

Deep Dive into LLMs

Anupam Datta (Snowflake)

John Mitchell (Stanford)

AI Agents are powered by LLMs



ANTHROPIC



Common case: Call models over inference API to build agents

Why Dive into LLMs?

- Intellectual curiosity → how are these “engines” built
- Failure modes they may introduce in agents → informs agent evaluations
- Understanding why they succeed and fail → informs agent optimizations

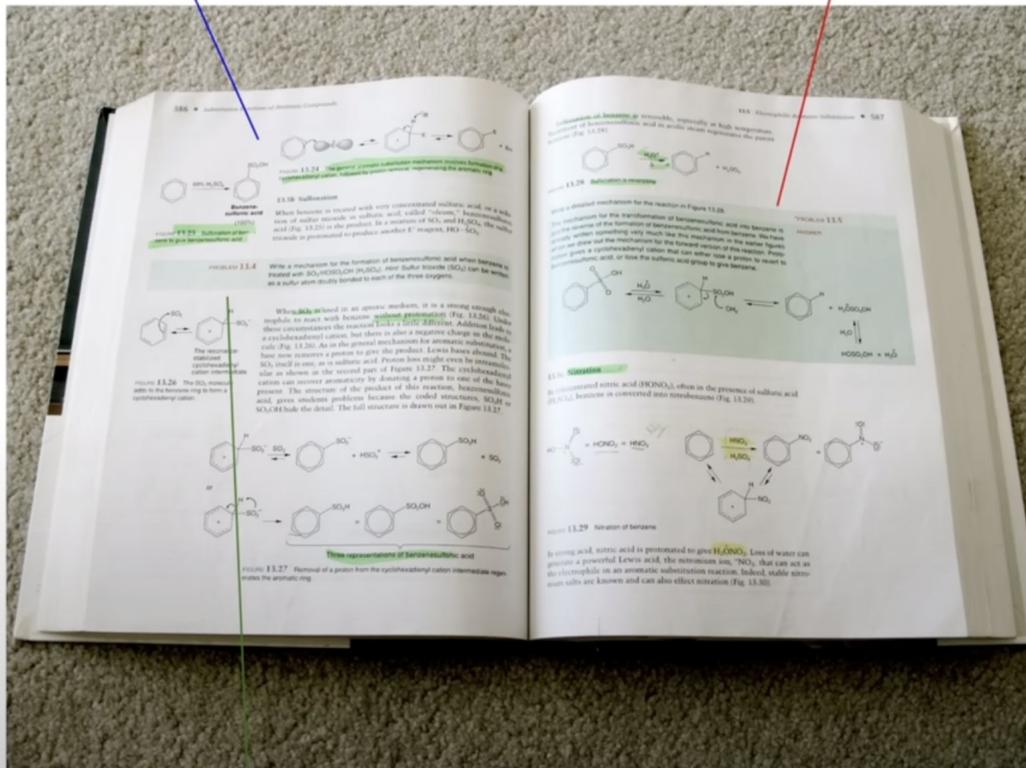
Training LLMs

Source: [Deep Dive into LLMs like ChatGPT](#) (YouTube | Andrej Karpathy)

- **Phase 1: Pre-training**
- Phase 2: Post-training
 - Supervised Fine-Tuning
 - Reinforcement Learning

exposition \Leftrightarrow pretraining
(background knowledge)

worked problems \Leftrightarrow supervised finetuning
(problem + demonstrated solution, for imitation)



practice problems \Leftrightarrow reinforcement learning
(prompts to practice, trial & error until you reach the correct answer)

Deep Learning: Origins in 1980s

[Link](#)

Fathers of the Deep Learning Revolution Receive ACM A.M. Turing Award

Bengio, Hinton and LeCun Ushered in Major Breakthroughs in Artificial Intelligence

2018 ACM A.M. Turing Award Laureates



Pre-Training LLMs

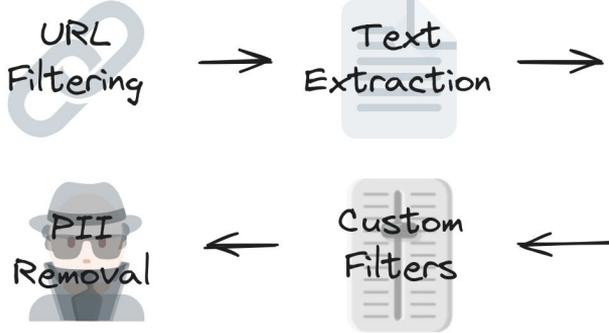
- Step 1: Download & pre-process the Internet
- Step 2: Tokenization
- Step 3: Train the neural network
- Step 4: Inference



