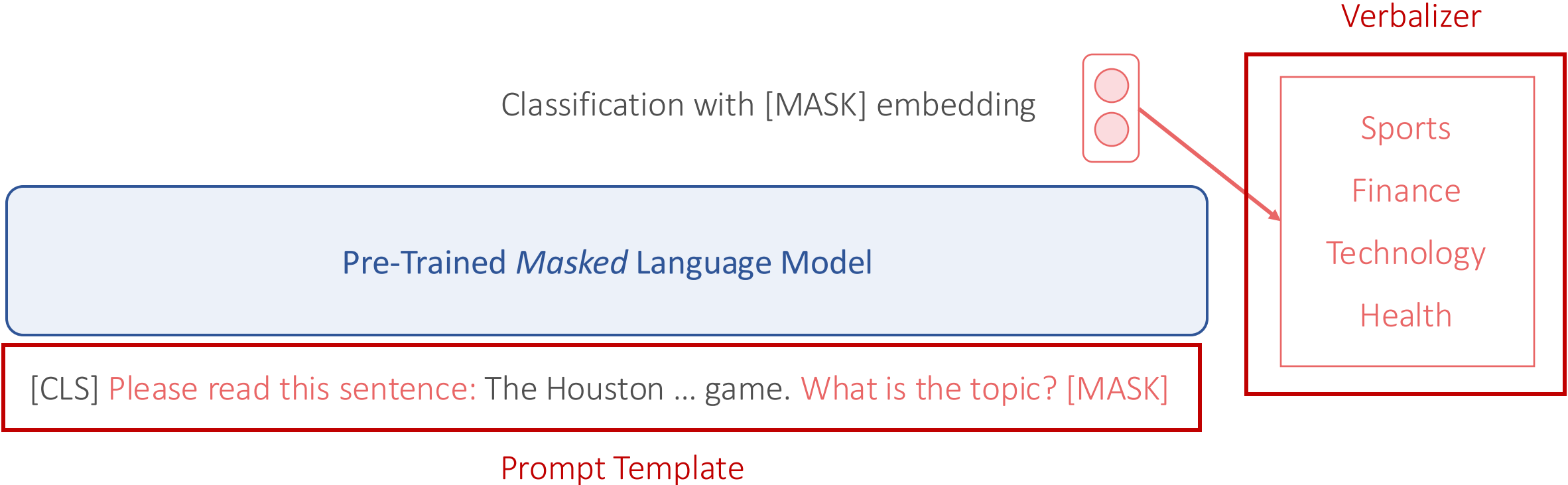


Pre-Trained *Masked* Language Model

- Sports
- Finance
- Technology
- Health

[CLS] The Houston Rockets won an ... overtime game. What is the topic? [MASK]

# Prompt Tuning



# Language Modeling

Binge ... on | - | and | of | is

Binge **drinking** ... is | and | had | in | was

Binge drinking **may** ... be | also | have | not | increase

Binge drinking may **not** ... be | have | cause | always | help

Binge drinking may not **necessarily** ... be | lead | cause | results | have

Binge drinking may not necessarily **kill** ... you | the | a | people | your

Binge drinking may not necessarily kill **or** ... even | injure | kill | cause | prevent

Binge drinking may not necessarily kill or **even** ... kill | prevent | cause | reduce | injure

Binge drinking may not necessarily kill or even **damage** ... your | the | a | you | someone

Binge drinking may not necessarily kill or even damage **brain** ... cells | functions | tissue | neurons

Binge drinking may not necessarily kill or even damage brain **cells,** ... some | it | the | is | long



# Zero-Shot Prompting

Prompt

This place is incredible! The lobster is the best I've ever had. The sentiment of the above sentence is

positive.

Completion

Prompt

Stephen Curry's clutch barrage seals another Olympic gold for USA. The topic of the above sentence is

sport.

Completion

# A New Way to Use NLP Models

- Task-specific features + task-specific model
- General embeddings + task-specific model
- General embeddings + general model + task-specific fine-tuning
- General embeddings + general model + **task-specific prompting**

# Zero-Shot Prompting

Prompt

This place is incredible! The lobster is the best I've ever had. The sentiment of the above sentence is

positive.

Completion

Prompt

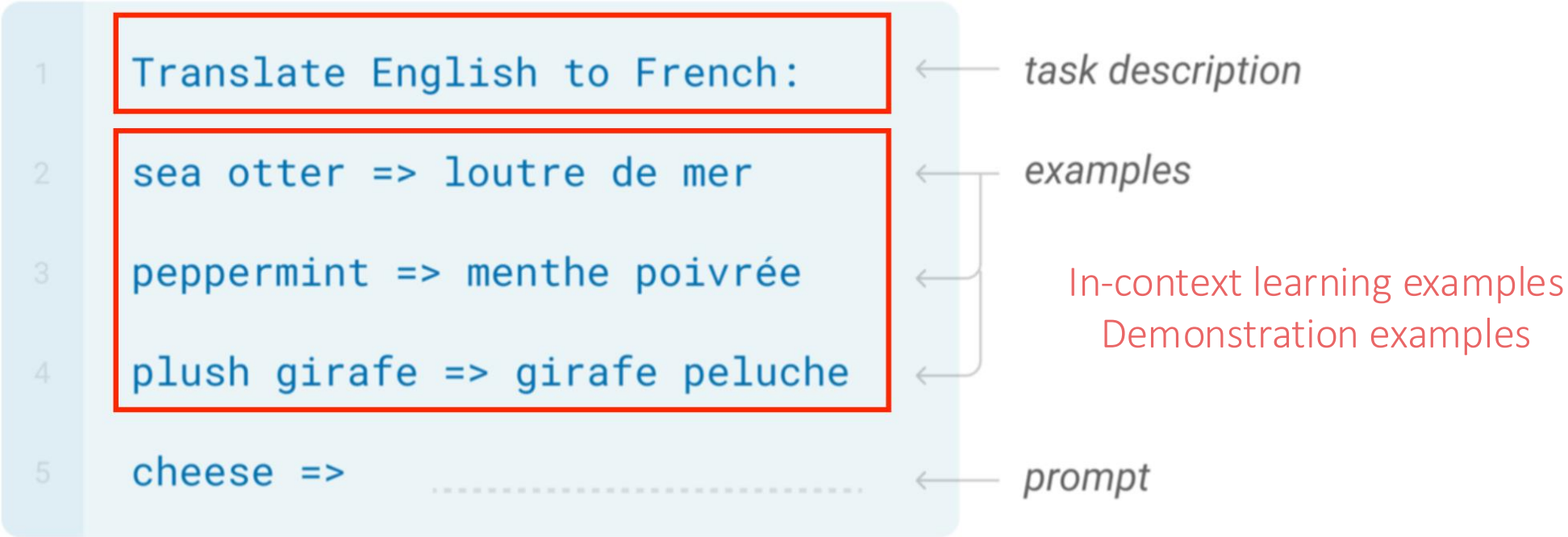
Stephen Curry's clutch barrage seals another Olympic gold for USA. The topic of the above sentence is

sport.

Completion

Any Issues?

# Few-Shot Prompting / In-Context Learning



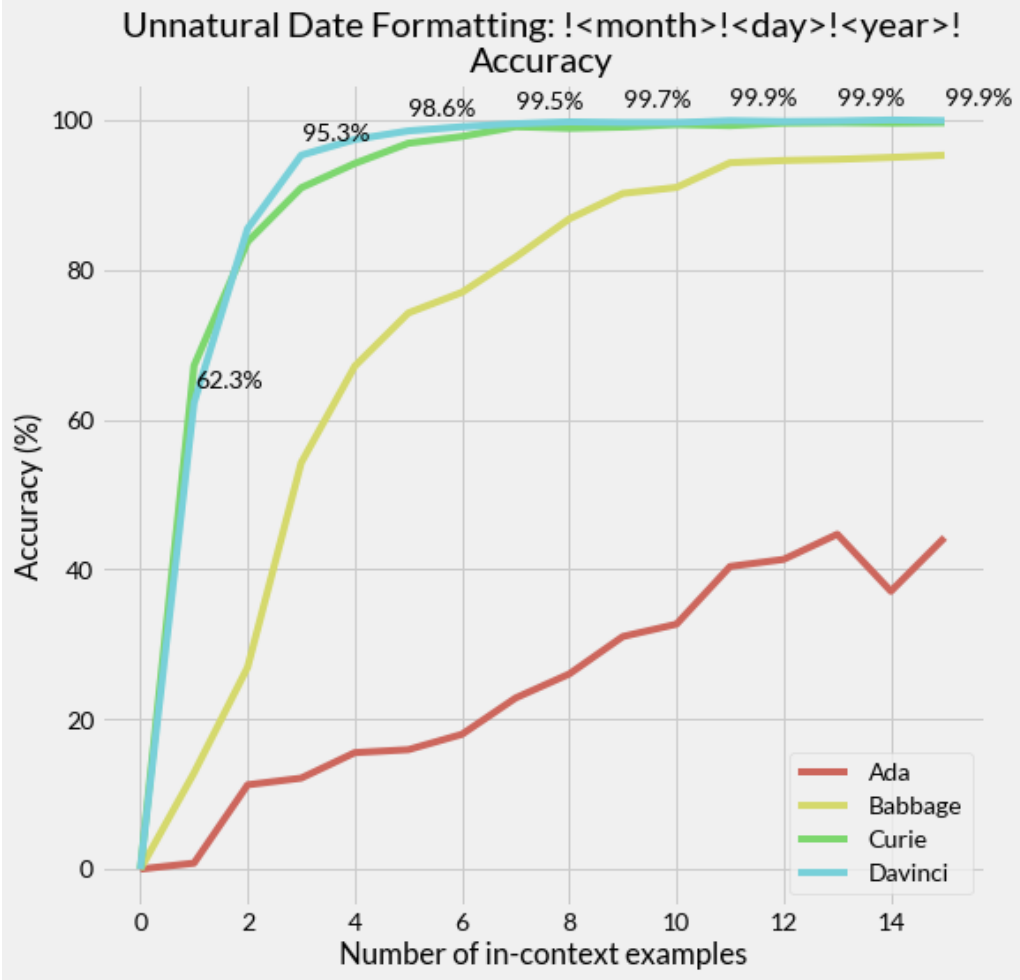
# Few-Shot Prompting / In-Context Learning

Input: 2014-06-01  
Output: !06!01!2014!  
Input: 2007-12-13  
Output: !12!13!2007!  
Input: 2010-09-23  
Output: !09!23!2010!  
Input: **2005-07-23**  
Output: **!07!23!2005!**

