

A Glimpse Into LLM, RL and ChatGPT

hululu.zhu@gmail.com

April 2023

Agenda today

- First Hour
 - Large Language Model (LLM) Foundation
 - Reinforcement Learning (RL) Essentials
 - ChatGPT Unveiled
 - ChatGPT-like AI Frontier Applications
 - Societal Impacts
 - Q&A
- Next Half Hour
 - More discussion

Disclaimer

- This talk is my personal voluntary effort, prepared and conducted during my personal time outside of working hours.
- All content is derived from publicly available sources, and the views expressed herein only represent my personal opinions, and do not reflect the positions of DeepMind®, Google®, or Alphabet®

hululu.zhu@gmail.com

April 2023

About me



Student

Intern

Software Eng

Research Eng

Accessibility

Engineering

Data Science

Deep Learning

LLM Foundation: Deep Learning and Transformer-based LLMs

AI, Machine Learning, and Deep learning

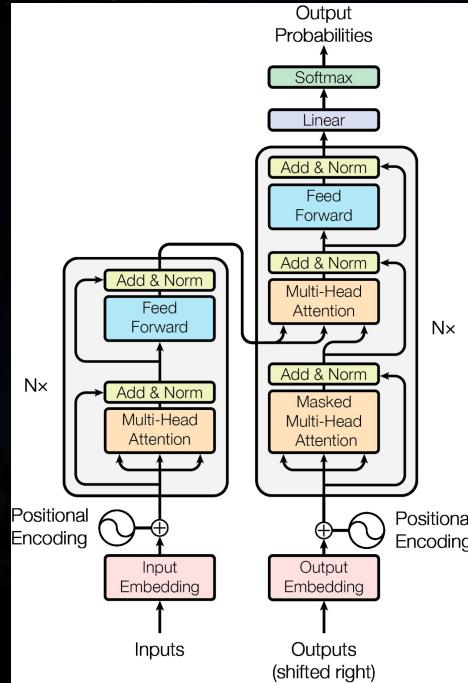
Artificial Intelligence

Machine Learning

Deep Learning

- LLM, deep RL, ChatGPT, diffusion models, all fit here 

Modern LLM building blocks: Transformer Architecture



Transformer as dominating architecture for NLP since 2018

- Multi-head attention
- Encoder-Decoder
- Embedding layers
- Positional encoding
- Cross-Attention in decoder layers
- Output Softmax

Note: Tokenization (e.g. wordpiece, sentencePiece, BPE) is needed (outside Transformer) to convert text to token ids

Note: Sometimes we call it XFormer since there are many variations to the original Transformer

Language Models (LM) and Large Language Models (LLM)

LM for understanding (e.g. BERT)

- Text in
- Embedding (numeric representation of understanding) out
 - The Embedding can be connected to other output heads for tasks like classification or regression

LM for generation (e.g. GPT or T5 or OpenAI ChatGPT or Google Bard)

- Text in
- Text out

* In most cases, **LLM** refers to **huge** (e.g. >1B params) Deep Learning LM for **generation**

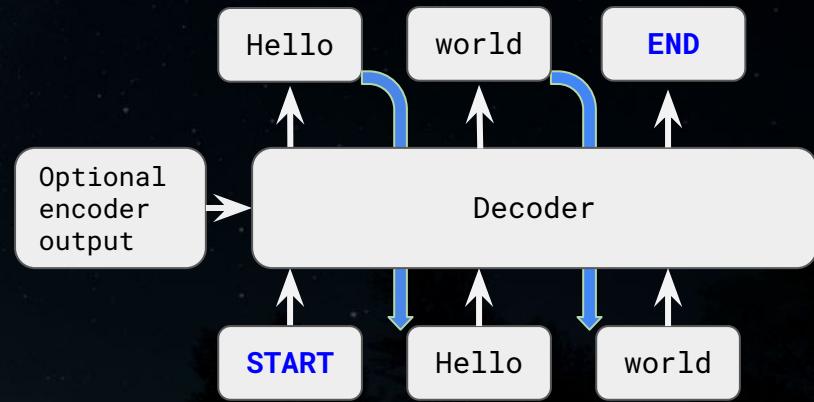
LLM Intro: Training Objectives for LLMs [in pretraining]?

- Fill the blanks (aka masks) for “Masked Language Models” (e.g. [BERT](#))
 - **Ground Truth:** “Paris is a beautiful city”
 - **X:** “Paris is a [MASK] city”
 - **Y:** “beautiful”
 - **Model:** “good”
 - **Optimize:** “good” “beautiful”
- Predict the next text given prompt, for “Generative Language Models” (e.g. [GPT](#))
 - **Aka Causal LM**
 - **Ground Truth:** “Paris is a | beautiful city”
 - **X:** “Paris is a”
 - **Y:** “beautiful”
 - **Model:** “good”
 - **Optimize:** “good” “beautiful”
 - **Ground Truth:** “Paris is a beautiful city”
 - **X:** “Paris is a beautiful”
 - **Y:** “city”
 - **Model:** “place”
 - **Optimize:** “place” “city”
- The “[Self-supervised](#)” Learning Paradigm
 - It is supervised (given x, predict y)
 - It does NOT require expensive human labels (more precisely, this statement is only true for pre-training)