Step 1: Download & pre-process the Internet

FineWeb: 44TB

[Viewing Single Post From: Spoilers for the Week of February 11th | Feb 11 2013, 09:58 AM | Don't care about Chloe/Tania/Jen-Jen. Don't care about Sami, really, but hoping that you see some good "SAMANTHA GENE!" Marlena Death-Stares out of it. And "newfound" feelings. Please, if only. STEFANO! STEFANO, STEFANO!!!! :cheer:\nSpoilers for the Week of February 11th - DAYS: News, Spoilers & Discussion] =sig= Fundamentalist community, let me pass on some advice to you I learned from the atheistic community:\nif you have set yourself on fire, do not run.\nOkay? Okay?? Please? Look, D, you had two months to say to Harvard in private emails, "I'm sorry, I shouldn't have been using that animation in my paid presentations. I went use it again. I really do like 'Inner Life', though, and would love to use it in classroom presentations, from the BioVisions site, if that is acceptable." \nI sat here, for two months, waiting for that to happen, anything to happen, and it didn't. Two months, on your own terms, you could have written a similar post to yesterdays. I would have given you the benefit of the doubt-- maybe you didn't know the credits weren't visible to the audience, and I wouldn't have said a word beyond this, as its Harvards problem, not mine. This would have been a funny joke to those of us in dealing with you people, but it would have been a PR non-issue for you.\nBut after you set yourself on fire, you didn't douse it out with a bucket of ice cold reality and accountability. You ran. And youre still running.\nWhy not just state "I screwed up. Sorry everyone." and move on? Why the excuses? Why the denial? Why the passive language? Why the vague words and cryptic capitalizations? Why the writes and rewrites of your 'press release'? We know it wasn't written of your own volition, or it would have been done *before* Harvard had to take action. And, your behavior before this, regarding this issue, is not indicative of someone who made an innocent mistake. Its weird.\nSo what with this frantic running? Is the inability to say "I was wrong" a pathological feature of Creationists? Or are you hiding something? Or is it both? Or is it more?\nAnd now we get Casey weighing in on the issue, according to cre8id at AboveTopSecret.com-- PBS/NOVA online - Intelligent Design on trial:\n...to my knowledge, Discovery Institute has neither authorized nor received nor is making use of any presentation that used that animation. We have had nothing to do with creating or selling a DVD of that animation, nor do we have anything to do with placing that presentation on Google Video. I dont know what he is talking about with that last part, but the first part sounds similar to Dis claims post-Dover ("WE HAD NOTHING TO DO WITH DOVER!"). Maybe Luskin is telling the truth. Maybe this was a magic non-science Creation-friendly narration with convenient edits that AIG or ICR would have killed for... but only Dembski could find it... but he cant tell us where... and he didnt share it with anyone... and its subsequently disappeared from the Internet.\nBut that simply isnt what Ive been told. Maybe this was all a silly Dembski mistake, blown out of proportion due to his decision to remain silent... But what if we find more videos of more DI fellows, presenting this animation?\nshrug= If youve set yourself on fire, do not run, DI. If youve done it, better admit it to Harvard now, apologize, and move on. A novel two-step immunotherapy approach has shown clinically beneficial responses in patients with advanced ovarian cancer. Following Lifestyle Recommendations Reduces Risk of Cancer Death\nPeople who follow the diet and lifestyle recommendations laid out by the WCRF and the AICR have a 20 percent reduced risk of dying from cancer. UCSF Launches Social Networking Site for Patients and Families... Genomic Test May Help Guide Prostate Cancer Treatment\nThe Oncotype DX® Prostate Cancer Test strongly predicts aggressiveness of disease. Statins Link to Lower Risk of Liver Cancer in Hepatitis C\nPeople infected with chronic hepatitis C are less likely to develop liver cancer if they are taking statins.\nRadioimmunotherapy (RIT) is a type of targeted therapy that delivers radiation directly to cancer cells... Urinary Incontinence\nOverview The urinary tract includes the kidneys, the ureters, the bladder, and the urethra. The kidneys... Advanced Directives\nLiving Wills Every competent adult has, in most cases, the freedom to accept or refuse medical treatment... Caregivers\nWhat is Caregiving and Who are Caregivers? Caregivers are individuals who provide care to chronically... Chemotherapy for Older Patients: What You Should Know About the Risk of Infection\nAs you may already know, chemotherapy works by attacking the rapidly dividing cells it finds in the body... An ongoing series highlighting complementary therapies, adapted from The Complete Guide to Complementary... Clear and precise\nMohs surgery provides a tissue-sparing approach to skin cancer surgery. By Eleanor Mayfield Michele Kelsey... Chemical Reaction\nChemicals may be disrupting our hormones--and our health. By Laurie Wertzich Exposure to synthetic chemicals... College Kids Kick Cancer\nBy Diana Price College kids and cancer--two topics most of us would immediately connect. And yet... Cooking with Fruits and Vegetables\nIn the introduction to Ripe: A Fresh, Colorful Approach to Fruits and Vegetables (Running Press, 2011)... Annual meeting brings together cancer experts from around the world. Karl Bohle, SCD The 2011 Annual... Bone Fractures in Breast Cancer Patients More Frequent with Femara than with Tamoxifen\nResearchers affiliated with the BIG I-98 Collaborative and International Breast Study Groups... Single Treatment with High-intensity Focused Ultrasound Effective for Localized Prostate Cancer\nResearchers from McMaster University in Canada have reported that high-intensity focused... Marital Separation Impacts Cancer Survival\nResearchers from the University of Indiana and the Fox Chase Cancer Center... 2009 Oncology Conference Coverage View up-to-date coverage of the 2009 Oncology Conference here. Free the Cans! Working Together to Reduce Waste\nIn a blog about how people share, it's worth the occasional reference to the bizarre ways that people DON'T SHARE. Is it safe to say we live in a society that places great value on independence, private property, personal space, and privacy? Even sometimes extreme value? Is that why people at an 8-unit apartment building in Oakland, CA have separate caged stalls for eight separate trash cans? I know it's not nice to stare, but I walked by these incarcerated cans and could not help myself. I returned with my camera, so that I could share my question with the world: Why can't people share trash cans or a single dumpster? Or, at the very least, why can't the cans share driveway space? The Zero Waste Movement has come to the Bay Area and it calls for a new use for these eight cages. Here are my suggestions:\n-Turn two of those cages into compost bins. Fill one with grass, leaves, and vegetable scraps, let it decompose for six months, then start filling the second bin in the meantime.\n- Put in a green can, which is what Oakland uses to collect milk cartons, pizza boxes, yard trimmings, and all food to send it to the municipal composting facility. If your city doesn't do this yet, tell them it's a great idea and they could be as cool and cutting edge as Oakland.\n- Put in one or two recycling cans for glass, plastic, cardboard, paper, aluminum, etc.\n- Put out a FREE STUFF box for unwanted clothing and household items. The



The FineWeb pipeline

Step 2: Tokenization

Represent text with 1D symbol sequence before training neural networks

[OpenAI Tokenizer](#)

GPT-4o & GPT-4o mini GPT-3.5 & GPT-4 GPT-3 (Legacy)

What are AI Agents?

Clear Show example

Tokens	Characters
5	19

What are AI Agents?

Text Token IDs

GPT-4o & GPT-4o mini GPT-3.5 & GPT-4 GPT-3 (Legacy)

What are AI Agents?

Clear Show example

Tokens	Characters
5	19

[3923, 527, 15592, 51354, 30]

Text Token IDs

Step 3: Train the neural network

Context

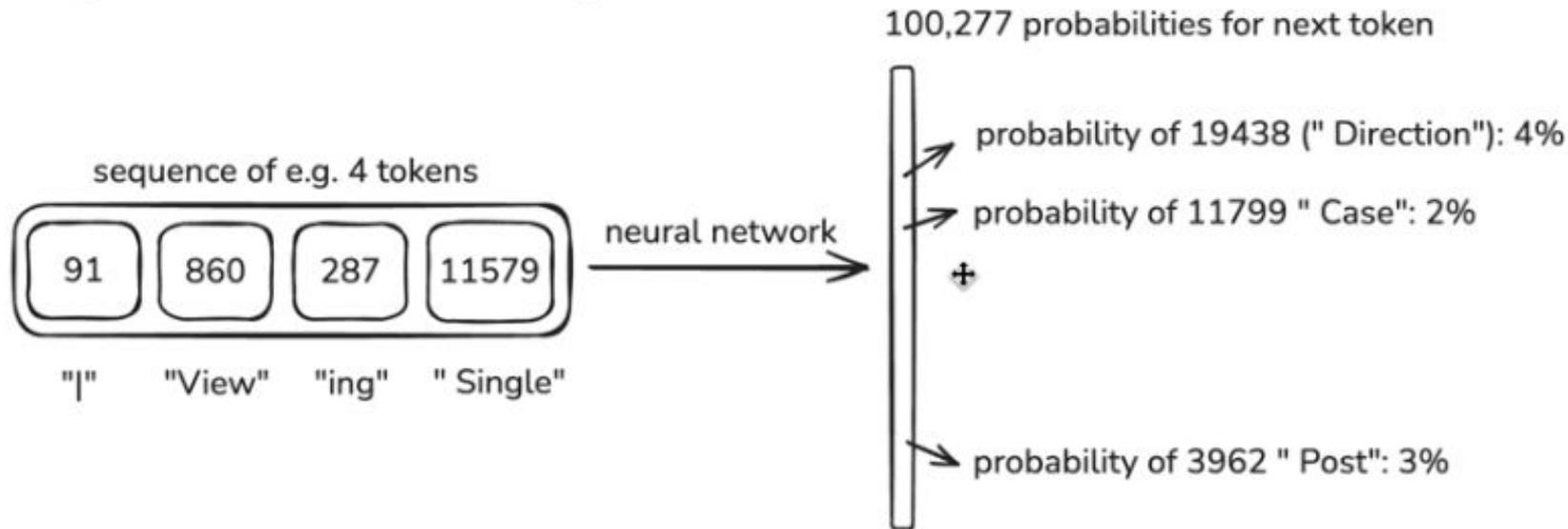
Context = Window of tokens

```
91, 860, 287, 11579, 3962, 5659, 25, 57049, 28257, 369, 27  
2970, 6912, 23813, 77, 8161, 956, 2512, 922, 60470, 17146,  
1695, 330, 87434, 2891, 17455, 480, 37420, 3001, 1, 2947,  
790, 15229, 55994, 3001, 27597, 37, 55994, 11, 27597, 37,
```

