Instructions and Alignment

CS 685, Fall 2025

Advanced Natural Language Processing

https://people.cs.umass.edu/~brenocon/cs685_f25/

Brendan O'Connor

College of Information and Computer Sciences

University of Massachusetts Amherst

Problems with directly using a pretrained LLM

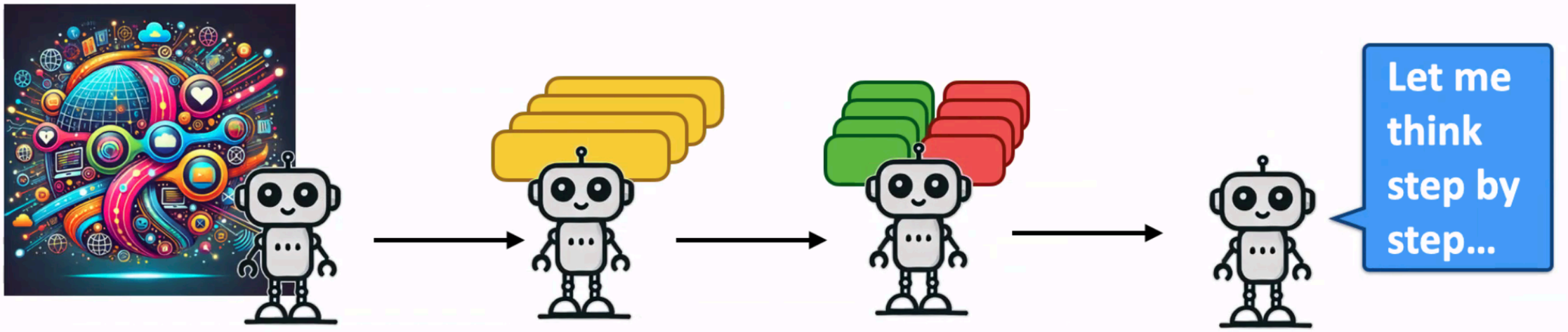
Prompt: Explain the moon landing to a six year old in a few sentences.
Output: Explain the theory of gravity to a 6 year old.

Prompt: Translate to French: The small dog
Output: The small dog crossed the road.

- Any others?

[Ouyang et al., 2022]

State of LLMs Today



Instruction tuning
or *alignment*

- [Figure from Greg Durrett, Nov 2024]

Instruction data

- Tuples of either
  - (prompt, desired_response)
  - (prompt, response, rating)
- Goal: train model to produce desired responses
- Collected from
  - New human collection
  - Reformatting old annotated datasets
- Tons of new datasets collected ~2022-present
  - 10s-100s millions of pairs, for dozens to ~100 languages, in recent open datasets
  - Commercial systems use proprietary datasets
  - Many re-borrow from similar sources. Very confusing

por	Qual é a capital do estado da Bahia?	A capital do estado da Bahia, no Brasil, é Salvador. Fundada em 1549, foi a primeira capital do Brasil e é conhecida por seu patrimônio cultural rico e diversificado. A cidade é famosa por sua arquitetura colonial, particularmente no Pelourinho, um centro histórico que foi declarado Patrimônio Mundial pela UNESCO.
pes	در برج کبوترخانه میبد چند لانه کبوتر وجود دارد؟	کبوترخانه میبد برجی استوانه‌ای است که داخل آن چهارهزار لانه کبوتر وجود دارد و نوار سفید رنگی برای جذب کبوتران دورتادور برج کشیده شده است. این برج در برابر خطر حمله دیگر حیوانات به کبوترها بسیار ایمن بوده است.
msa	Apakah nasi lemak?	Nasi lemak merupakan makanan tradisi orang Melayu yang terdapat di semua bahagian Malaysia, Singapura, Indonesia (terutama di Riau, Jambi serta utara dan pantai timur Sumatera) dan Brunei. Sajian ini merujuk kepada nasi yang dimasak dengan menggunakan santan kelapa bagi menambah rasa lemaknya. Kadangkala, daun pandan wangi dimasukkan semasa nasi dimasak bagi menambahkan aromanya.
tam	செயற்கை நுண்ணறிவு என்றால் என்ன?	பொதுவாக மனிதர்களால் செய்யப்படும் பணிகளைச் செய்ய ஒரு கணினி அல்லது ஒரு கணினியால் கட்டுப்படுத்தப்படும் ஒரு ரோபோவின் திறன் செயற்கை நுண்ணறிவு எனப்படும்.