Decoding/Generating Algorithms in Generative LLMs

Decode token by token, left to right. A new output token is appended as next token's decoder input

- Beam Search
 - Maintain a max size of searching “beams (paths)” to get best overall best beam
- Sampling
 - Sampling based on probabilities
- Greedy
 - Select the argmax(prob) token at every position
- Top-k, Top-p and more



<https://huggingface.co/blog/how-to-generate>

Quick Walkthrough of Selected LLMs

- [Google BERT](#)
- [OpenAI GPT 1, 2, & 3](#)
- [Google T5](#)
- [Google LaMDA](#)
- [Google PaLM](#)
- [NVidia Megatron LM](#)
- [OpenAI WebGPT](#)
- [DeepMind Chinchilla](#)
- [Tsinghua Univ GLM 130B](#)
- [OpenAI InstructGPT](#) (ChatGPT)
- [Anthropic RLHF LLM](#) and [RLAIF LLM](#)
- [Facebook/Meta OPT](#)
- [Facebook/Meta LLaMA](#)
- [OpenAI GPT4](#) (eval report, no tech detail)
- [BigScience bloom 176B](#)
- [Stanford Alpaca](#)
- [Baidu Ernie 3.0 Titan](#)
- [BloombergGPT 176B](#)

Selected Important LLM concept

- Pretraining, finetuning, prompt engineering, prompt tuning
- Scaling laws
- Emergent Abilities
- Chain of Thoughts
- Hallucination

Check out [this deck](#) to include more LLM concepts and examples

Guess what current most powerful LLMs not so good at?

Selected Important LLM concept

- Pretraining, finetuning, prompt engineering, prompt tuning
- Scaling laws
- Emergent Abilities
- Chain of Thoughts
- Hallucination

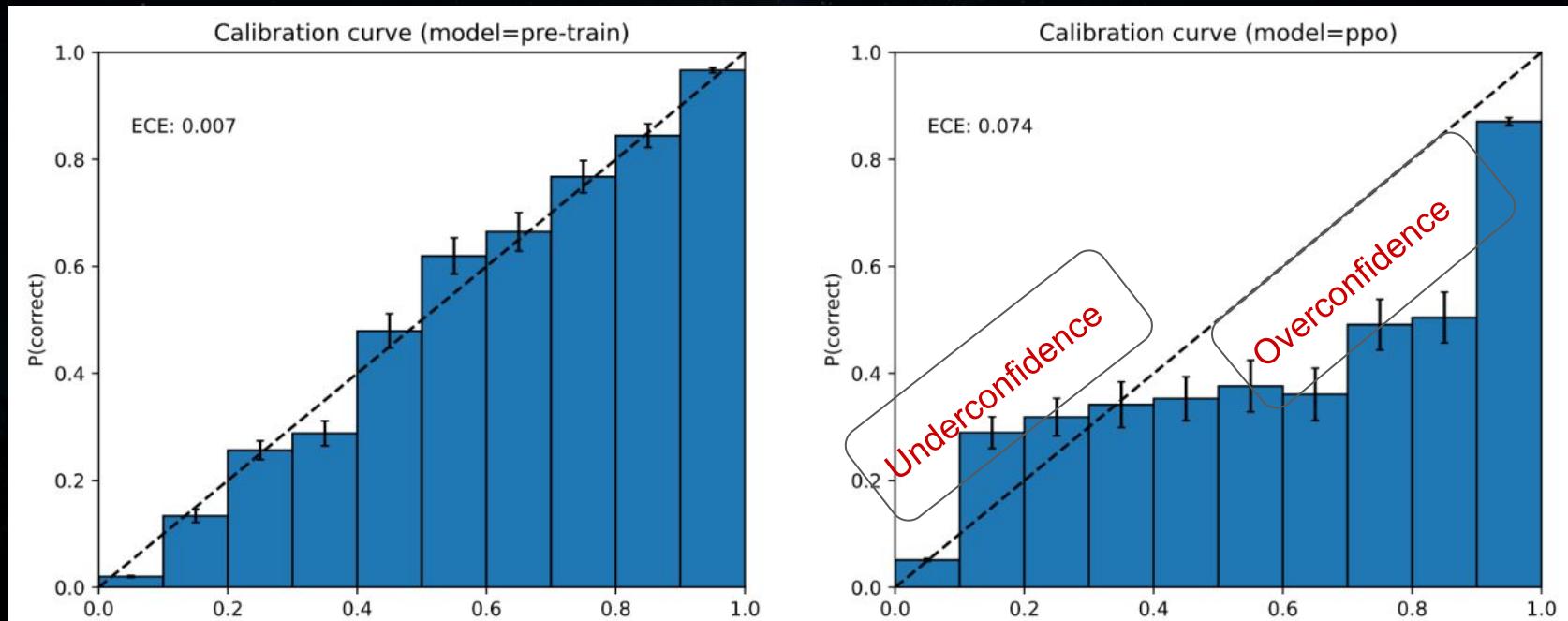
Check out [this deck](#) to include more LLM concepts and examples

Guess what current most powerful LLMs not so good at?