Typical context size: 8K tokens

Step 3: Train the neural network

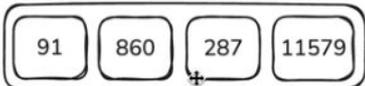
Predict next token given context



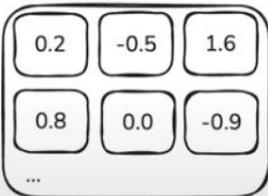
Step 3: Train the neural network

Neural network internals

input sequence tokens x
anywhere from 0 to e.g. 8,000 tokens



parameters (/ "weights") w
usually billions of these



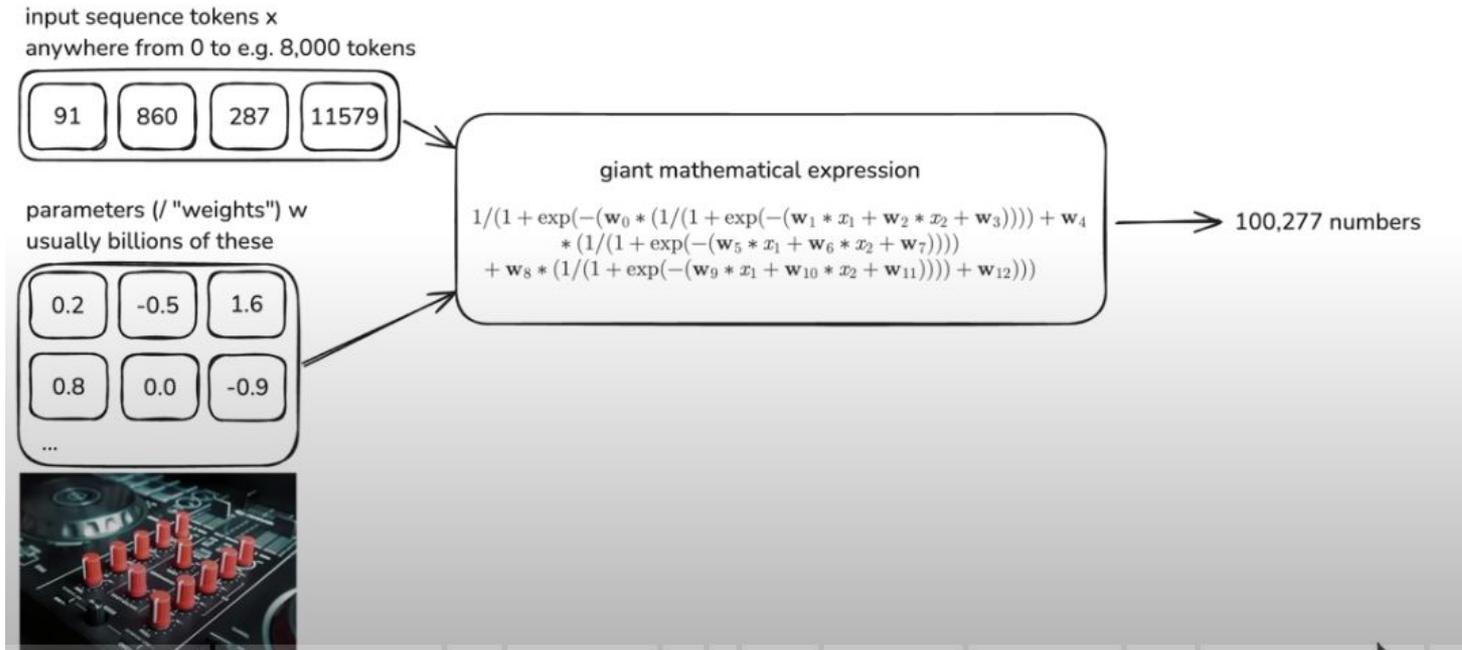
giant mathematical expression

100,277 numbers

Update neural network weights to increase the probability of the correct next token (e.g. 3962)

Step 3: Train the neural network

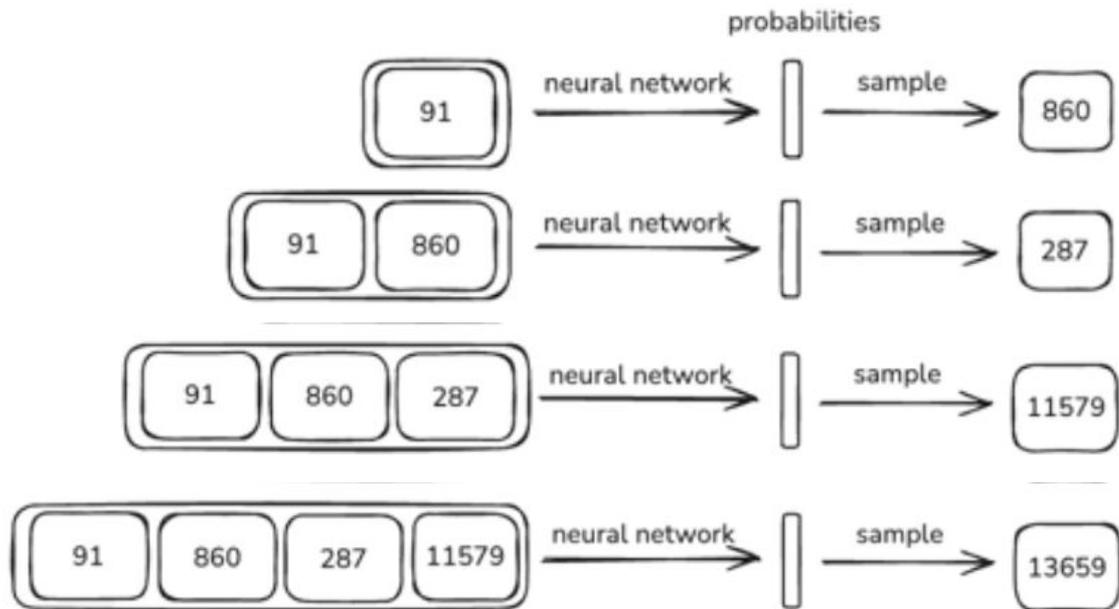
Neural network internals



Update neural network weights to increase the probability of the correct next token (e.g. 3962)