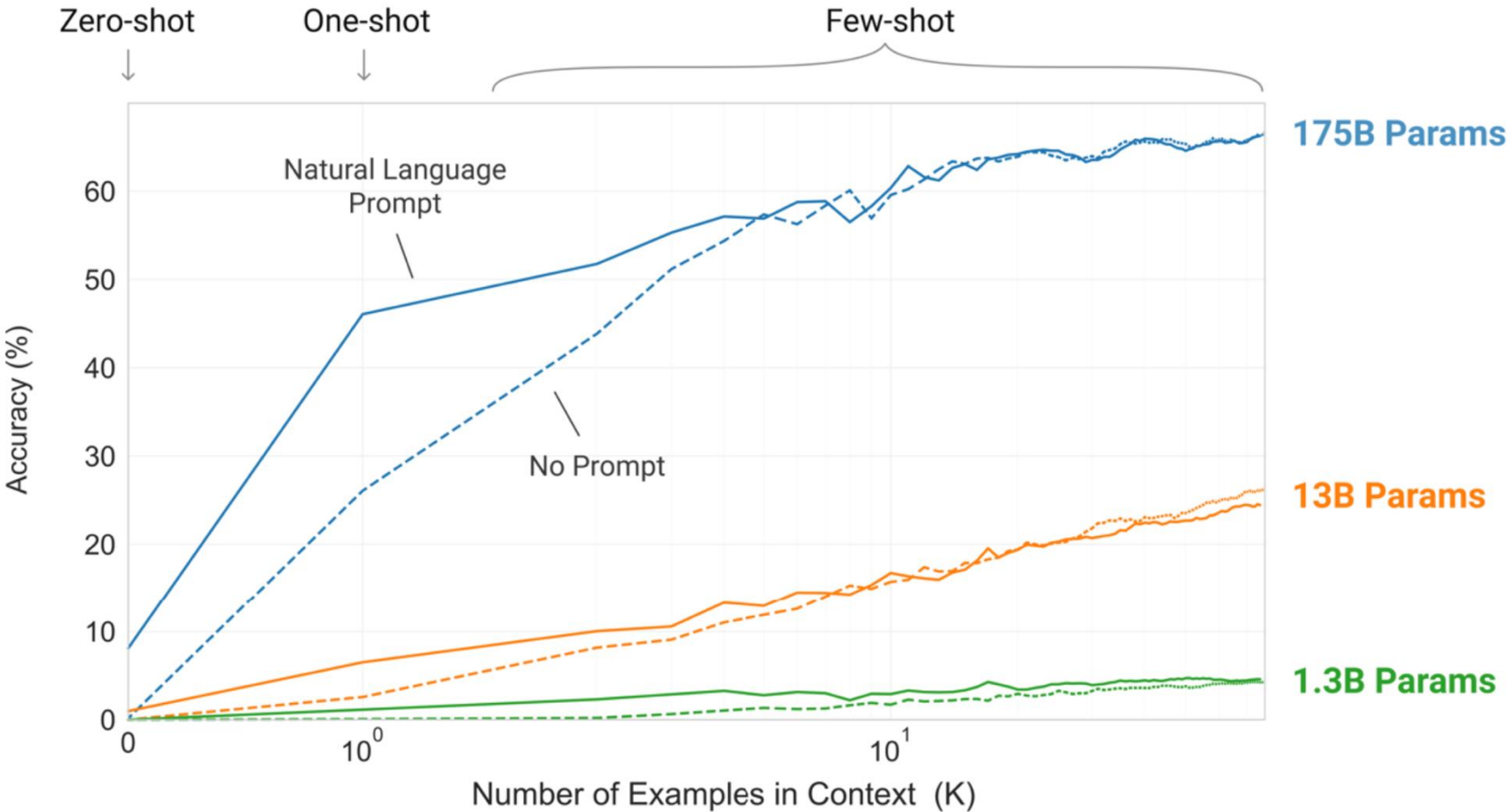
*in-context examples*

*test example*

*model completion*



# Few-Shot Prompting / In-Context Learning



# What Makes In-Context Learning Work?

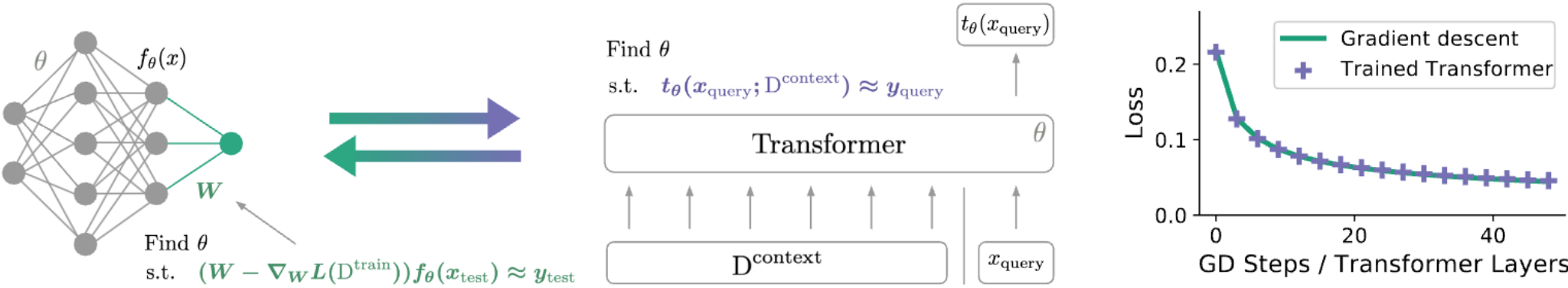
- Still an open research problem

---

## Transformers Learn In-Context by Gradient Descent

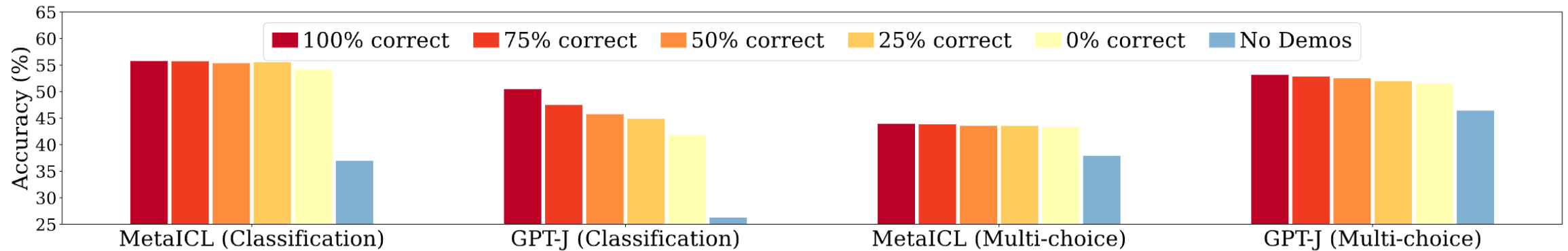
---

Johannes von Oswald<sup>1,2</sup> Eyvind Niklasson<sup>2</sup> Ettore Randazzo<sup>2</sup> João Sacramento<sup>1</sup>  
Alexander Mordvintsev<sup>2</sup> Andrey Zhmoginov<sup>2</sup> Max Vladymyrov<sup>2</sup>



# What Makes In-Context Learning Work?

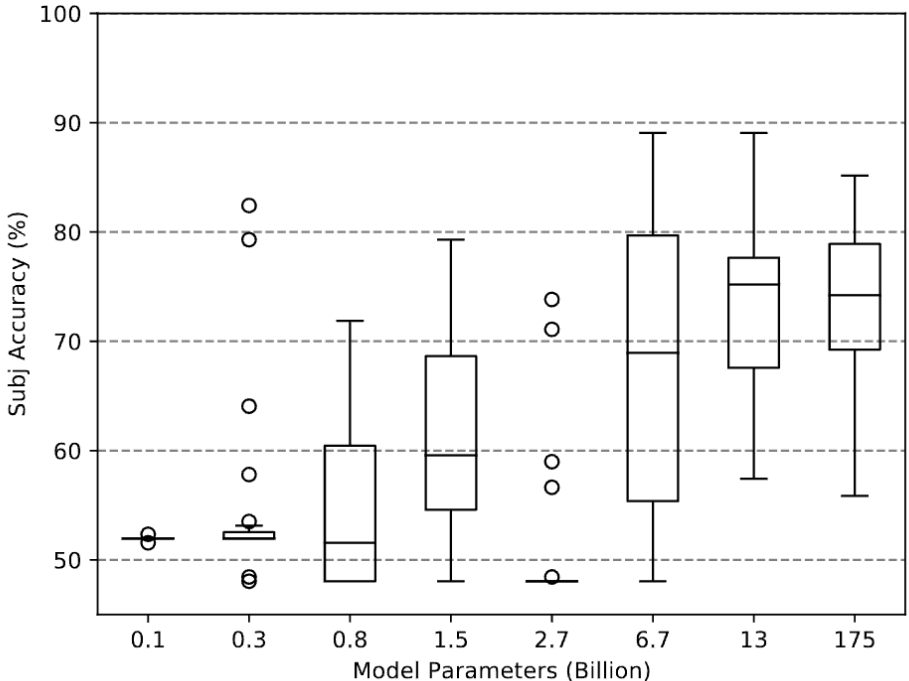
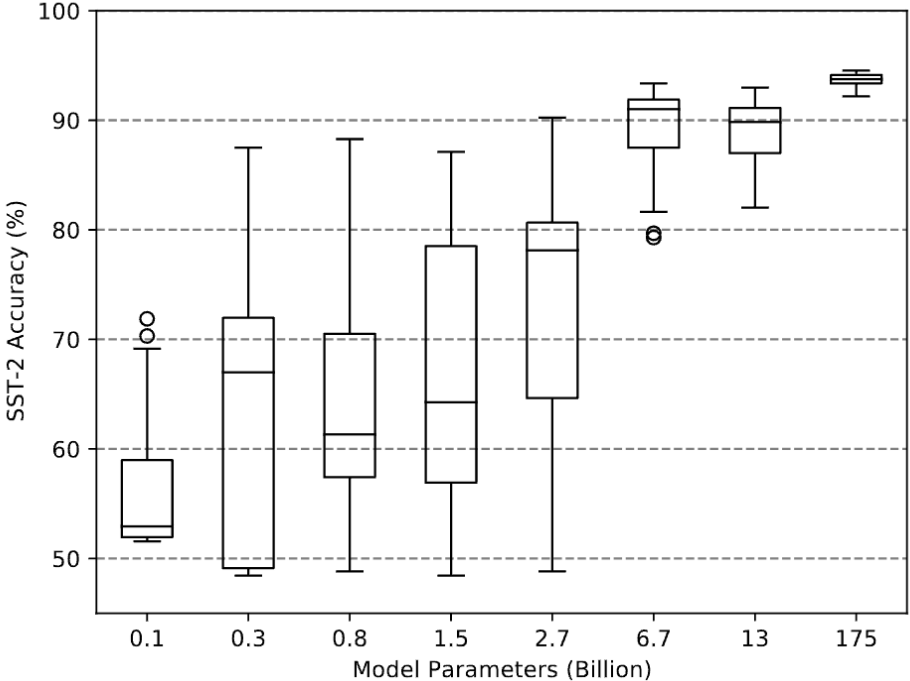
- Provide information more about format?
  - Give wrong in-context learning examples