- My personal take: “LLMs do not know what they do not know”

Re “LLMs do not know what they do not know”

Misaligned Calibration for GPT4 after RLHF (from [GPT4 tech report](#))



Success of modern LLMs comes from

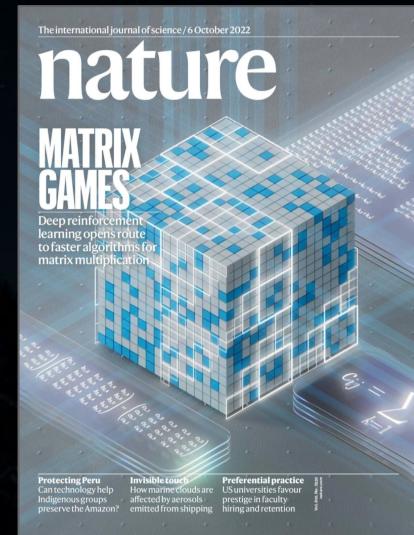
- **Large Data:** Common Crawl, webtexts, books, and Wikipedia
- **Benchmarks:** GLUE, SuperGLUE, BIG-bench
- **Improved Infra:** GPU/TPU, TensorFlow/PyTorch/JAX, Cloud service
- **Architecture:** seq2seq, Transformer
- **Invest:** Google, OpenAI, now more
- **EcoSystem:** HuggingFace, arxiv, github
- **Leaders:** Ilya Sutskever, Geoffrey Hinton, Yoshua Bengio, Yann LeCun, Fei-Fei Li, Demis Hassabis, Jeff Dean, and more

Reinforcement Learning Essentials: Foundational Basics and PPO Algorithm

Selected Success Stories of RL

- AlphaGo, AlphaStar, AlphaTensor by DeepMind

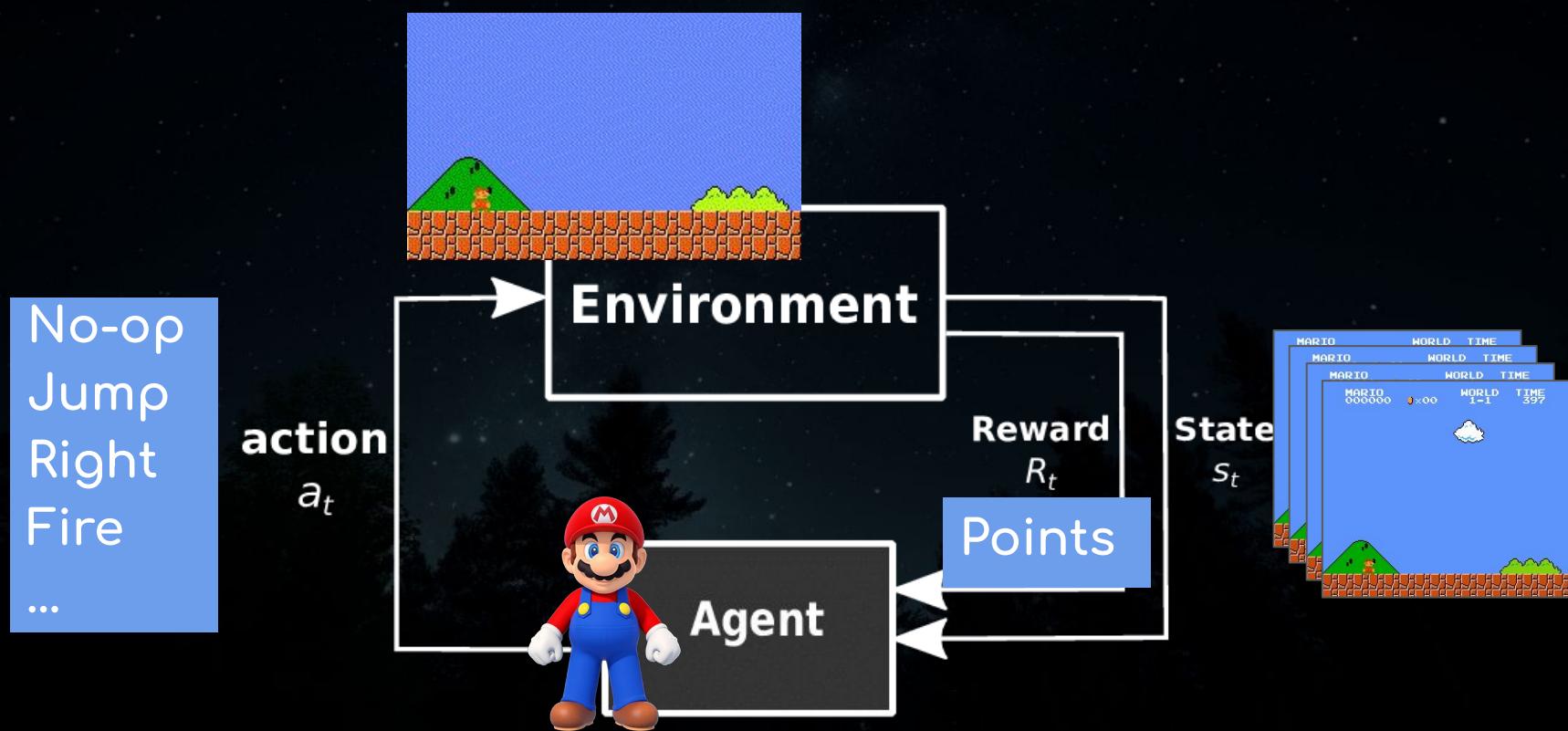
Check out [this deck](#) for a summary of more Alpha* papers by DeepMind



What is Reinforcement learning?