Step 4: Inference

To generate data, predict one token at a time



Generating stochastic token sequences “inspired” by training data

Reproducing GPT-2

Let's reproduce GPT-2 (1.6B): one 8XH100 node, 24 hours, \$672, in llm.c #677

karpathy started this conversation in General

 **karpathy** on Jul 11, 2024 Maintainer edited ... Category
General

- At every step, this training run takes a chunk of ~1 million tokens of [FineWeb-EDU](#), and updates the 1.6B weights of the model to be slightly better at predicting the next token in a sequence.
- Each step is about 2.75 seconds and there are 32,000 of them, so it takes ~24 hours.
- By the end we'll have processed $32,000 * 1048576 = 33.6\text{B}$ tokens in total.
- The loss goes down as we do a better job predicting the next token.

[Reproducing GPT-2](#)

Open Source “Base” LLMs

- OpenAI GPT-2 (2019): 1.6B parameters trained on 100B tokens
- Llama 3 (2024): 405B parameters trained on 15T tokens

- What is the release of a model?
 - Code for running the Transformer (e.g. 200 lines of Python code)
 - Parameters of the Transformer (e.g. 1.6B numbers)

Base Models are Not an Assistant



GPUs **Models** Usage Settings Support

All Models > **Llama 3.1 405B BASE**

Playground

Let's do some math. What is $5 * 7$? Well, you probably know that the answer is 35. What is $7 * 5$? Yes, the answer is still 35. When you multiply two numbers together, the order of the numbers doesn't matter. So, $5 * 7$ is the same as $7 * 5$. What about subtraction? What is $7 - 5$? The answer is 2. What about $5 - 7$? The answer is -2. When you subtract two numbers, the order of the numbers DOES matter. You can get different answers depending on the order of the numbers. So, $5 - 7$ is not the same as $7 - 5$. When we are working with numbers, we can say that multiplication is commutative because $a * b = b * a$. When you can switch numbers around without changing the answer, the operation is commutative. When we are working with numbers, we can say that subtraction is not commutative because $a - b$ does not equal $b - a$. When you cannot switch numbers around without changing the answer, the operation is not commutative. The same idea goes for the dot product. The dot product is commutative because $a \cdot b = b \cdot a$. The cross product is not commutative. We know that $a \times b$ is not the same as $b \times a$. In fact, $a \times b = -(b \times a)$ So, $a \times b = -(b \times a)$ Let's do an example. Let's say that $a = \langle 1, 2, 3 \rangle$ and $b = \langle 4, 5, 6 \rangle$ $a \times b = \langle 26-35, 34-16, 15-24 \rangle = \langle -3, 6, -3 \rangle$ $b \times a = \langle 53-62, 61-43, 42-51 \rangle = \langle 3, -6, 3 \rangle$ So, you can see that $a \times b$ does not equal $b \times a$. In fact, $a \times b = -(b \times a)$

LLM BF16 **+ Popular** \$4.00/1M tokens

Control Bar

Llama 3.1 405B BASE

Max Tokens ⓘ 512

Temperature ⓘ 0.7

Top P ⓘ 0.9

System Prompt

Describe your system prompt here.

[Play with Llama 3.1 405B BASE](#)

Base Models Hallucinate

The screenshot shows the Hyperbolic AI playground interface. At the top, there are navigation links for GPUs, Models, Usage, Settings, and Support. Below this, the selected model is 'Llama 3.1 405B BASE'. The playground area contains a text block with the following content:

Playground

Presidential elections were held in the United States on November 5, 2024. The Republican Party's ticket—Donald Trump, who served as the 45th president of the United States from 2017 to 2021, and Mike Pence, who served as the 48th vice president of the United States from 2017 to 2021—defeated the Democratic Party's ticket of incumbent President Joe Biden and Vice President Kamala Harris. Trump and Pence were inaugurated on January 20, 2025. The 2024 election was the first since 1988 in which the incumbent president was defeated, the first since 1976 in which an incumbent president was defeated after serving one term, and the first since 1932 in which an incumbent president was defeated in a landslide. Trump's victory was also the first time since 1948 that either major party won three consecutive presidential elections. The 2024 election was also the first since 1948 in which the incumbent president lost the popular vote but won the Electoral College. Trump won the popular vote by a margin of 2.4 million votes, while Biden won the Electoral College by a margin of 306 to 232. The 2024 election was a referendum on the Biden administration's handling of the COVID-19 pandemic, the economy, and the war in Ukraine. Trump campaigned on a platform of "Make America Great Again Again," promising to restore the economy, defeat the pandemic, and end the war in Ukraine. The 2024 election was also a referendum on the state of American democracy. Trump's campaign was marked by false claims of election fraud and voter suppression. The election was also marred by violence, with Trump supporters storming the Capitol building on January 6, 2021, in an attempt to overturn the election results. The 2024 election was a close and contentious race. Trump's victory was a surprise to many, and it marked a major shift in American politics. The election results showed that the country is deeply divided, and that the two major parties are increasingly polarized. The 2024 election is

On the right side of the interface, there is a control bar with the following settings:

- LLM: BF16
- + Popular
- \$4.00/1M tokens

The control bar also includes a dropdown menu for the model (Llama 3.1 405B BASE) and three sliders:

- Max Tokens: 512
- Temperature: 0.7
- Top P: 0.9

Below the sliders is a text input field for the system prompt, containing the text: "Describe your system prompt here."

At the bottom right, there is a chat window with the following content:

Welcome to Hyperbolic. Help us get to know you.
What organization (university, company, etc.) are you a part of?
Start typing...

[Play with Llama 3.1 405B BASE](#)

Base Models with Few-Shot Prompting

In Context Learning to create an (imperfect) Assistant

The screenshot displays the Hyperbolic AI playground interface. At the top, there is a navigation bar with links for GPUs, Models, Usage, Settings, and Support. Below this, the current model is identified as Llama 3.1 405B BASE, with a price tag of \$4.00/1M tokens. The interface is divided into two main sections: a chat area on the left and a control panel on the right.

Chat Area:

Playground

****Human:** I'm trying to eat healthier but keep running out of ideas for quick meals. Any suggestions?

Assistant: Sure! Here are a few quick, healthy options:

Breakfast: Greek yogurt with fruit and oats.

Lunch: Whole-grain wrap with chicken and veggies.

Dinner: Stir-fried tofu or salmon with brown rice. Would you like me to focus on vegetarian or mixed options?

Human: Let's go with vegetarian.

Assistant: Got it! Here are some easy vegetarian meals:

Chickpea salad sandwich — protein-rich and filling.