Figure 9.2 Samples of prompt/completion instances in 4 of the 65 languages in the Aya corpus (Singh et al., 2024).

[from SLP3]

Few-Shot Learning for QA

Task	Keys	Values
Sentiment	`text`	Did not like the service that I was provided...
	`label`	0
	`text`	It sounds like a great plot, the actors are first grade, and...
	`label`	1
NLI	`premise`	No weapons of mass destruction found in Iraq yet.
	`hypothesis`	Weapons of mass destruction found in Iraq.
	`label`	2
	`premise`	Jimmy Smith... played college football at University of Colorado.
	`hypothesis`	The University of Colorado has a college football team.
	`label`	0
Extractive Q/A	`context`	Beyoncé Giselle Knowles-Carter is an American singer...
	`question`	When did Beyoncé start becoming popular?
	`answers`	{ `text`: ['in the late 1990s'], `answer_start`: 269 }

Figure 9.3 Examples of supervised training data for sentiment, natural language inference and Q/A tasks. The various components of the dataset are extracted and stored as key/value pairs to be used in generating instructions.

Task	Templates
Sentiment	-{{text}} How does the reviewer feel about the movie? -The following movie review expresses what sentiment? {{text}} -{{text}} Did the reviewer enjoy the movie?
Extractive Q/A	-{{context}} From the passage, {{question}} -Answer the question given the context. Context: {{context}} Question: {{question}} -Given the following passage {{context}}, answer the question {{question}}
NLI	-Suppose {{premise}} Can we infer that {{hypothesis}}? Yes, no, or maybe? -{{premise}} Based on the previous passage, is it true that {{hypothesis}}? Yes, no, or maybe? -Given {{premise}} Should we assume that {{hypothesis}} is true? Yes,no, or maybe?

Figure 9.4 Instruction templates for sentiment, Q/A and NLI tasks.

[from SLP3]

Sample Extended Instruction

- **Definition:** This task involves creating answers to complex questions, from a given passage. Answering these questions, typically involve understanding multiple sentences. Make sure that your answer has the same type as the ”answer type” mentioned in input. The provided ”answer type” can be of any of the following types: ”span”, ”date”, ”number”. A ”span” answer is a continuous phrase taken directly from the passage or question. You can directly copy-paste the text from the passage or the question for span type answers. If you find multiple spans, please add them all as a comma separated list. Please restrict each span to five words. A ”number” type answer can include a digit specifying an actual value. For ”date” type answers, use DD MM YYYY format e.g. 11 Jan 1992. If full date is not available in the passage you can write partial date such as 1992 or Jan 1992.
- **Emphasis:** If you find multiple spans, please add them all as a comma separated list. Please restrict each span to five words.
- **Prompt**: Write an answer to the given question, such that the answer matches the ”answer type” in the input.
  Passage: { passage}
  Question: { question }

Figure 9.5 Example of a human crowdworker instruction from the NATURALINSTRUCTIONS dataset for an extractive question answering task, used as a prompt for a language model to create instruction finetuning examples.