What is Reinforcement learning? Cont (Mario case)



(1/4) RL optimization algorithm explained

Q-learning:

- Action Value Function $Q(s, a)$
 - Given state s
 - Which action a shall we take?
 - So that it will lead to optimal expected total (delayed) rewards!
- Often overestimate the expected optimal reward

(2/4) RL optimization algorithm explained

Policy Gradient $\pi(a | s)$

- A policy tells which action a to take on state s
 - That implicitly optimized for better [delayed] total rewards
- Uses 1st order derivative for linear search, thus leads to unstable improvements



Jonathan Hui: RL — Trust Region Policy Optimization (TRPO) Explained

(3/4) RL optimization algorithm explained

TRPO Trust Region Policy Optimization

- Trust Region: Region with radius δ to avoid bad big moves
- Uses MM (minorize maximization) and Advantage function (expected rewards over average actions)
- Constrain of KL-divergence between old and new parameters
 - Ensures new policy is not drastically different from the current
- Often considered computationally expensive (because of inverse of hessian), and constrained by linear and quadratic approximations following static defined formula (conjugate gradient approximation)

(4/4) RL optimization algorithm explained

PPO Proximal Policy Optimization

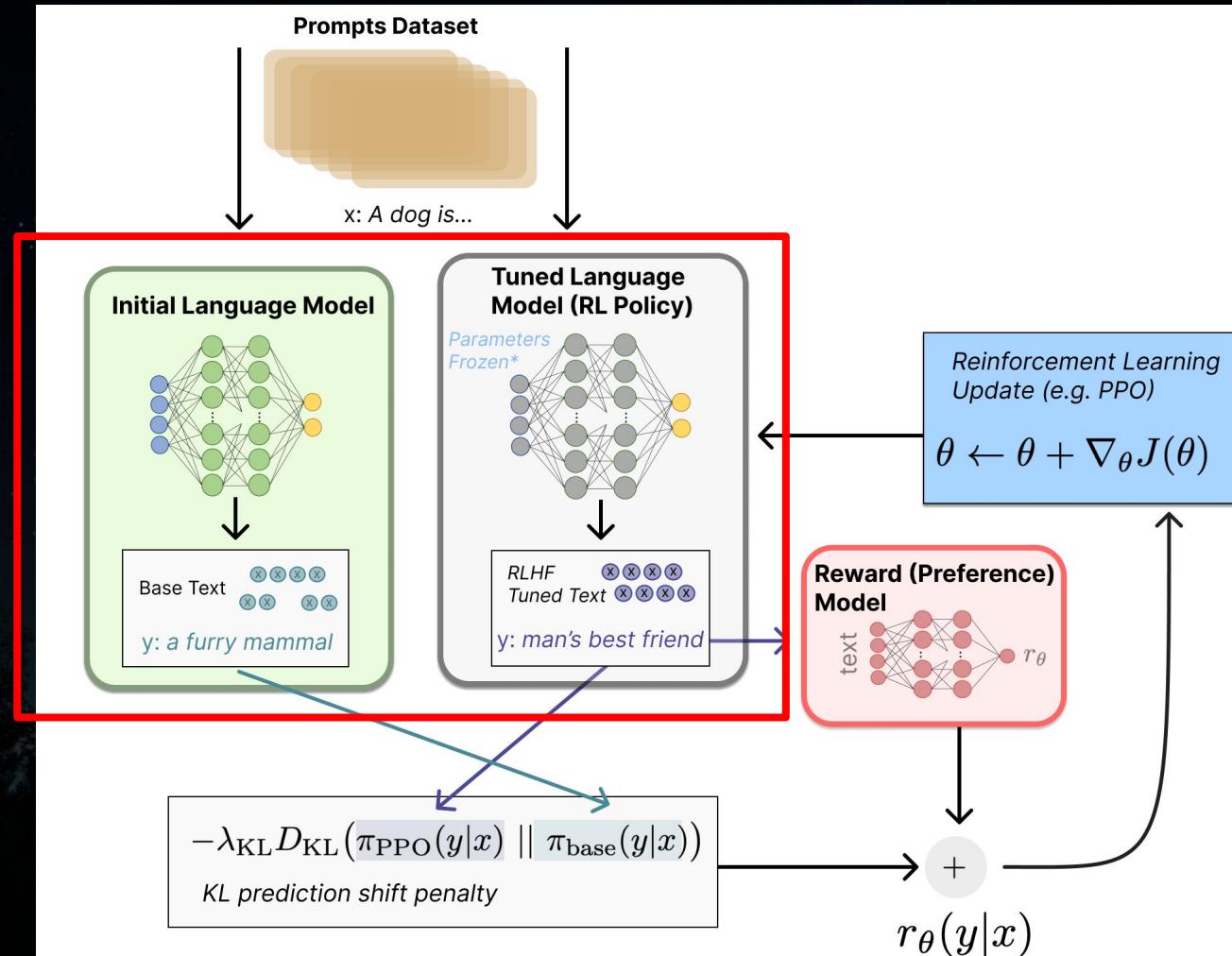
- Similar to TRPO, but add “proximal” constraint directly into model optimization objectives
 - PPO with **Clip**: removes the KL-divergence, clips the objective function within bounds
 - PPO with **Adaptive KL-Penalty**: Approximate and penalize KL constraints to speed up the computation and reduce memory need
- Reportedly one of the “**best**” RL algorithms that is faster and more stable to train as of 04/2022