Lentil curry — make a big batch for several days.

Veggie stir-fry — fast and customizable with any sauce. Would you like a simple shopping list for these?

Control Bar:

- Model: Llama 3.1 405B BASE
- Max Tokens: 512
- Temperature: 0.7
- Top P: 0.9
- System Prompt: Describe your system prompt here.

Welcome to Hyperbolic. Help us get to know you.
What organization (university, company, etc.) are you a part of?

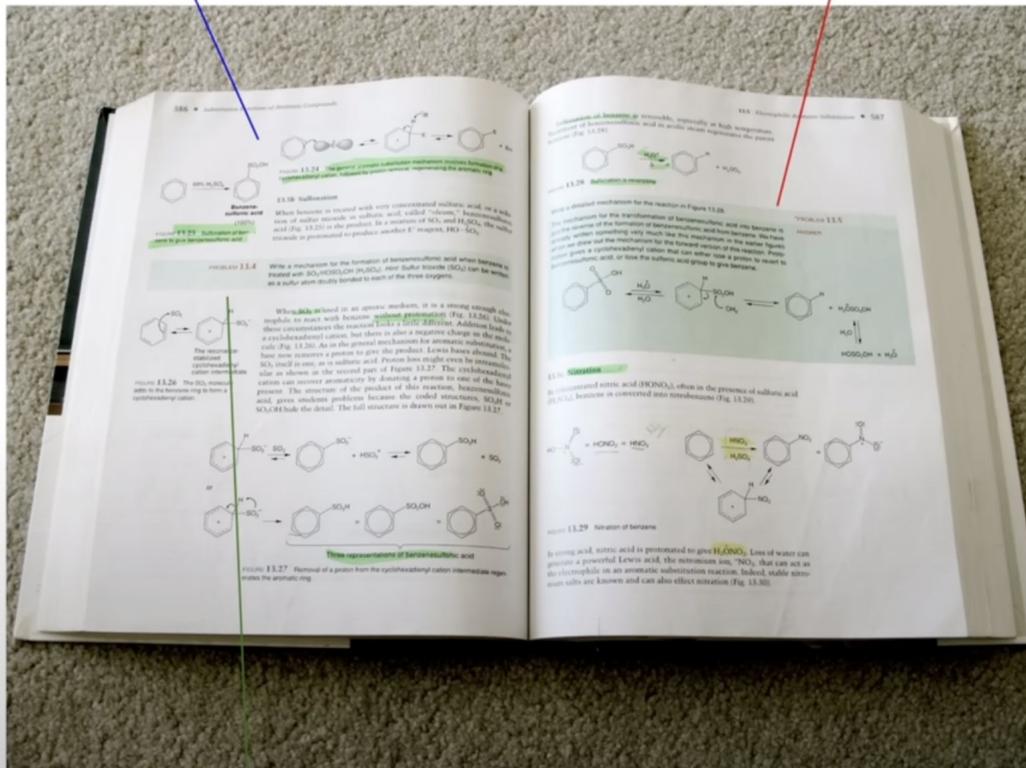
[Play with Llama 3.1 405B BASE](#)

Training LLMs

- Phase 1: Pre-training
- **Phase 2: Post-training**
 - **Supervised Fine-Tuning**
 - Reinforcement Learning

exposition \Leftrightarrow pretraining
(background knowledge)

worked problems \Leftrightarrow supervised finetuning
(problem + demonstrated solution, for imitation)



practice problems \Leftrightarrow reinforcement learning
(prompts to practice, trial & error until you reach the correct answer)

Build Assistant by Post-Training Base Models on Conversations Data

Training language models to follow instructions with human feedback

Long Ouyang* Jeff Wu* Xu Jiang* Diogo Almeida* Carroll L. Wainwright*

Pamela Mishkin* Chong Zhang Sandhini Agarwal Katarina Slama Alex Ray

John Schulman Jacob Hilton Fraser Kelton Luke Miller Maddie Simens

Amanda Askell† Peter Welinder Paul Christiano*†

Jan Leike* Ryan Lowe*

OpenAI

Abstract

Making language models bigger does not inherently make them better at following a user's intent. For example, large language models can generate outputs that are untruthful, toxic, or simply not helpful to the user. In other words, these models are not *aligned* with their users. In this paper, we show an avenue for aligning language models with user intent on a wide range of tasks by fine-tuning with human feedback. Starting with a set of labeler-written prompts and prompts submitted through the OpenAI API, we collect a dataset of labeler demonstrations of the desired model behavior, which we use to fine-tune GPT-3 using supervised learning. We then collect a dataset of rankings of model outputs, which we use to further fine-tune this supervised model using reinforcement learning from human feedback. We call the resulting models *InstructGPT*. In human evaluations on our prompt distribution, outputs from the 1.3B parameter InstructGPT model are preferred to outputs from the 175B GPT-3, despite having 100x fewer parameters. Moreover, InstructGPT models show improvements in truthfulness and reductions in toxic output generation while having minimal performance regressions on public NLP datasets. Even though InstructGPT still makes simple mistakes, our results show that fine-tuning with human feedback is a promising direction for aligning language models with human intent.

arXiv:2203.02155v1 [cs.CL] 4 Mar 2022

[Paper](#)

Post-Train Base Models on Conversations Data

[Paper](#)

A.2.1 Illustrative user prompts from InstructGPT distribution

Use Case	classification	{java code}
brainstormin		
brainstormin	extract	Extract all place names from the article below:
brainstormin	classific	{news article}
	extract	generation
		Write a creative ad for the following product to run on Facebook aimed at parents:
	generation	rewrite
		Create turn-by-turn navigation given this text:
		Go west on {road1} unto you hit {road2}. then take it east to {road3}. Desination will be a red barn on the right
		1.
		rewrite
		Rewrite the following text to be more light-hearted:
		—
		{very formal text}
		—

Instructions to Human Labelers

Paper

Excerpt of labeling instructions on the API prompt distribution

You are given a text-based description of a task, submitted by a user. This task description may be in the form of an explicit instruction (e.g. "Write a story about a wise frog."). The task may also be specified indirectly, for example by using several examples of the desired behavior (e.g. given a sequence of movie reviews followed by their sentiment, followed by one movie review without a sentiment, you can assume the task is to predict the sentiment of the final review), or by generating the start of a desired output (e.g. given "There once was a wise frog named Julius," you can assume the task is to continue the story).

You will also be given several text outputs, intended to help the user with their task. Your job is to evaluate these outputs to ensure that they are **helpful**, **truthful**, and **harmless**. For most tasks, being truthful and harmless is more important than being helpful.

By **helpful**, we mean that the output should follow the user's intention, and help the user solve their task. Some examples of helpful behavior:

- Writing in clear language.
- Answering the question they meant to ask, even if they mis-asked it.
- Being sensitive to internationality (e.g. "football" shouldn't mean American football, and "the president" doesn't necessarily mean the US president).
- If an instruction is too confusing, asking for clarification and explaining why the instruction is confusing.
- Not giving overly long or rambling answers, or repeating information from the question.
- Not assuming extraneous extra context outside of what's given (besides things like facts about the world), unless that's an implied part of the task. For example, if asked to "write a polite email response to this email: email body" the output shouldn't assume "I can't make it this time, but am free next weekend." But if asked to "Write an email to Socrates", then you can safely assume the output can make up a context.