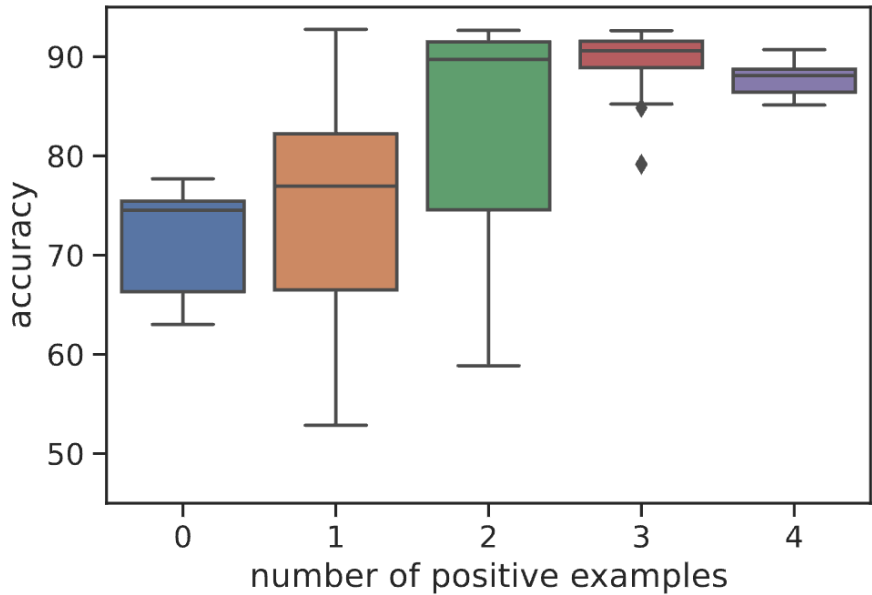
# LLMs are Sensitive to Small Changes in In-context Examples

Changing order of examples causes large variance in performance

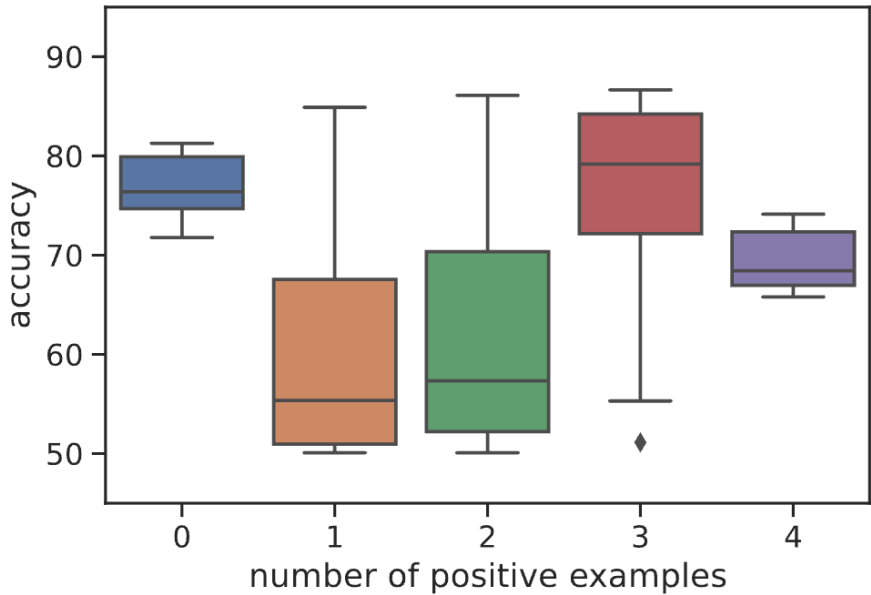


# LLMs are Sensitive to Small Changes in In-context Examples

Label balance causes large variance in performance



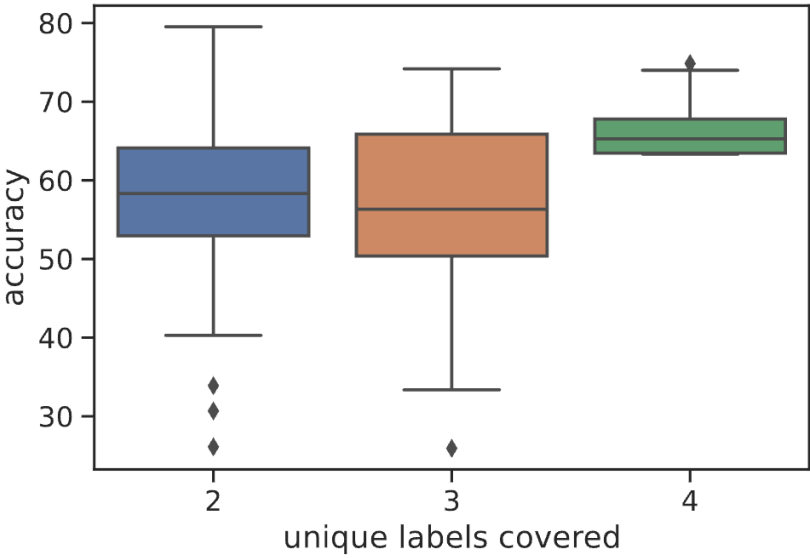
(a) Amazon



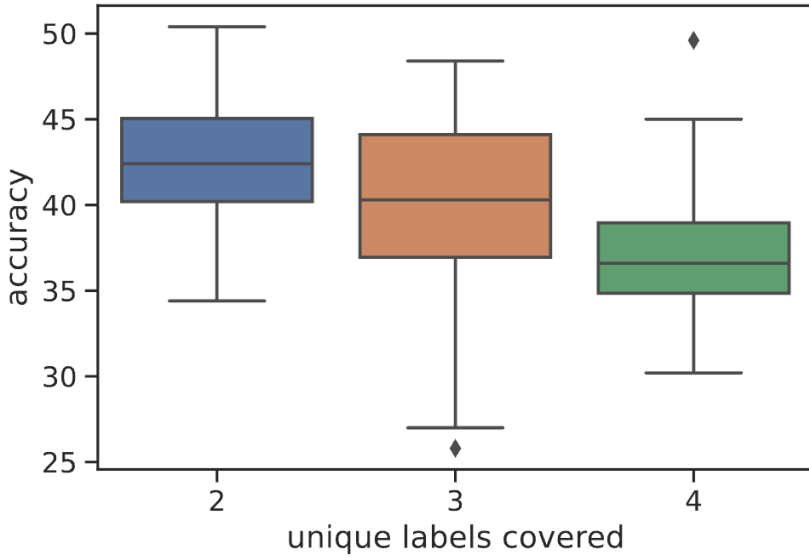
(b) SST-2

# LLMs are Sensitive to Small Changes in In-context Examples

Label coverage causes large variance in performance



(a) AGNews



(b) TREC

# How to Choose In-Context Learning Examples?

- Similarity?
- Coverage?
- Length?
- Confidence?
- Principles?

## Learning Principles (LEAP; this work)

### Generating Mistakes Intentionally (performed once)

Q: Evaluate  $\log_{1/3} 9$

A: To evaluate  $\log_{1/3} 9$ , you want to find the exponent to which  $\frac{1}{3}$  must be raised to get 9 as the result.

...

$$\frac{1^x}{(1/3)^x} = \frac{3^2}{1}$$

So  $\log_{1/3} 9 = \boxed{2}$



### Learning Principles (performed once)

It is crucial to accurately calculate the powers of the base in logarithm problems to correctly determine the exponent. Additionally understanding the definition of a logarithm is key: the logarithm base b of a number x is the exponent to which b must be raised to get x.

### Inference with Learned Principles:

A: First we can write  $125\sqrt{5}$  as  $5^3 \cdot 5^{1/2} = 5^{7/2}$ .

The logarithm base  $\sqrt{5}$  of  $5^{7/2}$  is asking to which power we must raise  $\sqrt{5}$  to get  $5^{7/2}$ .

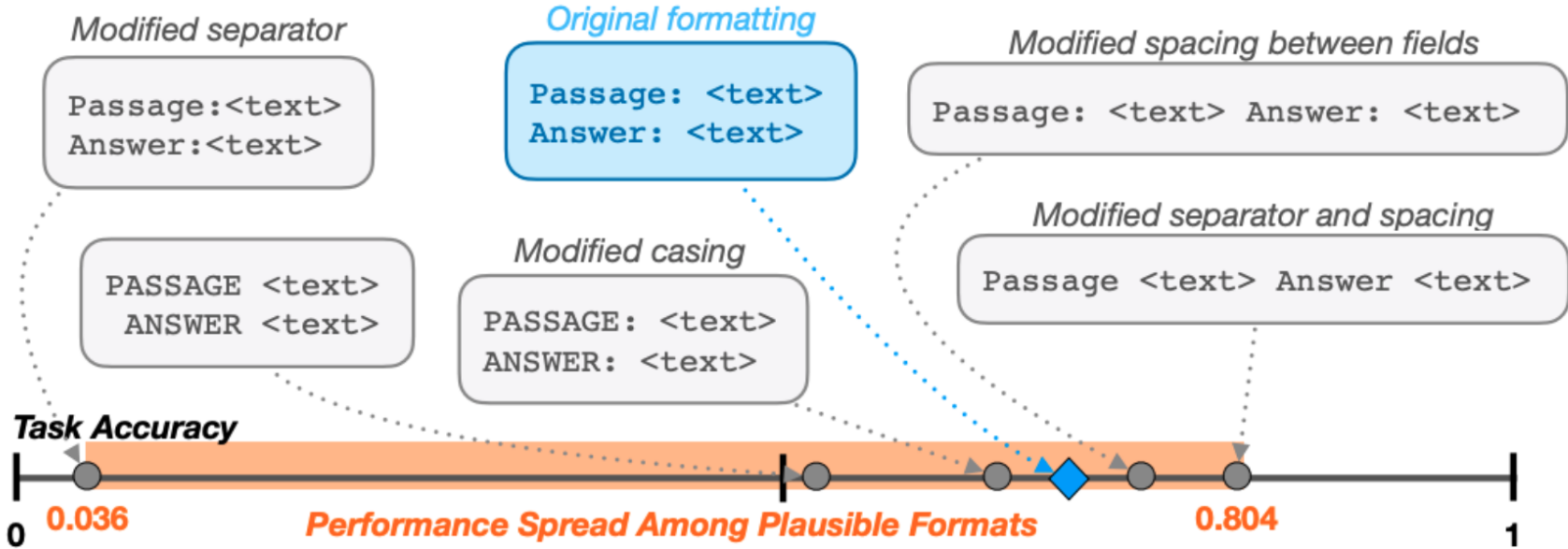
Since  $\sqrt{5} = 5^{1/2}$ , we can see that  $(5^{1/2})^7 = 5^{7/2}$ .