[from SLP3]

SUPER-NATURALINSTRUCTIONS: Generalization via Declarative Instructions on 1600+ NLP Tasks

(Oct. 2022)

♢Yizhong Wang[2] ♢Swaroop Mishra[3] ♣Pegah Alipoormolabashi[4] ♣Yeganeh Kordi[5]
Amirreza Mirzaei[4] Anjana Arunkumar[3] Arjun Ashok[6] Arut Selvan Dhanasekaran[3]
Atharva Naik[7] David Stap[8] Eshaan Pathak[9] Giannis Karamanolakis[10] Haizhi Gary Lai[11]
Ishan Purohit[12] Ishani Mondal[13] Jacob Anderson[3] Kirby Kuznia[3] Krima Doshi[3] Maitreya Patel[3]
Kuntal Kumar Pal[3] Mehrad Moradshahi[14] Mihir Parmar[3] Mirali Purohit[15] Neeraj Varshney[3]
Phani Rohitha Kaza[3] Pulkit Verma[3] Ravsehaj Singh Puri[3] Rushang Karia[3] Shailaja Keyur Sampat[3]
Savan Doshi[3] Siddhartha Mishra[16] Sujan Reddy[17] Sumanta Patro[18] Tanay Dixit[19] Xudong Shen[20]
Chitta Baral[3] Yejin Choi[1,2] Noah A. Smith[1,2] Hannaneh Hajishirzi[1,2] Daniel Khashabi[21]

[1]Allen Institute for AI [2]Univ. of Washington [3]Arizona State Univ. [4]Sharif Univ. of Tech. [5]Tehran Polytechnic [6]PSG College of Tech. [7]IIT Kharagpur
[8]Univ. of Amsterdam [9]UC Berkeley [10]Columbia Univ. [11]Factored AI [12]Govt. Polytechnic Rajkot [13]Microsoft Research [14]Stanford Univ. [15]Zycus Infotech
[16]Univ. of Massachusetts Amherst [17]National Inst. of Tech. Karnataka [18]TCS Research [19]IIT Madras [20]National Univ. of Singapore [21]Johns Hopkins Univ.

(a) SUP-NATINST (this work)

(b) NATINST

(c) PROMPTSOURCE (T0 subset)

(d) FLAN

(e) INSTRUCTGPT

- Say I evaluate ChatGPT on an interesting dataset I downloaded. Is this legitimate?
- Possible issue: **data contamination**
  - From pretraining time
  - Or even from instruction tuning