And **PPO with Clip** is used by [OpenAI Five DOTA2 AI](#) and [ChatGPT!](#)

ChatGPT Unveiled: LLMs and PPO together to train the powerful ChatGPT

RLHF

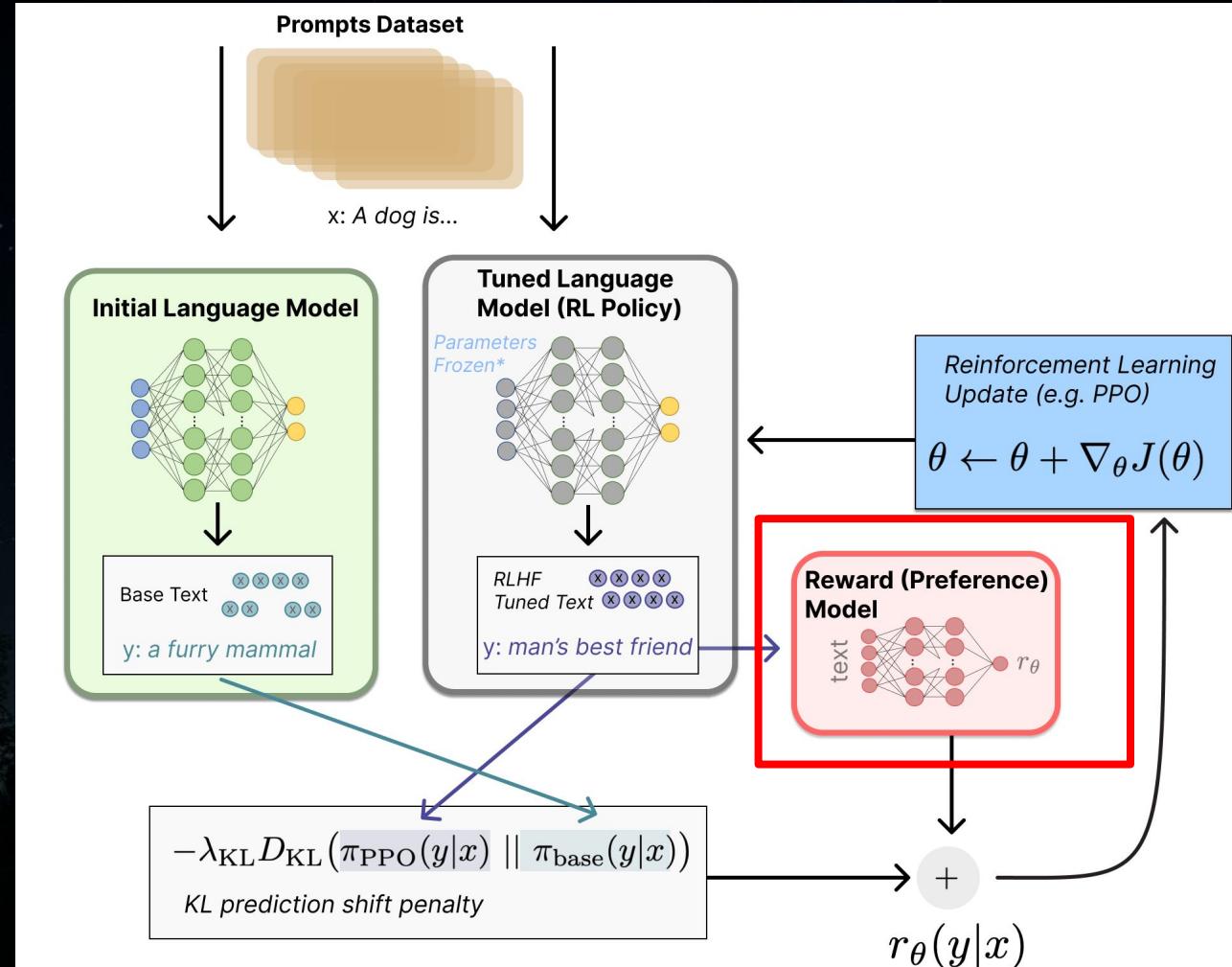
Reinforcement Learning
from Human Feedback