Therefore,  $\log_{\sqrt{5}} 125\sqrt{5} = \boxed{7}$



# Prompt Engineering

- Search for and design better prompts



# Chain-of-Thought (CoT) Prompting

- Ask the model to explain its reasoning before making an answer

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 **X**

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. ✓*

# Chain-of-Thought (CoT) Prompting

- Ask the model to explain its reasoning before making an answer

	Arithmetic					
	SingleEq	AddSub	MultiArith	GSM8K	AQUA	SVAMP
zero-shot	74.6/78.7	<b>72.2/77.0</b>	17.7/22.7	10.4/12.5	22.4/22.4	58.8/58.7
zero-shot-cot	<b>78.0/78.7</b>	69.6/74.7	<b>78.7/79.3</b>	<b>40.7/40.5</b>	<b>33.5/31.9</b>	<b>62.1/63.7</b>
	Common Sense		Other Reasoning Tasks		Symbolic Reasoning	
	Common SenseQA	Strategy QA	Date Understand	Shuffled Objects	Last Letter (4 words)	Coin Flip (4 times)
zero-shot	<b>68.8/72.6</b>	12.7/54.3	49.3/33.6	31.3/29.7	0.2/-	12.8/53.8
zero-shot-cot	64.6/64.0	<b>54.8/52.3</b>	<b>67.5/61.8</b>	<b>52.4/52.9</b>	<b>57.6/-</b>	<b>91.4/87.8</b>

# Chain-of-Thought (CoT) Prompting

- Ask the model to explain its reasoning before making an answer

	Arithmetic					
	SingleEq	AddSub	MultiArith	GSM8K	AQUA	SVAMP
zero-shot	74.6/78.7	<b>72.2/77.0</b>	17.7/22.7	10.4/12.5	22.4/22.4	58.8/58.7
zero-shot-cot	<b>78.0/78.7</b>	69.6/74.7	<b>78.7/79.3</b>	<b>40.7/40.5</b>	<b>33.5/31.9</b>	<b>62.1/63.7</b>
	Common Sense		Other Reasoning Tasks		Symbolic Reasoning	
	Common SenseQA	Strategy QA	Date Understand	Shuffled Objects	Last Letter (4 words)	Coin Flip (4 times)
zero-shot	<b>68.8/72.6</b>	12.7/54.3	49.3/33.6	31.3/29.7	0.2/-	12.8/53.8
zero-shot-cot	64.6/64.0	<b>54.8/52.3</b>	<b>67.5/61.8</b>	<b>52.4/52.9</b>	<b>57.6/-</b>	<b>91.4/87.8</b>



# Few-Shot Chain-of-Thought Prompting

(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. ✓

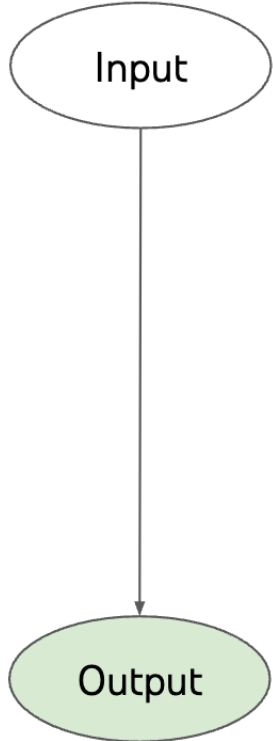
# Few-Shot Chain-of-Thought Prompting

	MultiArith	GSM8K
<b>Zero-Shot</b>	<b>17.7</b>	<b>10.4</b>
Few-Shot (2 samples)	33.7	15.6
Few-Shot (8 samples)	33.8	15.6
<b>Zero-Shot-CoT</b>	<b>78.7</b>	<b>40.7</b>
Few-Shot-CoT (2 samples)	84.8	41.3
Few-Shot-CoT (4 samples : First) (*1)	89.2	-
Few-Shot-CoT (4 samples : Second) (*1)	90.5	-
Few-Shot-CoT (8 samples)	93.0	48.7
<b>Zero-Plus-Few-Shot-CoT (8 samples) (*2)</b>	<b>92.8</b>	<b>51.5</b>
Finetuned GPT-3 175B [Wei et al., 2022]	-	33
Finetuned GPT-3 175B + verifier [Wei et al., 2022]	-	55
<b>PaLM 540B: Zero-Shot</b>	<b>25.5</b>	<b>12.5</b>
<b>PaLM 540B: Zero-Shot-CoT</b>	<b>66.1</b>	<b>43.0</b>
<b>PaLM 540B: Zero-Shot-CoT + self consistency</b>	<b>89.0</b>	<b>70.1</b>
PaLM 540B: Few-Shot [Wei et al., 2022]	-	17.9
PaLM 540B: Few-Shot-CoT [Wei et al., 2022]	-	56.9
PaLM 540B: Few-Shot-CoT + self consistency [Wang et al., 2022]	-	74.4

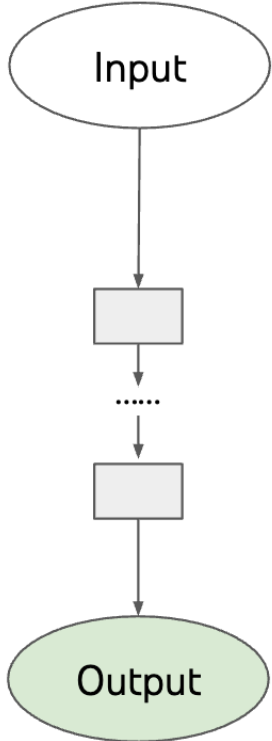
# What Makes Chain-of-Thought Work?

- Explicit reasoning steps
- Knowledge expansion
- Possibility to refine answers

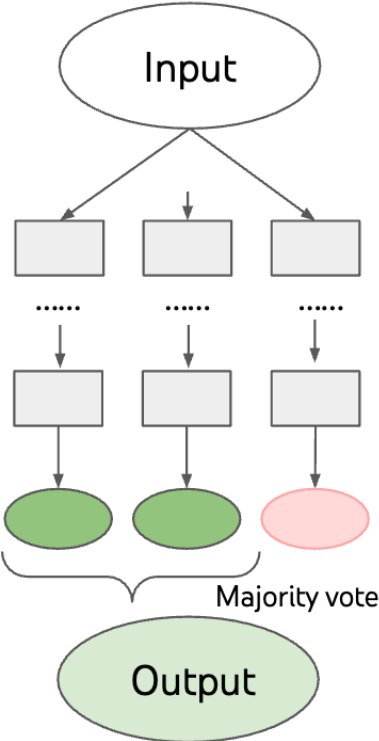
# Tree-of-Thoughts



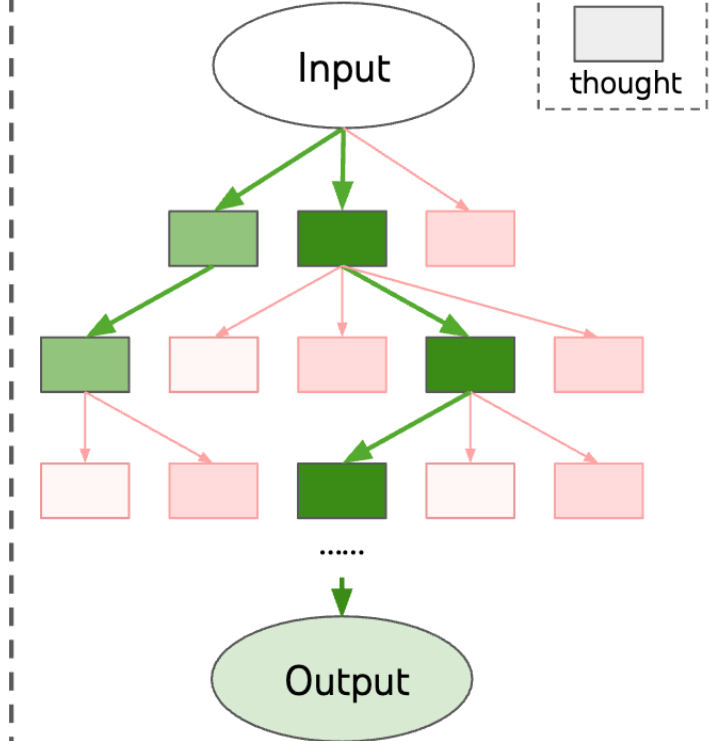
(a) Input-Output Prompting (IO)



(c) Chain of Thought Prompting (CoT)