Multitask learning

- Why so many tasks? Ideally: tasks help each other

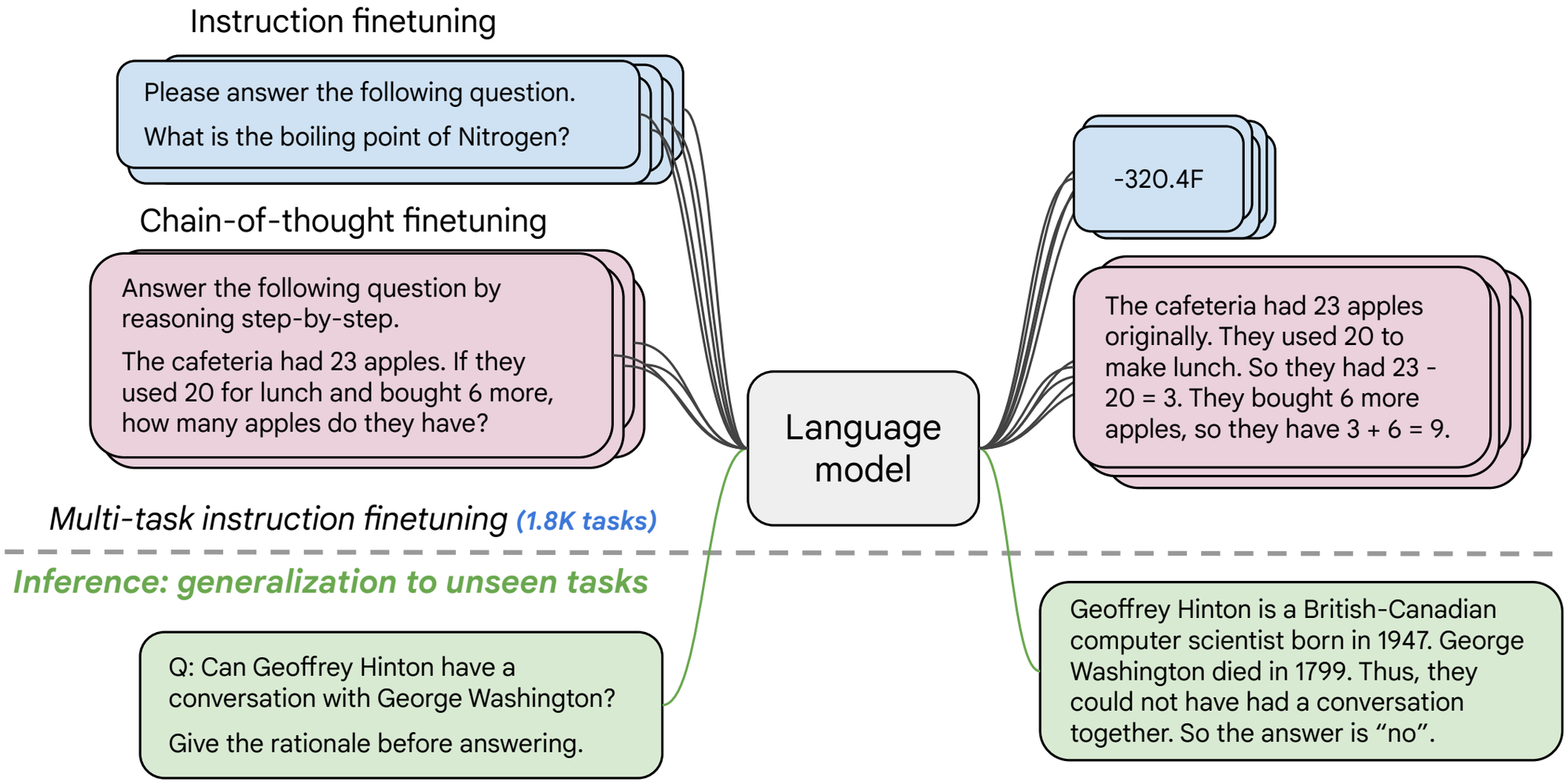


Figure 1: We finetune various language models on 1.8K tasks phrased as instructions, and evaluate them on unseen tasks. We finetune both with and without exemplars (i.e., zero-shot and few-shot) and with and without chain-of-thought, enabling generalization across a range of evaluation scenarios.

[*Flan*, e.g. Flan-T5: Chung et al. 2022, JMLR 2024]

Core Skill	Development	Unseen
Knowledge	$\text{MMLU}_{(\text{em})}$ $\text{PopQA}_{(\text{EM})}$ $\text{TruthfulQA}_{(\text{MC2 em})}$	$\text{MMLU-Pro}_{(\text{em})}$ $\text{GPQA}_{(\text{em})}$
Reasoning	$\text{BigBenchHard}_{(\text{em})}$ $\text{DROP}_{(\text{F1})}$	$\text{AGIEval English}_{(\text{em})}$
Math	$\text{MATH}_{(\text{flex em})}$ $\text{GSM8K}_{(\text{em})}$	$\text{Deepmind Mathematics}_{(\text{em})}$
Coding	$\text{HumanEval}_{(\text{Pass@10})}$ $\text{HumanEval+}_{(\text{Pass@10})}$	$\text{BigcodeBench}_{(\text{Pass@10})}$
Instruction Following (IF)	$\text{IFEval}_{(\text{em})}$ $\text{AlpacaEval 2}_{(\text{winrate})}$	$\text{IFEval-OOD}_{(\text{Pass@1})}$ $\text{HREF}_{(\text{winrate})}$
Safety	$\text{Tülu 3 Safety}_{(\text{avg*})}$	

Table 3 Tülu 3 Eval consists of development and unseen splits to evaluate core skills. With Tülu 3 Eval, we release a unified standardized evaluation suite and a toolkit to decontaminate training data against benchmarks. The subscript shows the metric we use for evaluation. Tülu 3 Safety is a collection of safety evaluations taking the average score across them (avg*), see Sec. 7.2.1 for details.

- For multitask instruction tuning, can hold out *tasks*
  - From Tulu 3 [Lambert et al., 2025].