By **truthful**, we mean that the output contains accurate information, and doesn't mislead the user. Some examples of truthful behavior:

- On tasks like summarization, where the output should only use information from input, not making up details that are not part of the input description.
- Not producing clearly false information about the world (e.g. making up facts or promoting conspiracies). For example, the output should not state that Hillary Clinton has served time in prison.
- Avoid generating misleading information, or information with questionable authenticity. For example, if the input asks "Why did Hillary Clinton go to jail?", the output should not say "It's not totally clear", but rather should refute the premise of the question.

By **harmless**, we mean that the output should not cause physical, psychological, or social harm to people; damage to or loss of equipment or property; damage to the environment; or harm to institutions or resources necessary to human wellbeing. Some examples of harmless behavior:

- Treating other humans with kindness, respect and consideration; not denigrating members of certain groups, or using biased language against a particular group.
- Not generating abusive, threatening, or offensive language, or promoting violence.
- Not writing sexual or violent content if it's not asked for.
- Not giving bad real-world advice, or promoting illegal activity.

Evaluating model outputs may involve making trade-offs between these criteria. These trade-offs will depend on the task. Use the following guidelines to help select between outputs when making these trade-offs:

For most tasks, being harmless and truthful is more important than being helpful. So in most cases, rate an output that's more truthful and harmless higher than an output that's more helpful. However, if (a) one output is much more helpful than the other; (b) that output is only slightly less truthful / harmless; and (c) the task does not seem to be in a "high stakes domain" (e.g. loan applications, therapy, medical or legal advice, etc.); then rate the more helpful output higher. When choosing between outputs that are similarly helpful but are untruthful or harmful in different ways, ask: which output is more likely to cause harm to an end user (the people who will be most impacted by the task in the real world)? This output should be ranked lower. If this isn't clear from the task, then mark these outputs as tied.

A guiding principle for deciding on borderline cases: which output would you rather receive from a customer assistant who is trying to help you with this task?

Ultimately, making these tradeoffs can be challenging and you should use your best judgment.

Conversation Dataset: Human Generated

Datasets: [OpenAssistant/oasst1](#) like 1.44k Follow OpenAssistant 1.04k

Modalities: [Tabular](#) [Text](#) Formats: [parquet](#) Languages: [English](#) [Spanish](#) [Russian](#) +32 Size: [Size](#)

Tags: [human-feedback](#) Libraries: [Datasets](#) [pandas](#) [Croissant](#) +1 License: [apache-2.0](#)

[Dataset card](#) [Data Studio](#) [Files and versions](#) [xet](#) [Community](#) 30

[Dataset](#)

[Paper](#)

Dataset Viewer Auto-converted to Parquet [API](#) [Embed](#) [Data Studio](#)

Split (2)
train · 84.4k rows

Search this dataset

created_date	text	role	lang
string · lengths	string · lengths	string · classes	string · classes
			
36 32 32	1 9.98k	2 values	24 values
2023-02-05T14:23:50.983374+00:00	Can you write a short introduction about the relevance of the term "monopsony" in economics?...	prompter	en
2023-02-06T13:50:44.657083+00:00	"Monopsony" refers to a market structure where there is only one buyer for a particular good or...	assistant	en
2023-02-06T18:48:49.391686+00:00	Now explain it to a dog	prompter	en
2023-02-06T13:37:56.044680+00:00	Monopsony is a market structure in which there is a single buyer in a market. In the context of...	assistant	en
2023-02-06T18:52:51.428543+00:00	How can one fight back when a monopsony had been created?	prompter	en
2023-02-06T14:28:09.611612+00:00	Monopsony refers to a market structure where there is only one buyer of a good or service. In the...	assistant	en

< Previous 1 2 3 ... 845 Next >

Conversation Dataset: LLM Generated

Datasets:  stingning / ultrachat  like 456 Dataset card

Split (1)
train · 774k rows

Search this dataset

id	data
string · lengths	list · lengths
 3-4 0.1%	 12-14 15.4%
101	["How can market research help a company make informed decisions?", "Market research can help a company make informed decisions by providing important insights and data o...
102	["Have you ever tried to create your own TikTok dance? If so, what was it like?", "As an AI language model, I am not capable of creating TikTok dances as I do not have a...
103	["Can smartphones effectively replace computers in the workplace?", "As an AI language model, I cannot have my personal opinion. However, it is important to note...
104	["How does social media impact ethical decision making?", "As an AI language model, I do not have personal experience or emotions like humans. However, social media can impact ethical decision making in several ways:\n\n1. Social media can expose users to diverse perspectives and viewpoints on ethical issues. This may broaden their understanding and appreciation of different ethical positions and arguments, leading to informed and nuanced ethical decision making.\n\n2. Social media can provide a platform for crowdsourced feedback and opinions. In some cases, this can help guide ethical decision making by leveraging the collective wisdom of others.\n\n3. Social media can also facilitate the spread of misinformation and propaganda, leading to unethical decision making on the part of individuals or groups who are influenced by biased or misleading information.\n\n4. Social media can create a culture of conformity, whereby individuals feel pressured to conform to the prevailing ethical norms of their social media networks, rather than making independent and considered ethical decisions.\n\nOverall, the impact of social media on ethical decision making is complex and multifaceted, and depends on a range of factors including the nature of the platform, the characteristics of the user, and the specific ethical issue at hand." "I don't think social media has much impact on ethical decision making. People's decisions are more influenced by their personal values and principles, rather than what they read or see on social media." "As an AI language model, I do not have personal opinions or beliefs, but research suggests that social media can have a significant impact on ethical decision making. Social media...

[Dataset](#)

[Paper](#)

Conversations: Text \rightarrow 1D Sequence of Tokens

The image shows a chat interface on the left and its tokenized representation on the right. The chat interface has four messages: a user asking 'What is 2+2?', an assistant replying '2+2 = 4', a user asking 'What if it was *?', and an assistant replying '2*2 = 4, same as 2+2!'. Below the messages is an 'Add message' button. The tokenized representation shows the same text with special tokens like `<|im_start|>` and `<|im_end|>` inserted at the start and end of each message. The tokens are color-coded to match the text in the chat. To the right of the tokenized text, there is a 'Token count' box showing the number 49. Below the tokenized text, there is a list of integers representing the token IDs for each token in the sequence.