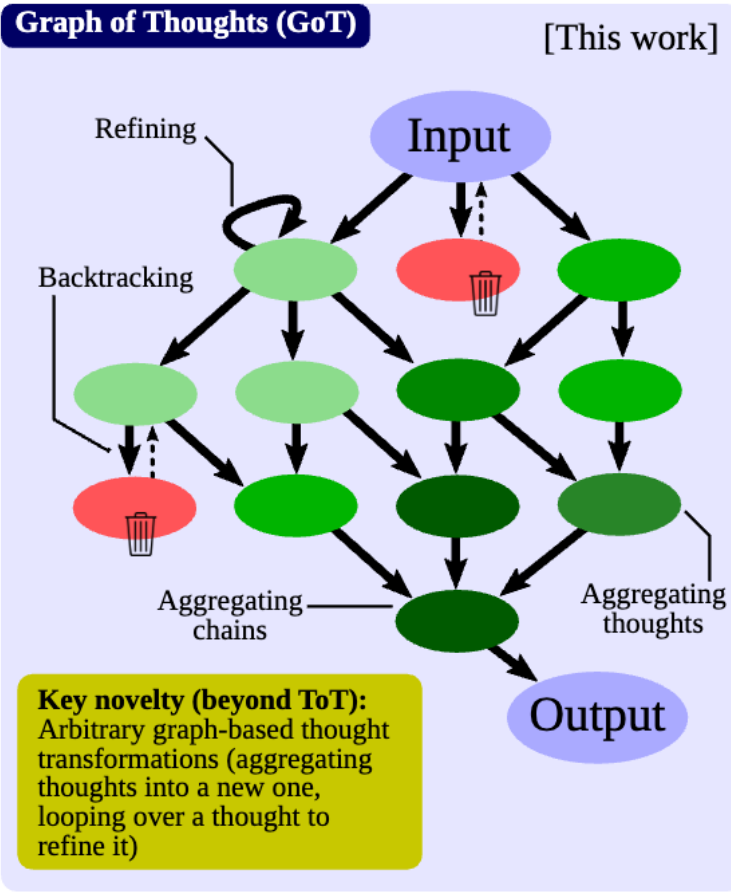
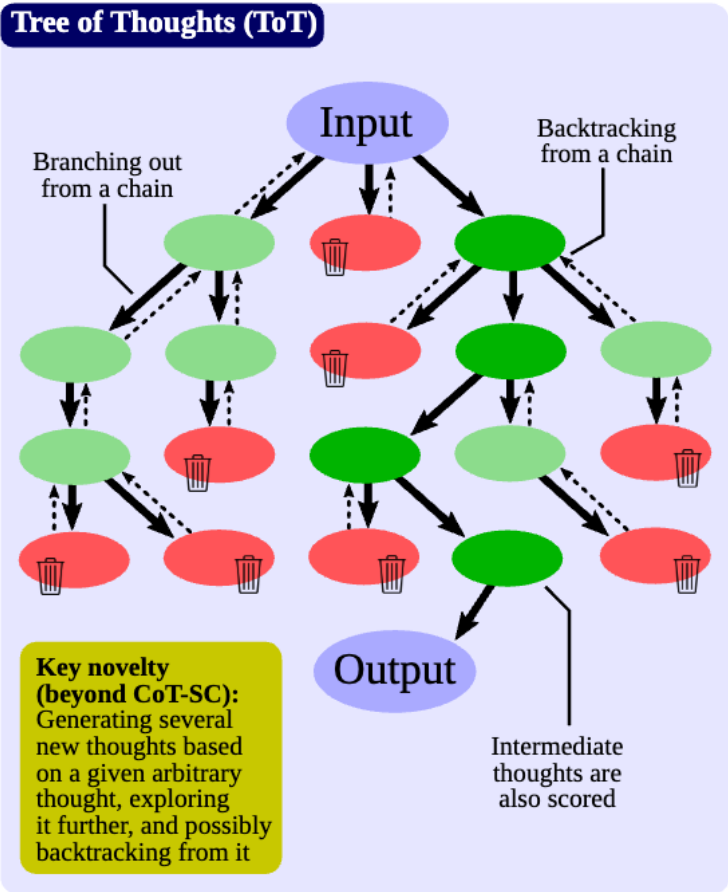
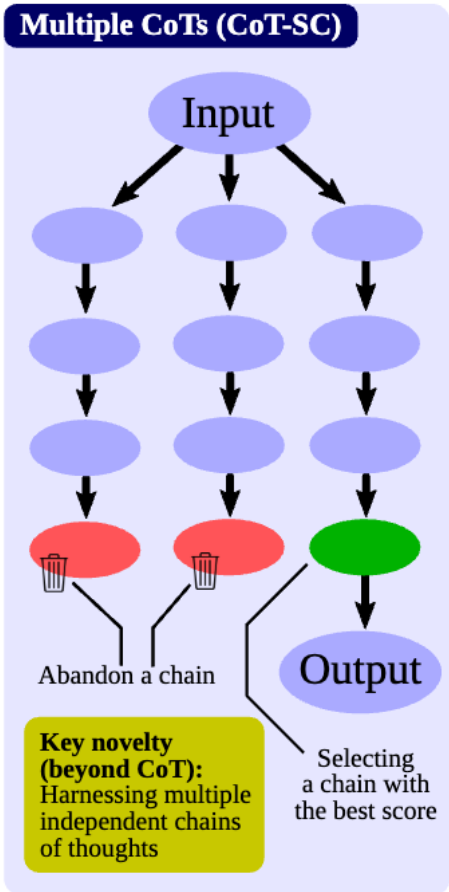
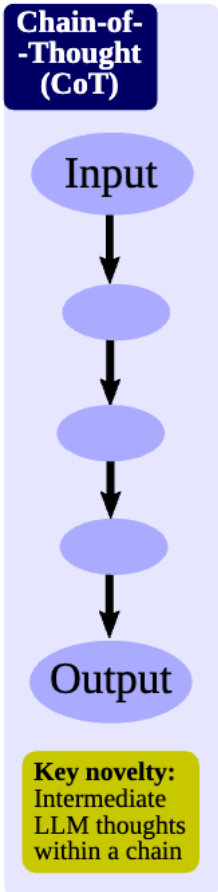


(c) Self Consistency with CoT (CoT-SC)

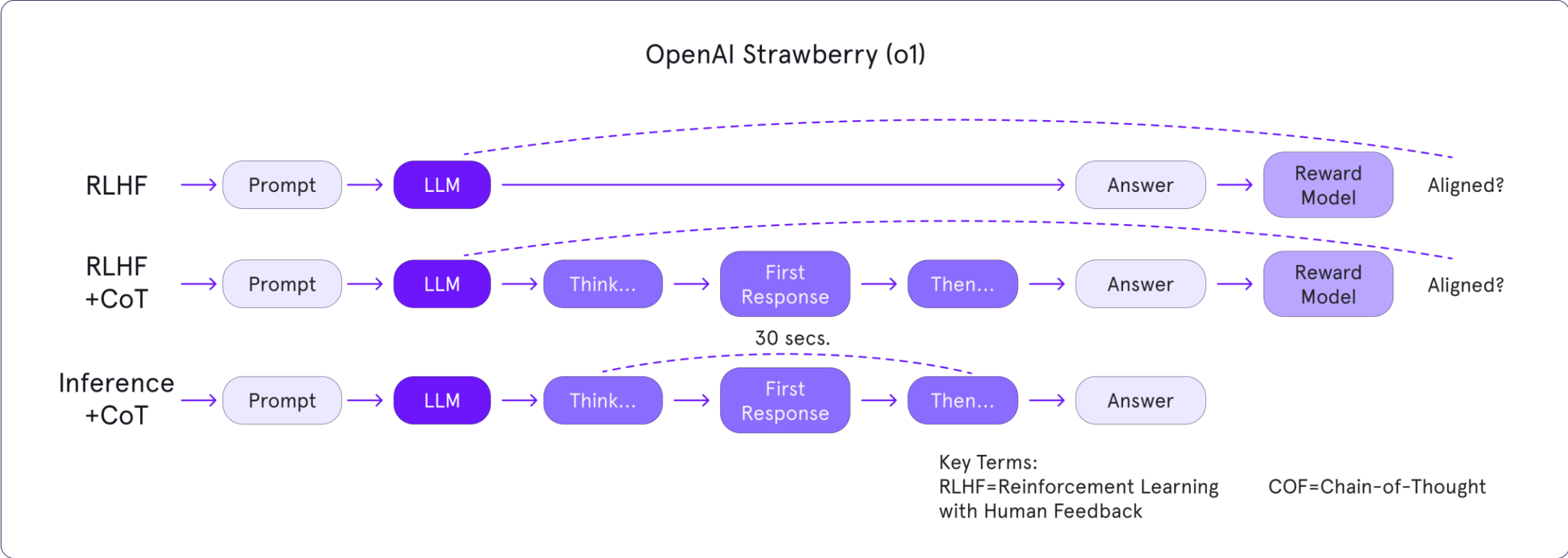


**(d) Tree of Thoughts (ToT)**

# Graph-of-Thoughts



# OpenAI o1



# Do LLMs Know They are Wrong?

- It's hard to say, but most of the time yes
  - Verifying if easier than generating
- Correction-based and refinement-based approaches

## Training Language Models to Self-Correct via Reinforcement Learning

Aviral Kumar<sup>\*,+,1</sup>, Vincent Zhuang<sup>\*,+,1</sup>, Rishabh Agarwal<sup>\*,1</sup>, Yi Su<sup>\*,1</sup>, JD Co-Reyes<sup>1</sup>, Avi Singh<sup>1</sup>, Kate Baumli<sup>1</sup>, Shariq Iqbal<sup>1</sup>, Colton Bishop<sup>1</sup>, Rebecca Roelofs<sup>1</sup>, Lei M Zhang<sup>1</sup>, Kay McKinney<sup>1</sup>, Disha Shrivastava<sup>1</sup>, Cosmin Paduraru<sup>1</sup>, George Tucker<sup>1</sup>, Doina Precup<sup>1</sup>, Feryal Behbahani<sup>†,1</sup> and Aleksandra Faust<sup>†,1</sup>

<sup>1</sup>Google DeepMind, <sup>\*</sup>Equal Contribution, <sup>+</sup>Randomly ordered via coin flip, <sup>†</sup>Jointly supervised.

### GENERATING SEQUENCES BY LEARNING TO [SELF-]CORRECT

Sean Welleck<sup>1,3,\*</sup> Ximing Lu<sup>1,\*</sup>

Peter West<sup>3,†</sup> Faeze Brahman<sup>1,†</sup>

Tianxiao Shen<sup>3</sup> Daniel Khashabi<sup>2</sup> Yejin Choi<sup>1,3</sup>

<sup>1</sup>Allen Institute for Artificial Intelligence

<sup>2</sup>Center for Language and Speech Processing, Johns Hopkins University

<sup>3</sup>Paul G. Allen School of Computer Science & Engineering, University of Washington

---

## SELF-REFINE: Iterative Refinement with Self-Feedback

---

Aman Madaan<sup>1</sup>, Niket Tandon<sup>2</sup>, Prakhar Gupta<sup>1</sup>, Skyler Hallinan<sup>3</sup>, Luyu Gao<sup>1</sup>, Sarah Wiegrefe<sup>2</sup>, Uri Alon<sup>1</sup>, Nouha Dziri<sup>2</sup>, Shrimai Prabhumoye<sup>4</sup>, Yiming Yang<sup>1</sup>, Shashank Gupta<sup>2</sup>, Bodhisattwa Prasad Majumder<sup>5</sup>, Katherine Hermann<sup>6</sup>, Sean Welleck<sup>2,3</sup>, Amir Yazdanbakhsh<sup>6</sup>, Peter Clark<sup>2</sup>

<sup>1</sup>Language Technologies Institute, Carnegie Mellon University

<sup>2</sup>Allen Institute for Artificial Intelligence

<sup>3</sup>University of Washington <sup>4</sup>NVIDIA <sup>5</sup>UC San Diego <sup>6</sup>Google Research, Brain Team

[amadaan@cs.cmu.edu](mailto:amadaan@cs.cmu.edu), [nikett@allenai.org](mailto:nikett@allenai.org)

# Do LLMs Know They are Wrong?

- It's hard to say, but most of the time yes
  - Verifying is easier than generating
- Correction-based and refinement-based approaches

## LARGE LANGUAGE MODELS CANNOT SELF-CORRECT REASONING YET

**Jie Huang<sup>1,2\*</sup> Xinyun Chen<sup>1\*</sup> Swaroop Mishra<sup>1</sup> Huaixiu Steven Zheng<sup>1</sup> Adams Wei Yu<sup>1</sup>  
Xinying Song<sup>1</sup> Denny Zhou<sup>1</sup>**

<sup>1</sup>Google DeepMind    <sup>2</sup>University of Illinois at Urbana-Champaign

jeffhj@illinois.edu, {xinyunchen, dennyzhou}@google.com

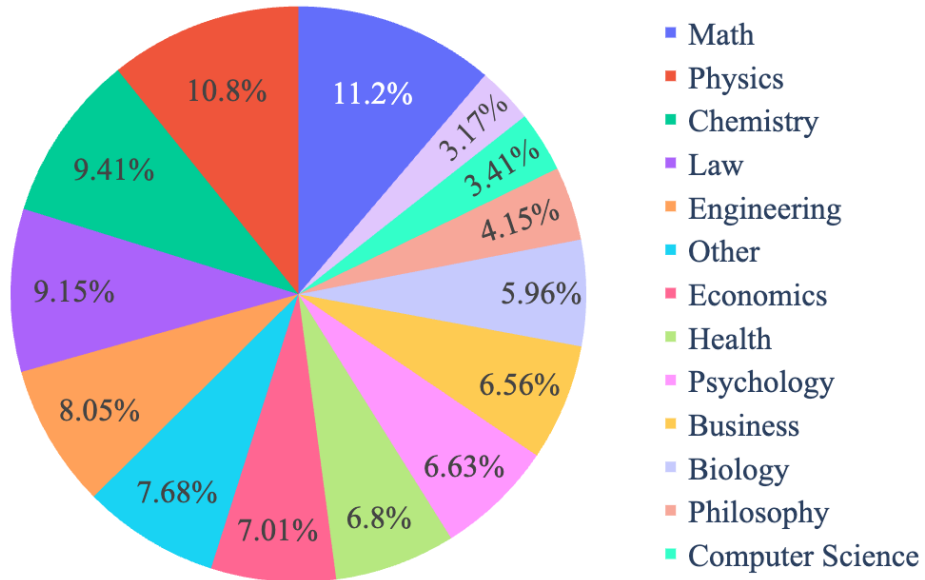


# Lecture Plan

- Large Language Models
  - Prompting
  - In-Context Learning
  - Chain-of-Thought Prompting
- Evaluation of Large Language Models
- Alignment
  - Instruction Tuning
  - Human Preference Optimization

# Benchmark

- MMLU and MMLU-pro



Find all  $c$  in  $\mathbb{Z}_3$  such that  $\mathbb{Z}_3[x]/(x^2 + c)$  is a field.  
 (A) 0 (B) 1 (C) 2 (D) 3

Figure 14: An Abstract Algebra example.