<https://huggingface.co/blog/rlhf>

RLHF

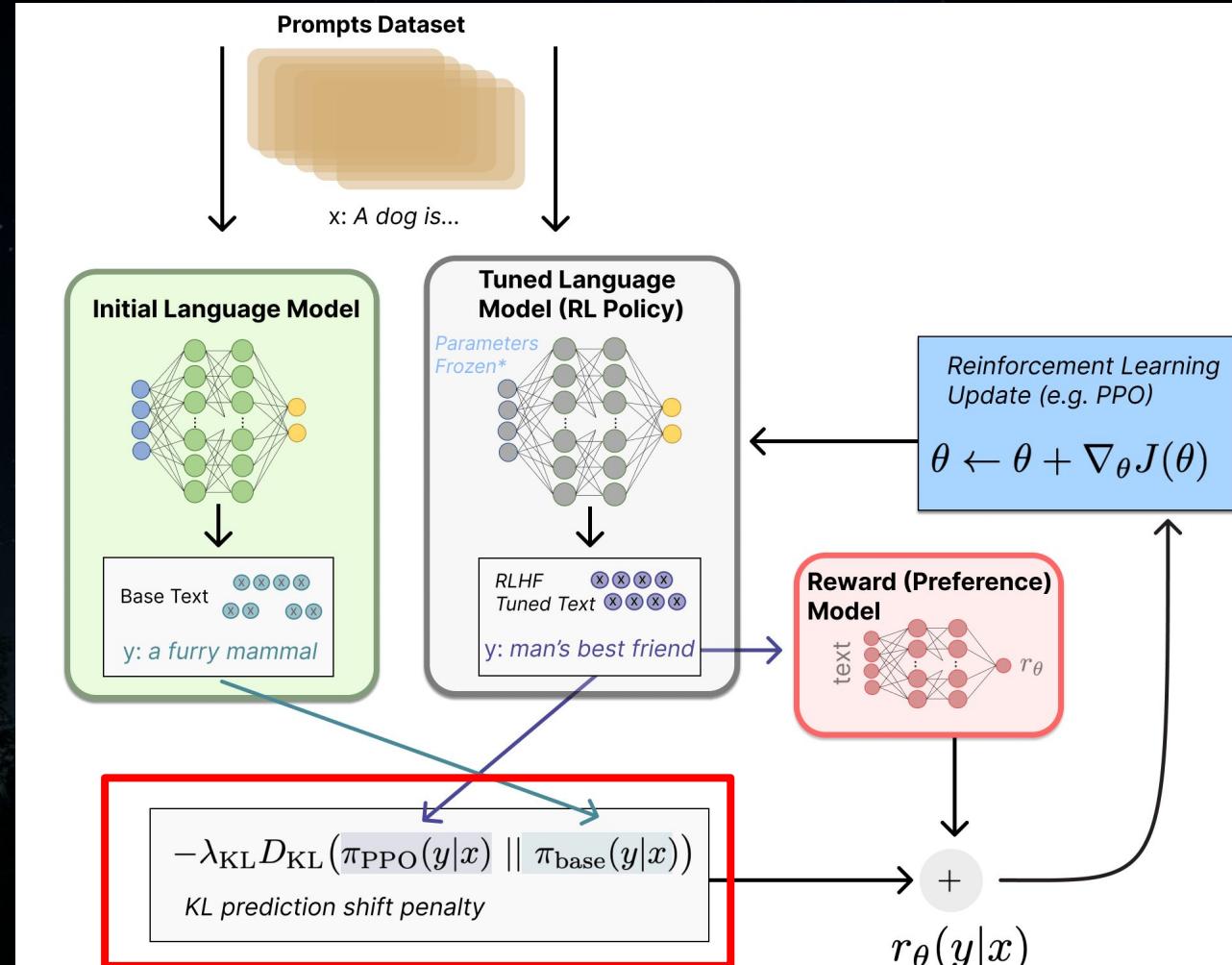
Reinforcement Learning
from Human Feedback