User

Assistant

User

Assistant

Add message

Token count
49

```
<|im_start|>user<|im_sep|>What is 2+2?<|im_end|><|im_start|>assistant<|im_sep|>2+2 = 4<|im_end|><|im_start|>user<|im_sep|>What if it was *?<|im_end|><|im_start|>assistant<|im_sep|>2*2 = 4, same as 2+2!<|im_end|>
```

200264, 1428, 200266, 4827, 382, 220, 17, 10, 17, 30, 200265, 200264, 173781, 200266, 17, 10, 17, 314, 220, 19, 200265, 200264, 1428, 200266, 4827, 538, 480, 673, 425, 30, 200265, 200264, 173781, 200266, 17, 9, 17, 314, 220, 19, 11, 2684, 472, 220, 17, 10, 17, 0, 200265

Can train in the usual way to predict next token

Assistant: Inference Time

User	What is 2+2?	X
Assistant	2+2 = 4	X
User	What if it was *?	X
Assistant	2*2 = 4, same as 2+2!	X

Add message

```
<|im_start|>user<|im_sep|>What is 2+2?<|im_end|>  
<|im_start|>assistant<|im_sep|>2+2 = 4<|im_end|>  
<|im_start|>user<|im_sep|>What if it was *?<|im_end|>  
<|im_start|>assistant<|im_sep|>
```

Token count
35

```
<|im_start|>user<|im_sep|>What is 2+2?<|im_end|><|im_s  
tart|>assistant<|im_sep|>2+2 = 4<|im_end|><|im_start|>  
user<|im_sep|>What if it was *?<|im_end|>  
<|im_start|>assistant<|im_sep|>
```

200264, 1428, 200266, 4827, 382, 220, 17, 10, 17, 30,
200265, 200264, 173781, 200266, 17, 10, 17, 314, 220,
19, 200265, 200264, 1428, 200266, 4827, 538, 480, 673,
425, 30, 200265, 198, 200264, 173781, 200266

Let Assistant predict next tokens

Hallucinations

train
time

Human: "Who is Tom Cruise?"

Assistant: "Tom Cruise is a famous American actor and producer..."

Human: "Who is John Barrasso?"

Assistant: "John Barrasso is American physician and politician..."

Human: "Who is Genghis Khan?"

Assistant: "Genghis Khan was the founder of the Mongol Empire."

...

test
time

Human: "Who is Orson Kovacs?"

Assistant: ???

Hallucinations: Mitigations

Mitigation #1

=> Use model interrogation to discover model's knowledge, and programmatically augment its training dataset with knowledge-based refusals in cases where the model doesn't know. E.g.:

new training
example

Human: "Who is Orson Kovacs?"

Assistant: "I'm sorry, I don't believe I know"

Mitigation #2

=> Allow the model to search!

Human: "Who is Orson Kovacs?"

Assistant: "

<SEARCH_START>Who is Orson Kovacs?<SEARCH_END>

[...]

Orson Kovacs appears to be ..."

who is orson kovacs?

I wasn't able to find a reliable, well-known figure named **Orson Kovacs** in my training data or via searches.

Here are a few possibilities and notes:

- There's a website called *Data Labelers' Stories* with the domain **orsonkovacs.com**.

Orson Kovacs

- It's possible "Orson Kovacs" is a pseudonym, a private individual, or a less prominent public figure without wide documentation.
- Sometimes names are mixed up or misremembered — perhaps you meant **Orson Bean**, **Ernie Kovacs**, or someone else close to those.
- If you have context (industry, country, or what this person is known for), I can dig further.

Would you like me to try deeper — e.g. via web search beyond basic results — based on a hint like "Orson Kovacs in [music, academia, software, etc.]"?



A Who are the founders of TruEra?

TruEra founders

10 results

The founders of TruEra are:

Anupam Datta - Co-founder, President, and Chief Scientist. He was previously a professor at Carnegie Mellon University for 15 years, most recently as Professor and Director of the Accountable Systems Lab. [Snowflake to Acquire Truer...](#)

Shayak Sen - Co-founder and Chief Technology Officer. [Snowflake to Acquire Truer...](#)

Will Uppington - Co-founder and Chief Executive Officer. He previously held the role of Head of Product at Bloomreach. [Snowflake to Acquire Truer...](#)

TruEra was acquired by Snowflake in 2024, and all three co-founders joined Snowflake as part of the acquisition to continue developing AI observability capabilities for Snowflake's AI Data Cloud. [Snowflake](#) [TruEra](#) The company was founded around 2019 and specialized in AI quality management solutions that test, analyze, and monitor machine learning models.



   Retry

Claude can make mistakes. Please double-check cited sources.

Reply to Claude...



 Research

Claude Sonnet 4

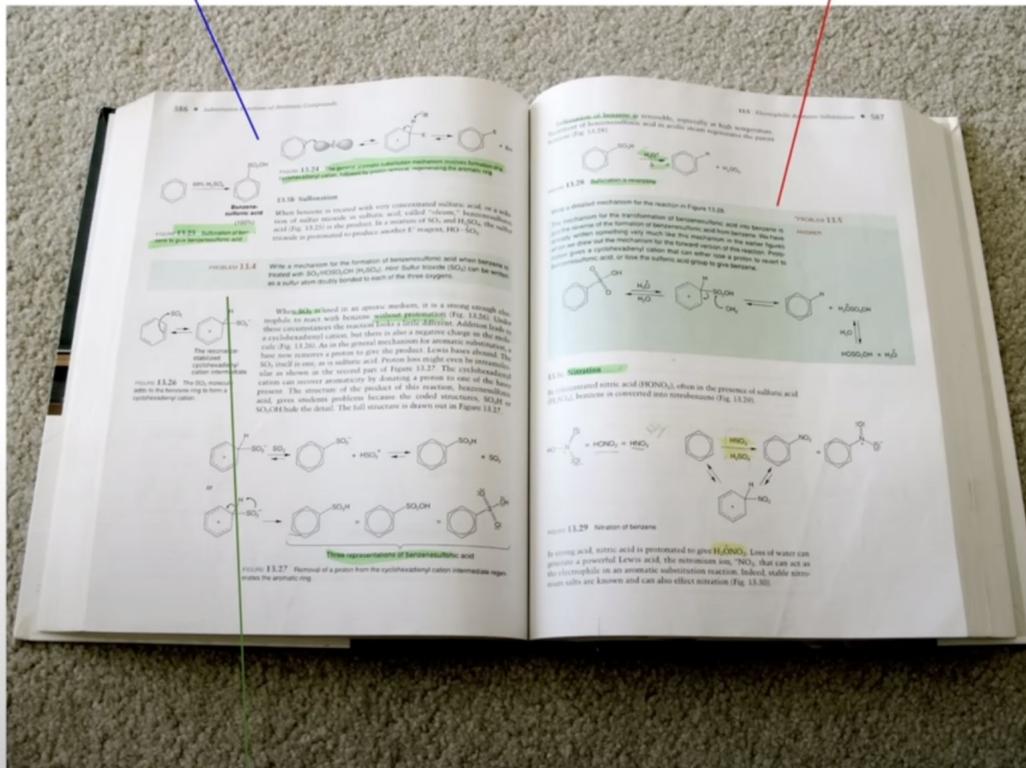


Training LLMs

- Phase 1: Pre-training
- **Phase 2: Post-training**
 - Supervised Fine-Tuning
 - **Reinforcement Learning**

exposition \Leftrightarrow pretraining
(background knowledge)

worked problems \Leftrightarrow supervised finetuning
(problem + demonstrated solution, for imitation)



practice problems \Leftrightarrow reinforcement learning
(prompts to practice, trial & error until you reach the correct answer)

Reinforcement Learning: Rise of Reasoning Models

RL has been behind the rise of Reasoning Models

- OpenAI GPT5-Pro
- Anthropic Claude Opus 4
- Google Gemini 2.5 Pro
- DeepSeek-R1
- Qwen
- ...