What is the embryological origin of the hyoid bone?  
 (A) The first pharyngeal arch  
 (B) The first and second pharyngeal arches  
 (C) The second pharyngeal arch  
 (D) **The second and third pharyngeal arches**

Figure 15: An Anatomy example.

Why isn't there a planet where the asteroid belt is located?  
 (A) A planet once formed here but it was broken apart by a catastrophic collision.  
 (B) There was not enough material in this part of the solar nebula to form a planet.  
 (C) There was too much rocky material to form a terrestrial planet but not enough gaseous material to form a jovian planet.  
 (D) **Resonance with Jupiter prevented material from collecting together to form a planet.**

Figure 16: An Astronomy example.

Three contrasting tactics that CSO's can engage in to meet their aims are \_\_\_\_\_ which typically involves research and communication, \_\_\_\_\_, which may involve physically attacking a company's operations or \_\_\_\_\_, often involving some form of \_\_\_\_\_.  
 (A) Non-violent direct action, Violent direct action, Indirect action, Boycott  
 (B) Indirect action, Instrumental action, Non-violent direct action, Information campaign  
 (C) **Indirect action, Violent direct action, Non-violent direct-action Boycott.**  
 (D) Non-violent direct action, Instrumental action, Indirect action, Information campaign

Figure 17: A Business Ethics example.

# Benchmark

- BIG-Bench Hard

## Model Input (Navigate)

<Few shot chain-of-thought exemplars omitted for brevity>

**Q:** If you follow these instructions, do you return to the starting point? Turn left. Turn right. Take 5 steps. Take 4 steps. Turn around. Take 9 steps.

Options:

- Yes
- No

**A:** Let's think step by step.

## Model Output

We start at the origin (0, 0), facing the positive y-axis.  
(1) Turn left: (0, 0), facing the negative x-axis.  
(2) Turn right: (0, 0), facing the positive y-axis.  
(3) Take 5 steps: (0, 5), facing the positive y-axis.  
(4) Take 4 steps: (0, 9), facing the positive y-axis.  
(5) Turn around: (0, 9), facing the negative y-axis.  
(6) Take 9 steps: (0, 0), facing the negative y-axis.  
Since (0, 0) is (0, 0), we are indeed where we started. So the answer is **Yes**. ✓

## Model Input (Word Sorting)

<Few shot chain-of-thought exemplars omitted for brevity>

**Q:** Sort the following words alphabetically: List: burley bela arapaho bacteria bock

**A:** Let's think step by step.

## Model Output

The first letter: "burley": "b" (2). "bela": "b" (2). "arapaho": "a" (1). "bacteria": "b" (2). "bock": "b" (2).  
We now have: (1) "arapaho" < (2) ["burley" ? "bela" ? "bacteria" ? "bock"]. Now let's sort this subpart ["burley" ? "bela" ? "bacteria" ? "bock"] by looking at their second letters.  
The second letter: "burley": "u" (21). "bela": "e" (5). "bacteria": "a" (1). "bock": "o" (15). We now have: (1) "bacteria" < (5) "bela" < (15) "bock" < (21) "burley". Hence, we have "arapaho" < ["bacteria" < "bela" < "bock" < "burley"]. So the answer is **arapaho bacteria bela bock burley**. ✓

# Benchmark

- MATH

- Algebra
- Geometry
- Probability
- Calculus

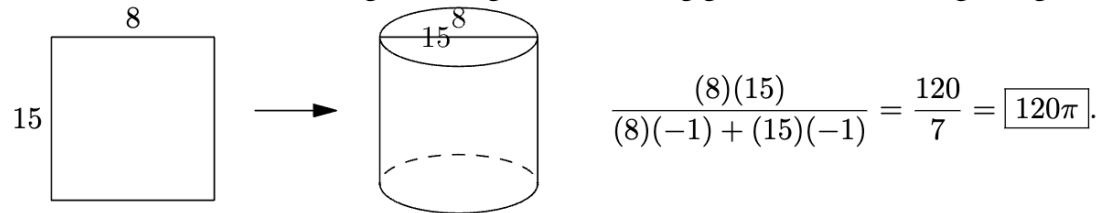
**Problem:** Suppose  $a$  and  $b$  are positive real numbers with  $a > b$  and  $ab = 8$ . Find the minimum value of  $\frac{a^2+b^2}{a-b}$ .

**Generated solution:** Let  $s = a + b$ . By QM-AM,  $\sqrt{\frac{a^2+b^2}{2}} \geq \frac{a+b}{2} = \frac{s}{2}$ . Then  $\frac{a^2+b^2}{2} \geq \frac{s^2}{4}$ , so  $a^2 + b^2 \geq \frac{s^2}{2}$ . Hence,  $\frac{a^2+b^2}{a-b} \geq \frac{s^2}{2} + \frac{2s^2}{a-b} = \frac{s^2}{2} + 2s^2 = \boxed{8}$ .

**Ground truth solution:** We can write  $\frac{a^2+b^2}{a-b} = \frac{a^2+b^2-2ab+16}{a-b} = \frac{(a-b)^2+16}{a-b} = a - b + \frac{16}{a-b}$ . By AM-GM,  $a - b + \frac{16}{a-b} \geq 2\sqrt{(a-b) \cdot \frac{16}{a-b}} = 8$ . Equality occurs when  $a - b = 4$  and  $ab = 8$ . We can solve these equations to find  $a = 2\sqrt{3} + 2$  and  $b = 2\sqrt{3} - 2$ . Thus, the minimum value is  $\boxed{8}$ .

**Problem:** Right  $\triangle ABC$  has legs measuring 8 cm and 15 cm. The triangle is rotated about one of its legs. What is the number of cubic centimeters in the maximum possible volume of the resulting solid? Express your answer in terms of  $\pi$ .

**Generated solution:** Rotating the triangle about one leg gives us the following triangle:



**Ground truth solution:** If the triangle is rotated about the shorter leg, then the radius is the longer leg and the height is the shorter leg, and the volume is  $\frac{1}{3} \cdot (15^2\pi)(8) = 600\pi$  cubic centimeters. If the triangle is rotated about the longer leg, then the radius is the shorter leg and the height is the longer leg, and the volume is  $\frac{1}{3}(8^2\pi)(15)$ , which is  $\frac{8}{15}$  of the volume we found earlier. So, the maximum possible volume is  $\boxed{600\pi}$  cubic centimeters.

Figure 3: Problems, step-by-step solutions generated by our GPT-2 1.5B model, and ground truth solutions. Observe that models can provide the right answer yet generate a misleading and wrong explanation. The second generated solution demonstrates that models are capable of generating Asymptote commands to create figures and graphics.

# Benchmark

- GSM8K
  - Math problems

**Problem:** Beth bakes 4, 2 dozen batches of cookies in a week. If these cookies are shared amongst 16 people equally, how many cookies does each person consume?

**Solution:** Beth bakes 4 2 dozen batches of cookies for a total of  $4 \times 2 = 8$  dozen cookies  
There are 12 cookies in a dozen and she makes 8 dozen cookies for a total of  $12 \times 8 = 96$  cookies  
She splits the 96 cookies equally amongst 16 people so they each eat  $96/16 = 6$  cookies

**Final Answer:** 6

**Problem:** Mrs. Lim milks her cows twice a day. Yesterday morning, she got 68 gallons of milk and in the evening, she got 82 gallons. This morning, she got 18 gallons fewer than she had yesterday morning. After selling some gallons of milk in the afternoon, Mrs. Lim has only 24 gallons left. How much was her revenue for the milk if each gallon costs \$3.50?

Mrs. Lim got 68 gallons - 18 gallons = 50 gallons this morning.

So she was able to get a total of 68 gallons + 82 gallons + 50 gallons = 200 gallons.

She was able to sell 200 gallons - 24 gallons = 176 gallons.

Thus, her total revenue for the milk is  $\$3.50/\text{gallon} \times 176 \text{ gallons} = \$616$ .

**Final Answer:** 616

**Problem:** Tina buys 3 12-packs of soda for a party. Including Tina, 6 people are at the party. Half of the people at the party have 3 sodas each, 2 of the people have 4, and 1 person has 5. How many sodas are left over when the party is over?

**Solution:** Tina buys 3 12-packs of soda, for  $3 \times 12 = 36$  sodas

6 people attend the party, so half of them is  $6/2 = 3$  people

Each of those people drinks 3 sodas, so they drink  $3 \times 3 = 9$  sodas

Two people drink 4 sodas, which means they drink  $2 \times 4 = 8$  sodas

With one person drinking 5, that brings the total drank to  $5 + 9 + 8 + 3 = 25$  sodas

As Tina started off with 36 sodas, that means there are  $36 - 25 = 11$  sodas left

**Final Answer:** 11

# Benchmark

Category Benchmark	Llama 3.1 70B	Llama 3.3 70B	Amazon Nova Pro	Llama 3.1 405B	Gemini Pro 1.5	GPT-4o	Claude 3.5 Sonnet
General							
MMLU Chat (0-shot, CoT)	86.0	<b>86.0</b>	85.9	<b>88.6</b>	87.1	87.5	<b>88.9</b>
MMLU PRO (5-shot, CoT)	66.4	<b>68.9</b>	-	<b>73.4</b>	76.1	73.8	<b>77.8</b>
Instruction Following							
IFEval	87.5	<b>92.1</b>	<b>92.1</b>	<b>88.6</b>	81.9	84.6	<b>89.3</b>
Code							
HumanEval (0-shot)	80.5	<b>88.4</b>	<b>89.0</b>	<b>89.0</b>	89.0	86.0	<b>93.7</b>
MBPP EvalPlus (base) (0-shot)	86.0	<b>87.6</b>	-	<b>88.6</b>	87.8	83.9	86.8
Math							
MATH (0-shot, CoT)	67.8	<b>77.0</b>	76.6	<b>73.9</b>	<b>82.9</b>	76.9	78.3
Reasoning							
GPQA Diamond (0-shot, CoT)	48.0	<b>50.5</b>	-	<b>49.0</b>	53.5	47.5	<b>65.0</b>
Tool use							
BFCL v2 (0-shot)	77.5	<b>77.3</b>	-	<b>81.1</b>	80.3	74.0	79.3
Long context							
NIH/Multi-needle	97.5	<b>97.5</b>	-	<b>98.1</b>	94.7	-	<b>99.4</b>
Multilingual							
Multilingual MGSM (0-shot)	86.9	<b>91.1</b>	-	<b>91.6</b>	89.6	90.6	<b>92.8</b>

# Chatbot Arena

## 🗄️ Chatbot Arena (formerly LMSYS): Free AI Chat to Compare & Test Best AI Chatbots

小红书 | [Twitter](#) | [Discord](#) | [Blog](#) | [GitHub](#) | [Paper](#) | [Dataset](#) | [Kaggle Competition](#)

Grok-3 result is released here: [https://x.com/lmarena\\_ai/status/1891706264800936307!](https://x.com/lmarena_ai/status/1891706264800936307!)