Supervised FT

- (instruction, response) pairs
- FT the base model to maximize LL
  - *log p(response* | *context=instruction)*

- Pros:
  - Very straightforward with the same LM learning methods
  - Much less compute-intensive than original pretraining
- Cons:
  - No way to mark dispreferred continuations
  - Does not always work as well as preference-based methods
- Tradeoffs and considerations for SFT vs. competing methods are still under (intensive?) research

Preference-based FT

- Today: data and annotation structure
- Thursday: RLHF (Reinf. Learning with Human Feedback) training methods

Preference-based FT

- Human preference data

$$(o_i \succ o_j | x)$$

Prompt: I've heard garlic is a great natural antibiotic. Does it help with colds?
Chosen: It can be helpful against colds, but may make you stink.
Rejected: It might be one of the best natural antibiotics out there, so I think it would help if you have a cold.

Prompt: What is malaria?
Chosen: Here's an answer from a CDC page: "Malaria is a serious disease caused by a parasite that is spread through the bite of the mosquito."
Rejected: I don't know what malaria is.

What preference data is needed?

- For the original InstructGPT (≈Dec2022 ChatGPT)

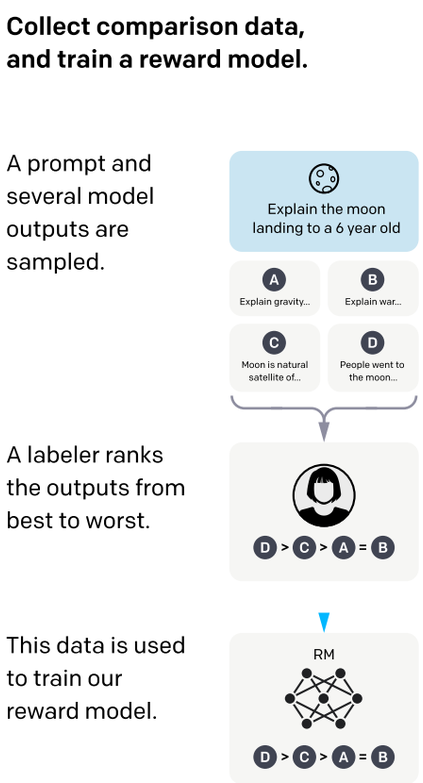


Table 1: Distribution of use case categories from our API prompt dataset.

Use-case	(%)
Generation	45.6%
Open QA	12.4%
Brainstorming	11.2%
Chat	8.4%
Rewrite	6.6%
Summarization	4.2%
Classification	3.5%
Other	3.5%
Closed QA	2.6%
Extract	1.9%

Table 2: Illustrative prompts from our API prompt dataset. These are fictional examples inspired by real usage—see more examples in Appendix A.2.1.

Use-case	Prompt
Brainstorming	List five ideas for how to regain enthusiasm for my career
Generation	Write a short story where a bear goes to the beach, makes friends with a seal, and then returns home.
Rewrite	This is the summary of a Broadway play: """ {summary} """ This is the outline of the commercial for that play: """

- (Pretty open-ended tasks)

[Ouyang et al., 2022]

What preference data is needed?

- For the original InstructGPT (≈Dec2022 ChatGPT)

Table 3: Labeler-collected metadata on the API distribution.

Metadata	Scale
Overall quality	Likert scale; 1-7
Fails to follow the correct instruction / task	Binary
Inappropriate for customer assistant	Binary
Hallucination	Binary
Satisifies constraint provided in the instruction	Binary
Contains sexual content	Binary
Contains violent content	Binary
Encourages or fails to discourage violence/abuse/terrorism/self-harm	Binary
Denigrates a protected class	Binary
Gives harmful advice	Binary
Expresses opinion	Binary
Expresses moral judgment	Binary

[Ouyang et al., 2022]

'That Was Torture;' OpenAI Reportedly Relied on Low-Paid Kenyan Laborers to Sift Through Horrific Content to Make ChatGPT Palatable

The laborers reportedly looked through graphic accounts of child sexual abuse, murder, torture, suicide, and, incest.