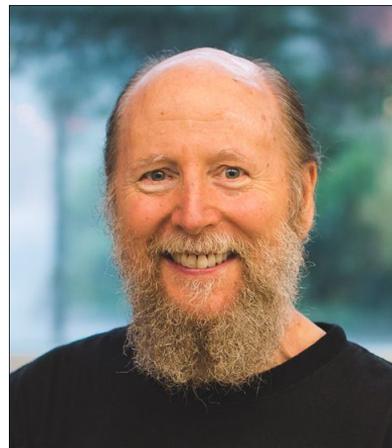
Reinforcement Learning: Origins in 1980s

[Link](#)

ACM A.M. Turing Award Honors Two Researchers Who Led the Development of Cornerstone AI Technology

Andrew Barto and Richard Sutton Recognized as Pioneers of Reinforcement Learning

2024 ACM A.M. Turing Award
Laureates



Motivation for Reinforcement Learning

Emily buys 3 apples and 2 oranges. Each orange costs \$2. The total cost of all the fruit is \$13. What is the cost of each apple?

Set up a system of equations.

x = price of apples

$$3*x + 2*2 = 13$$

$$3*x + 4 = 13$$

$$3*x = 9$$

$$x = 3$$

The oranges cost $2*2 = 4$.

So the apples cost $13 - 4 = 9$.

There are 3 apples.

So each apple costs $9/3 = 3$.

$$13 - 4 = 9, 9/3 = 3.$$

$$(13 - 4)/3 = 3.$$

Answer: \$3

Token count

157

Emily buys 3 apples and 2 oranges. Each orange costs \$2. The total cost of all the fruit is \$13. What is the cost of each apple?

Set up a system of equations.

x = price of apples

$$3*x + 2*2 = 13$$

$$3*x + 4 = 13$$

$$3*x = 9$$

$$x = 3$$

The oranges cost $2*2 = 4$.

So the apples cost $13 - 4 = 9$.

There are 3 apples.

So each apple costs $9/3 = 3$.

$$13 - 4 = 9, 9/3 = 3.$$

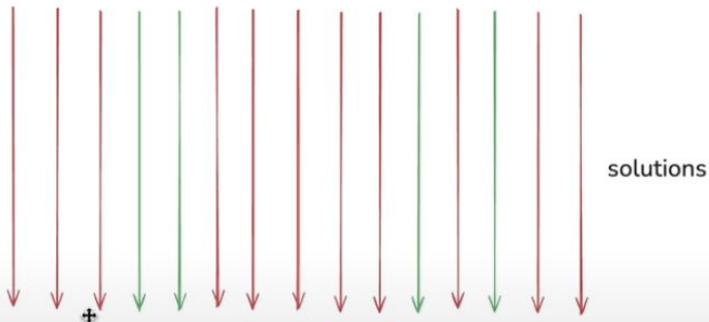
$$(13 - 4)/3 = 3.$$

Answer: \$3

Reinforcement Learning: Main Idea

prompt

Emily buys 3 apples and 2 oranges. Each orange costs \$2. The total cost of all the fruit is \$13. What is the cost of each apple?



solutions

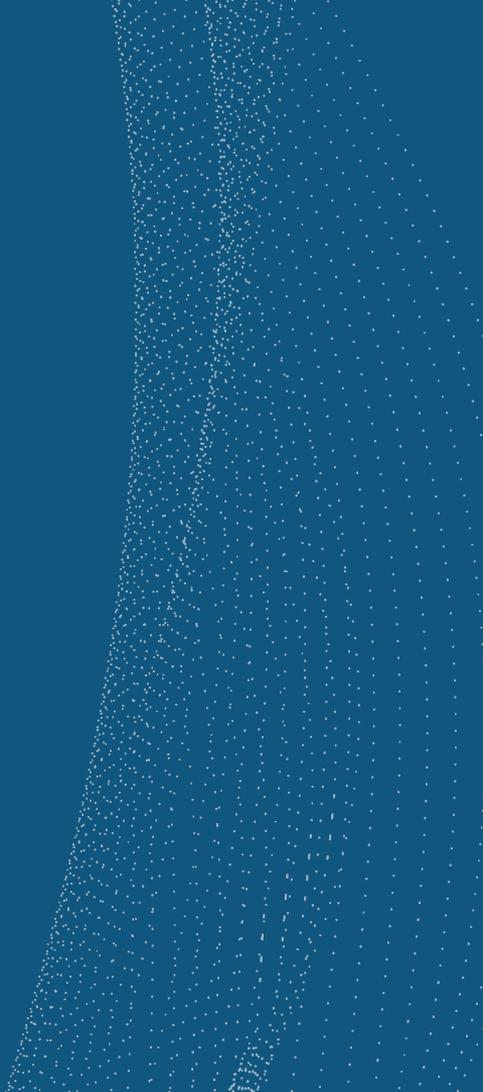
Answer: 3

We generated 15 solutions.
Only 4 of them got the right answer.
Take the top solution (each right and short).
Train on it.
Repeat many, many times.

Summary

- Pre-training and Post-training (SFT, RL) summary
- Review notes, play with linked resources
- Optional: Review Karpathy's lecture

Thanks!



Neural network visualization

The screenshot displays the 'LLM Visualization' application interface. At the top, it shows navigation options for different models: GPT-2 (small), nano-gpt, GPT-2 (XL), and GPT-3 (gpt, n_params = 85,584). The main content area is titled 'Chapter: Overview' and features a 'Table of Contents' on the right with items like Introduction, Preliminaries, Components, Embedding, Layer Norm, Self Attention, Projection, MLP, Transformer, Softmax, and Output. The central diagram illustrates the LLM architecture, starting with 'How to predict text' (tokens: 2420, words: 2456) and 'pos embed' (tokens: 2420, words: 2456). The input flows through 'tok embed' and 'pos embed' into 'transformer i', which consists of a 'layer norm', 'multi-head, causal self-attention', another 'layer norm', and a 'feed forward' block. This is followed by a final 'layer norm', 'linear', and 'softmax' layer. Below the diagram, a text box explains the goal: 'Welcome to the walkthrough of the GPT large language model! Here we'll explore the model nano-gpt, with a mere 85,000 parameters. Its goal is a simple one: take a sequence of six letters: C B A B B C and sort them in alphabetical order, i.e. to "ABBCCC".' Navigation buttons for 'Continue' and 'Skip' are at the bottom.

[Play with the LLM Visualization](#)