### 📖 How It Works

- **Blind Test:** Ask any question to two anonymous AI chatbots (ChatGPT, Gemini, Claude, Llama, and more).
- **Vote for the Best:** Choose the best response. You can keep chatting until you find a winner.
- **Play Fair:** If AI identity reveals, your vote won't count.
- **NEW features:** Upload an image 🖼️ and chat, or use 🎨 **Text-to-Image** models like DALL-E 3, Flux, Ideogram to generate images! Use 🐙 **RepoChat** tab to chat with Github repos.

### 🏆 Chatbot Arena LLM Leaderboard

- Backed by over 1,000,000+ community votes, our platform ranks the best LLM and AI chatbots. Explore the top AI models on our LLM [leaderboard!](#)

### 📌 Chat now!

🔍 Expand to see the descriptions of 89 models

💬 Model A	💬 Model B
-----------	-----------

# Chatbot Arena Leaderboard

Rank* (UB)	Rank (StyleCtrl)	Model	Arena Score	95% CI	Votes	Organization	License
1	1	<a href="#">chocolate...(Early..Grok-3)</a>	1402	+7/-6	7829	xAI	Proprietary
2	4	<a href="#">Gemini-2.0-Flash-Thinking-Exp-01-21</a>	1385	+5/-5	13336	Google	Proprietary
2	2	<a href="#">Gemini-2.0-Pro-Exp-02-05</a>	1379	+5/-6	11197	Google	Proprietary
2	1	<a href="#">ChatGPT-4o-latest...(2025-01-29)</a>	1377	+5/-6	10529	OpenAI	Proprietary
5	2	<a href="#">DeepSeek-R1</a>	1361	+8/-7	5079	DeepSeek	MIT
5	8	<a href="#">Gemini-2.0-Flash-001</a>	1356	+6/-5	9092	Google	Proprietary
5	2	<a href="#">o1-2024-12-17</a>	1353	+6/-5	15437	OpenAI	Proprietary
8	6	<a href="#">o1-preview</a>	1335	+4/-4	33169	OpenAI	Proprietary
8	8	<a href="#">Qwen2.5-Max</a>	1332	+7/-7	7370	Alibaba	Proprietary
10	9	<a href="#">DeepSeek-V3</a>	1317	+4/-4	17717	DeepSeek	DeepSeek
10	11	<a href="#">Qwen-Plus-0125</a>	1313	+8/-10	3682	Alibaba	Proprietary
10	11	<a href="#">Gemini-2.0-Flash-Lite-Preview-02-05</a>	1310	+6/-6	8465	Google	Proprietary
10	14	<a href="#">GLM-4-Plus-0111</a>	1308	+8/-8	4171	Zhipu	Proprietary
11	11	<a href="#">o3-mini</a>	1305	+6/-7	9338	OpenAI	Proprietary
11	16	<a href="#">Step-2-16K-Exp</a>	1304	+7/-11	5133	StepFun	Proprietary



# Lecture Plan

- Large Language Models
  - Prompting
  - In-Context Learning
  - Chain-of-Thought Prompting
- Evaluation of Large Language Models
- Alignment
  - Instruction Tuning
  - Human Preference Optimization

# Why Alignment?

- Language modeling  $\neq$  assisting users

**PROMPT** *Explain the moon landing to a 6 year old in a few sentences.*

**COMPLETION**

GPT-3

Explain the theory of gravity to a 6 year old.

Explain the theory of relativity to a 6 year old in a few sentences.

Explain the big bang theory to a 6 year old.

Explain evolution to a 6 year old.

# Why Alignment?

- Language modeling  $\neq$  assisting users

**PROMPT** *Explain the moon landing to a 6 year old in a few sentences.*

**COMPLETION**

**Human**

A giant rocket ship blasted off from Earth carrying astronauts to the moon. The astronauts landed their spaceship on the moon and walked around exploring the lunar surface. Then they returned safely back to Earth, bringing home moon rocks to show everyone.

# Why Alignment?

- Continuing writing does not always work

Prompt

This place is incredible! The lobster is the best I've ever had. The sentiment of the above sentence is

positive.

Completion

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 **X**

# Instruction Tuning

- LLMs have knowledge, but don't always generate the outputs we want
- Training LLMs to following **human instructions**

### Annotated task definitions

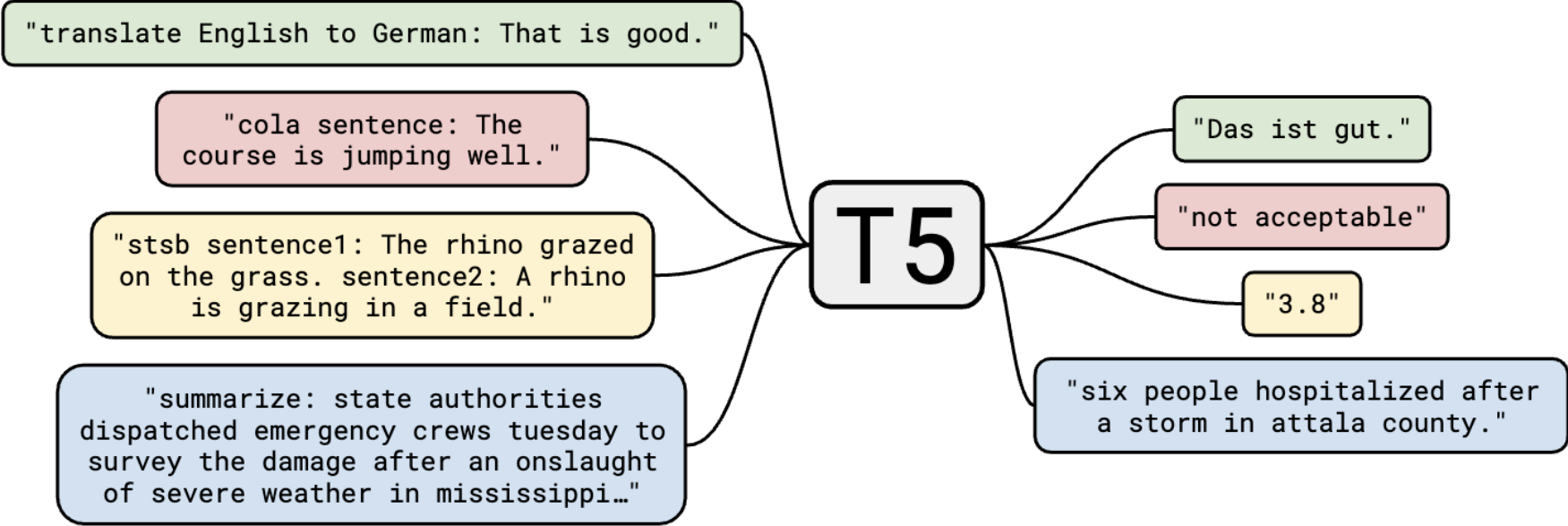
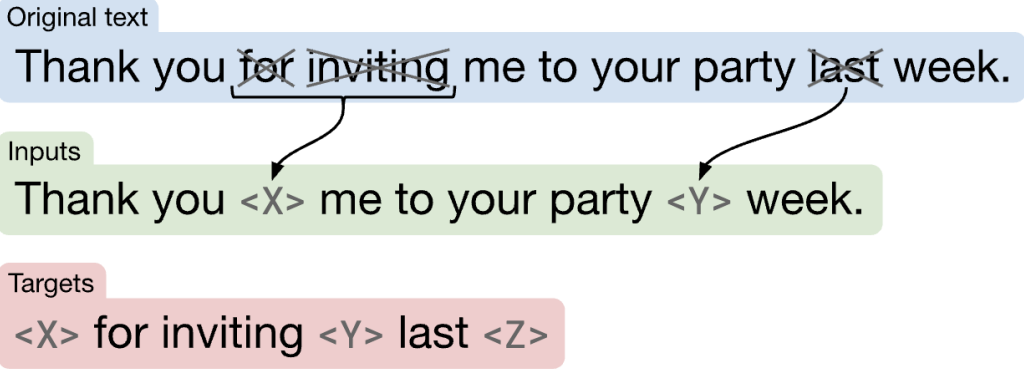
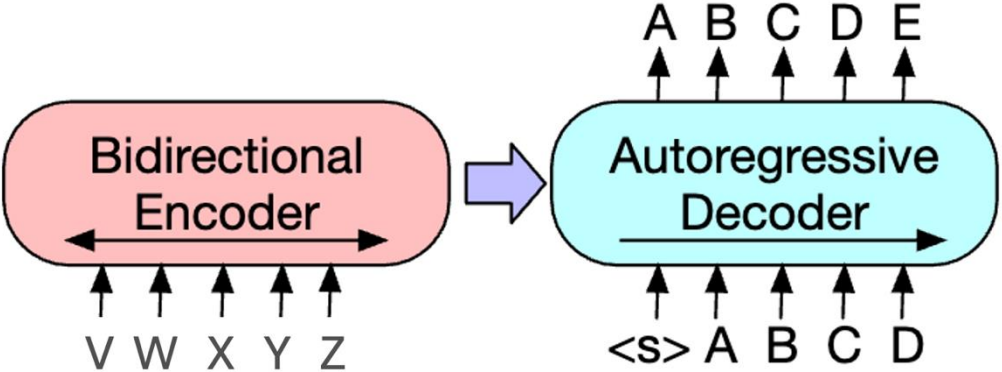
You will be given two pieces of text... One of them is simpler ...  
You are expected to output 'Text one' if the first sentence is simpler.  
Otherwise output 'Text two'.

Given a sentence with a missing word, pick the answer option that best fills out the missing word in the sentence. Indicate each answer with its index ('a', 'b', 'c', 'd').

Given a document, generate a short title of the document. The title should convey the main idea/event/topic about which the document is being written.