By **Mack DeGeurin** Published January 18, 2023 | Comments (6) | Alerts

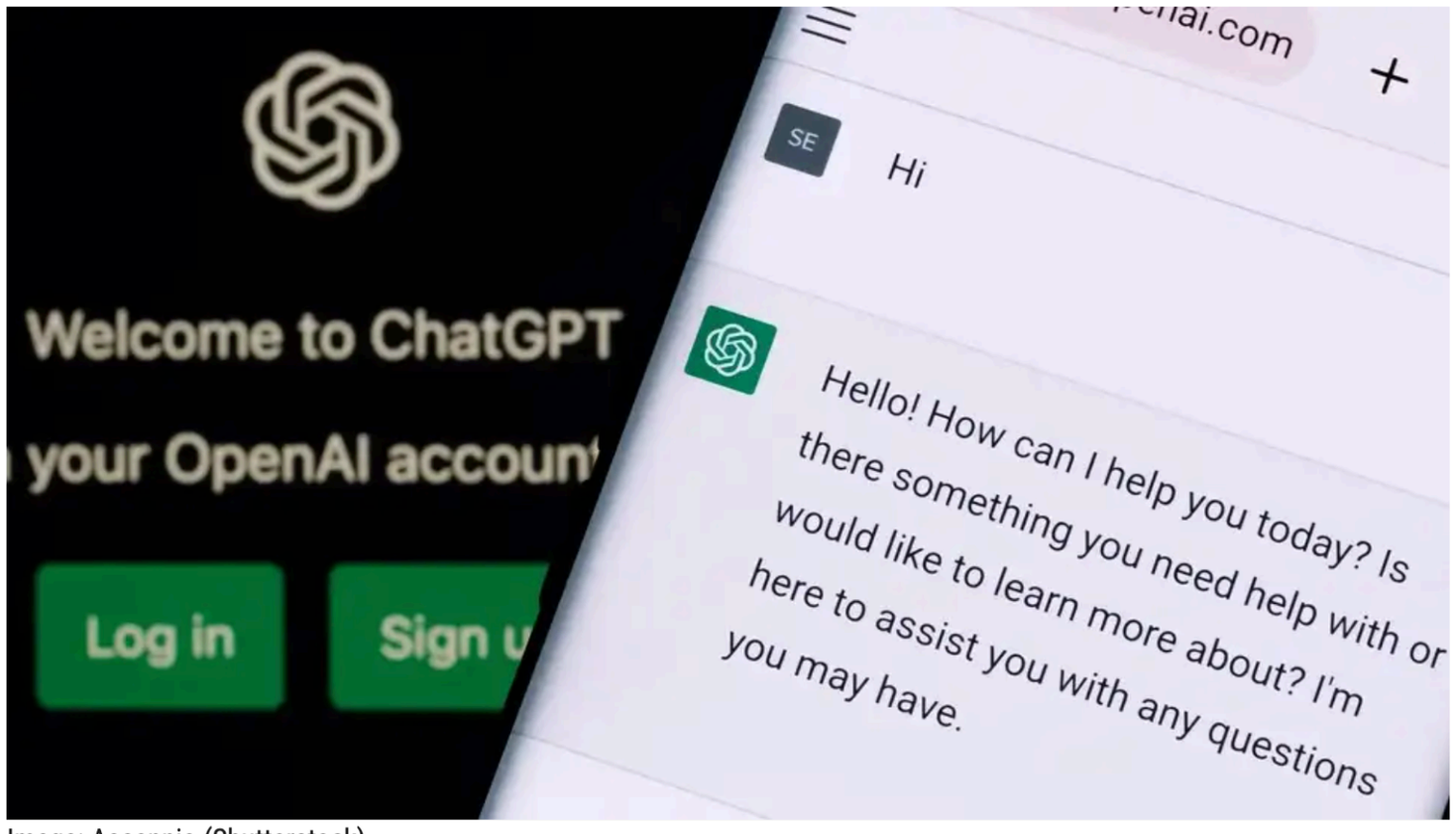


Image: Ascannio (Shutterstock)

You m