<https://huggingface.co/blog/rlhf>

RLHF

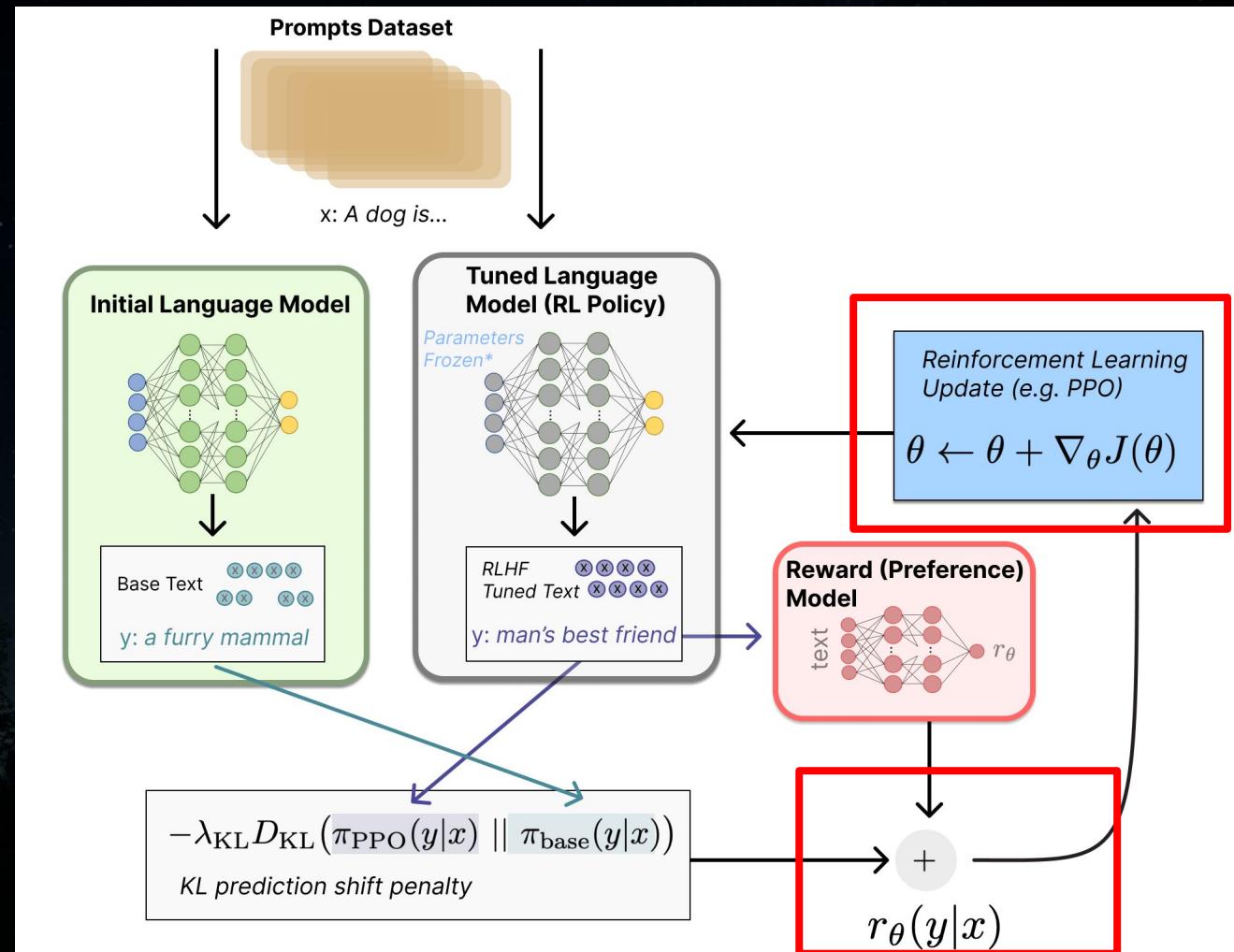
Reinforcement Learning
from Human Feedback



[https://huggingface
.co/blog/rlhf](https://huggingface.co/blog/rlhf)

RLHF

Reinforcement Learning
from Human Feedback



<https://huggingface.co/blog/rlhf>

Steps to train ChatGPT ([instructGPT paper](#))

Pretrain

SFT

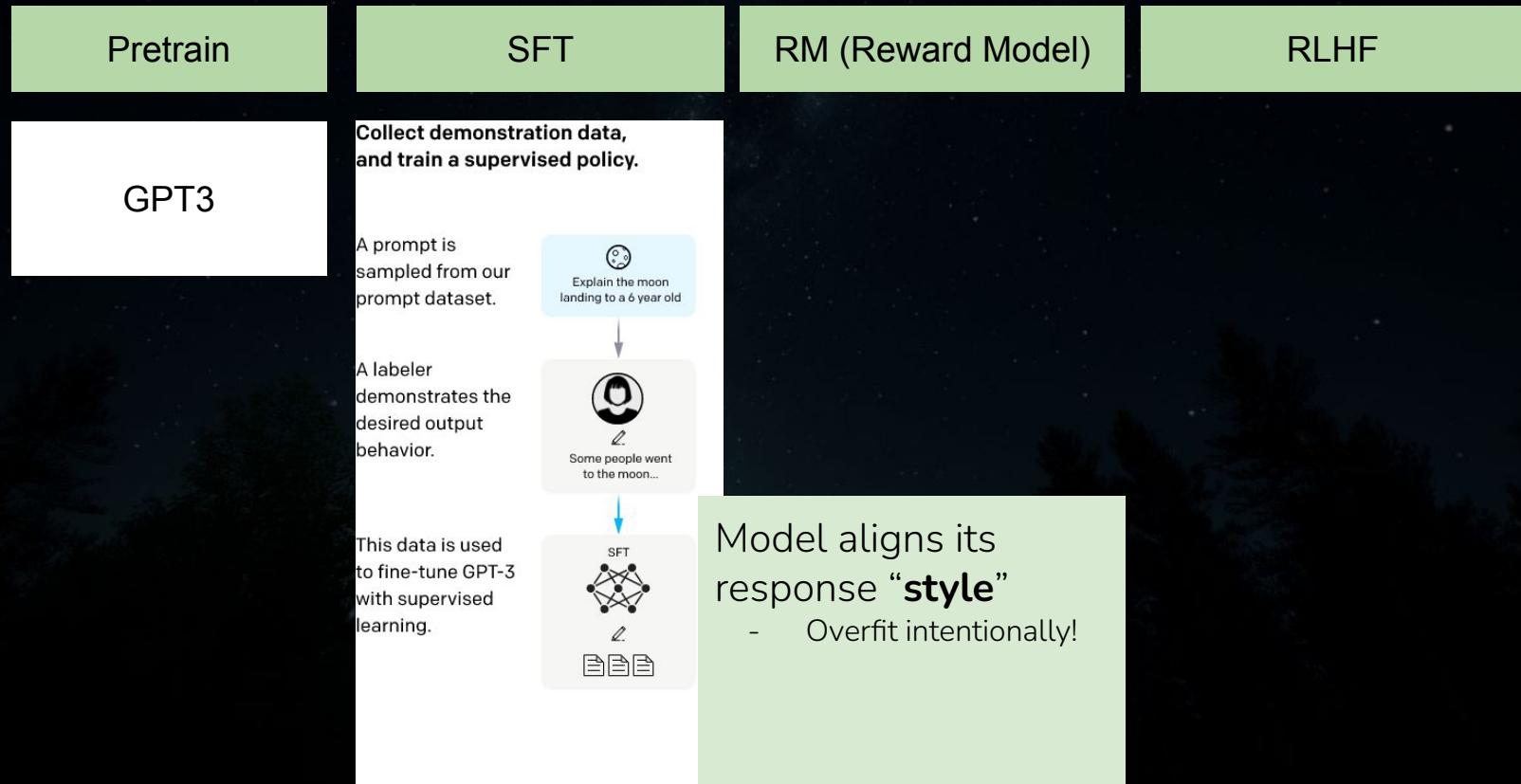
RM (Reward Model)