Category	Description
Input Content	Primary description of the task input
<u>Additional Input Content</u>	Additional details on task input
Action Content	Action to perform for task
Input Mention	Mentions of input within action content
Output Content	Primary description of task output
<u>Additional Output Content</u>	Additional details on task output
Label List	Task output labels (classification only)
Label Definition	Task Label definitions (classification only)

# Recap: T5



# Instruction Tuning

- Convert existing tasks to (input, output) format
- Create many prompts and collect human answers

### Annotated task definitions

You will be given two pieces of text... One of them is simpler ...  
You are expected to output 'Text one' if the first sentence is simpler.  
Otherwise output 'Text two'.

Given a sentence with a missing word, pick the answer option that best fills out the missing word in the sentence. Indicate each answer with its index ('a', 'b', 'c', 'd').

Given a document, generate a short title of the document. The title should convey the main idea/event/topic about which the document is being written.

Category	Description
Input Content	Primary description of the task input
Additional Input Content	Additional details on task input
Action Content	Action to perform for task
Input Mention	Mentions of input within action content
Output Content	Primary description of task output
Additional Output Content	Additional details on task output
Label List	Task output labels (classification only)
Label Definition	Task Label definitions (classification only)

# Sidenote: Why Decoder-Only Instead of Encoder-Decoder?

Original text

Thank you ~~for inviting~~ me to your party ~~last~~ week.

Inputs

Thank you <X> me to your party <Y> week.

Targets

<X> for inviting <Y> last <Z>

Binge ... on | - | and | of | is

Binge **drinking** ... is | and | had | in | was

Binge drinking **may** ... be | also | have | not | increase

Binge drinking may **not** ... be | have | cause | always | help

Binge drinking may not **necessarily** ... be | lead | cause | results | have

Binge drinking may not necessarily **kill** ... you | the | a | people | your

Binge drinking may not necessarily kill **or** ... even | injure | kill | cause | prevent

Binge drinking may not necessarily kill or **even** ... kill | prevent | cause | reduce | injure

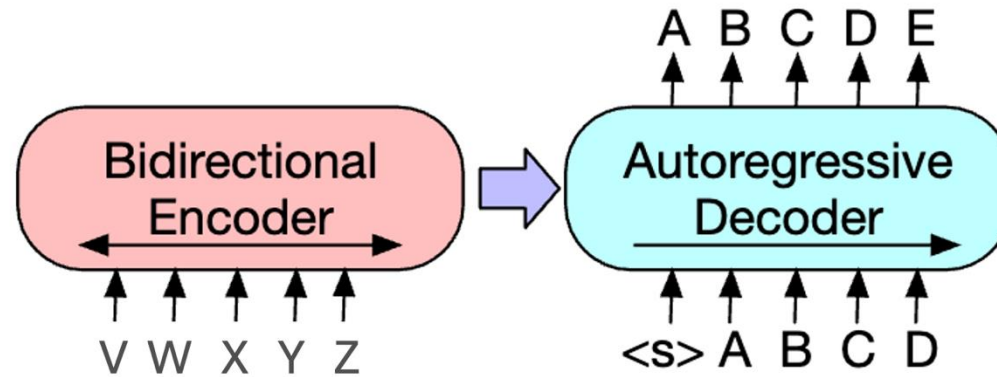
Binge drinking may not necessarily kill or even **damage** ... your | the | a | you | someone

Binge drinking may not necessarily kill or even damage **brain** ... cells | functions | tissue | neurons

Binge drinking may not necessarily kill or even damage brain **cells**, ... some | it | the | is | long



# Sidenote: Why Decoder-Only Instead of Encoder-Decoder?

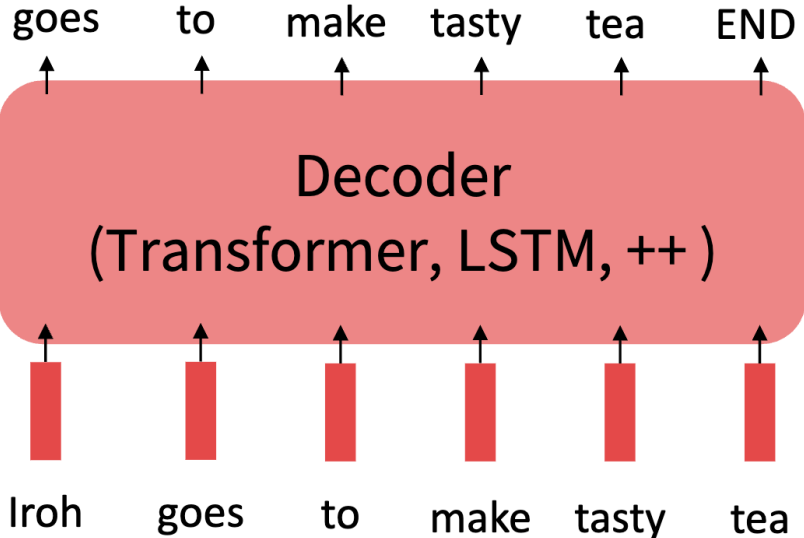


Binge ... on | - | and | of | is  
Binge **drinking** ... is | and | had | in | was  
Binge drinking **may** ... be | also | have | not | increase  
Binge drinking may **not** ... be | have | cause | always | help  
Binge drinking may not **necessarily** ... be | lead | cause | results | have  
Binge drinking may not necessarily **kill** ... you | the | a | people | your  
Binge drinking may not necessarily kill **or** ... even | injure | kill | cause | prevent  
Binge drinking may not necessarily kill or **even** ... kill | prevent | cause | reduce | injure  
Binge drinking may not necessarily kill or even **damage** ... your | the | a | you | someone  
Binge drinking may not necessarily kill or even damage **brain** ... cells | functions | tissue | neurons  
Binge drinking may not necessarily kill or even damage brain **cells**, ... some | it | the | is | long

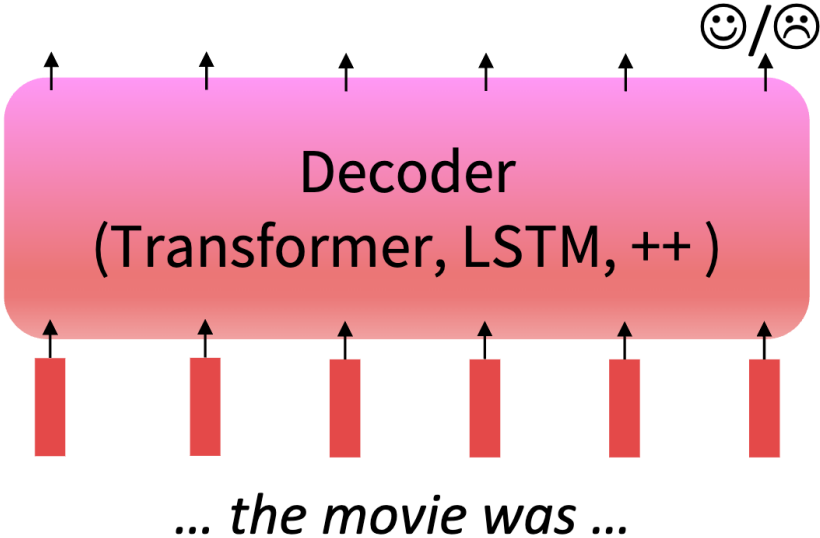
# Scaling Up Instruction Tuning

## Step 1: Pretrain (on language modeling)

Lots of text; learn general things!



**Step 2: Finetune (on many tasks)**  
Not many labels; adapt to the tasks!



# Instruction Tuning → Instruction Pre-Training

- Instruction fine-tuning for many tasks



# Instruction Tuning

## Model input (Disambiguation QA)

Q: In the following sentences, explain the antecedent of the pronoun (which thing the pronoun refers to), or state that it is ambiguous.

Sentence: The reporter and the chef will discuss their favorite dishes.

Options:

- (A) They will discuss the reporter's favorite dishes
- (B) They will discuss the chef's favorite dishes
- (C) Ambiguous

A: Let's think step by step.

## Before instruction finetuning

The reporter and the chef will discuss their favorite dishes.

The reporter and the chef will discuss the reporter's favorite dishes.

The reporter and the chef will discuss the chef's favorite dishes.

The reporter and the chef will discuss the reporter's and the chef's favorite dishes.

✘ (doesn't answer question)

# Instruction Tuning

## Model input (Disambiguation QA)

Q: In the following sentences, explain the antecedent of the pronoun (which thing the pronoun refers to), or state that it is ambiguous.

Sentence: The reporter and the chef will discuss their favorite dishes.

Options:

- (A) They will discuss the reporter's favorite dishes
- (B) They will discuss the chef's favorite dishes
- (C) Ambiguous

A: Let's think step by step.

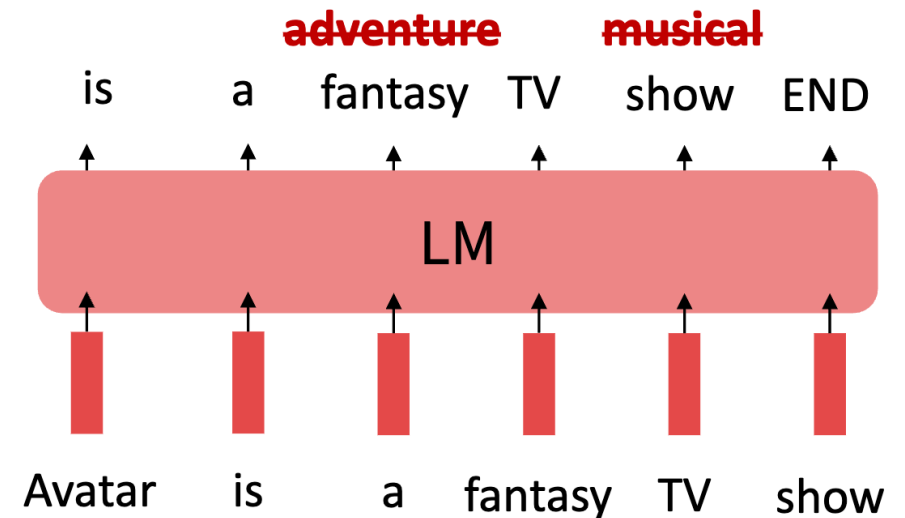
## After instruction finetuning

The reporter and the chef will discuss their favorite dishes does not indicate whose favorite dishes they will discuss. So, the answer is (C). ✓

# Limitations of Instruction Fine-Tuning

- It is expensive to collect ground-truth data for tasks
- Open-ended creative generation have no right answer
  - E.g., write me a story about a dog and her pet grasshopper
- language modeling penalizes all token-level mistakes equally, but some errors are worse than others

Even with instruction finetuning, there is still a mismatch between the LM objective and “satisfying human preferences”!



# Lecture Plan

- Large Language Models
  - Prompting
  - In-Context Learning
  - Chain-of-Thought Prompting
- Evaluation of Large Language Models
- Alignment
  - Instruction Tuning
  - Human Preference Optimization