RLHF

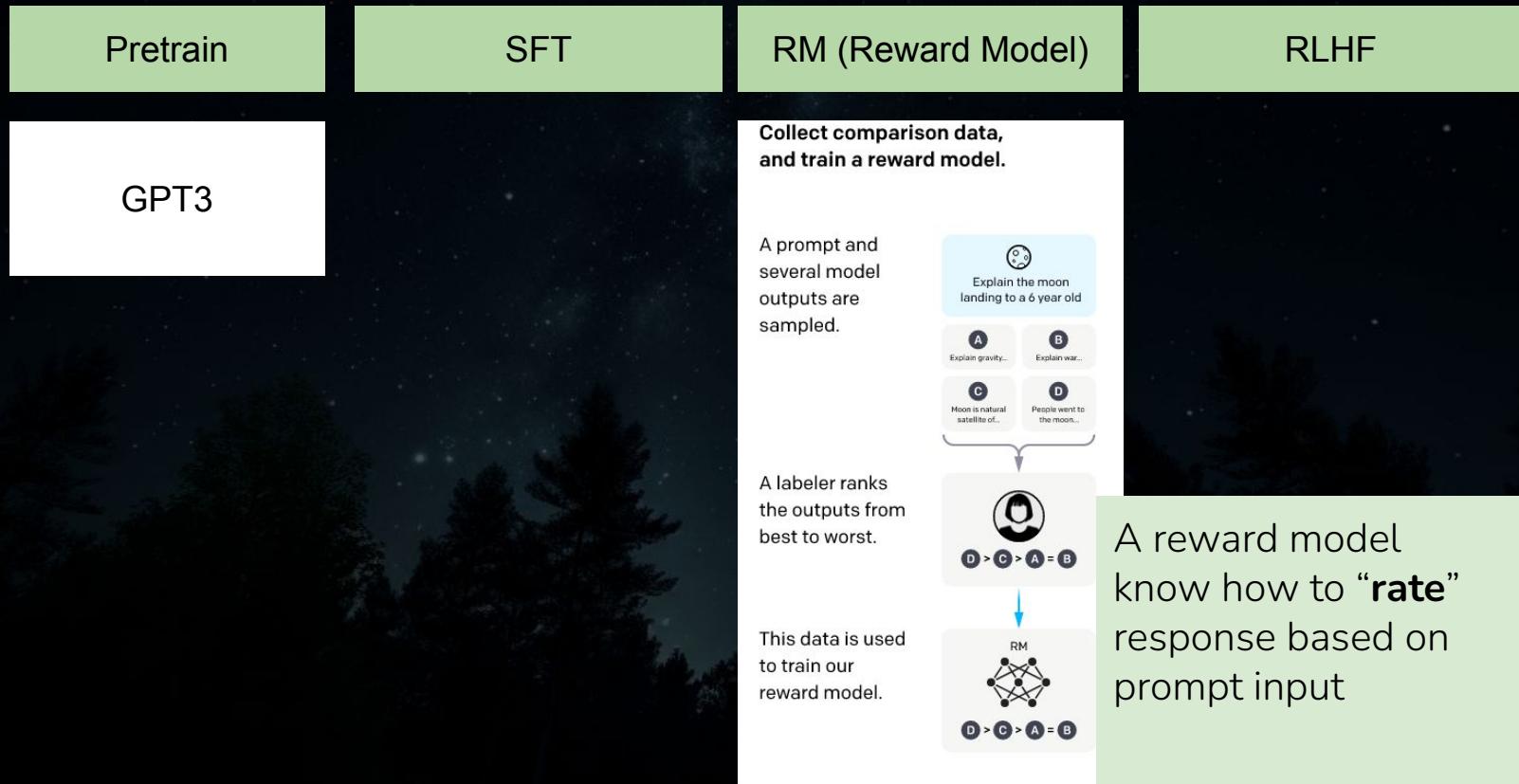
GPT3

Model gains
“knowledge”

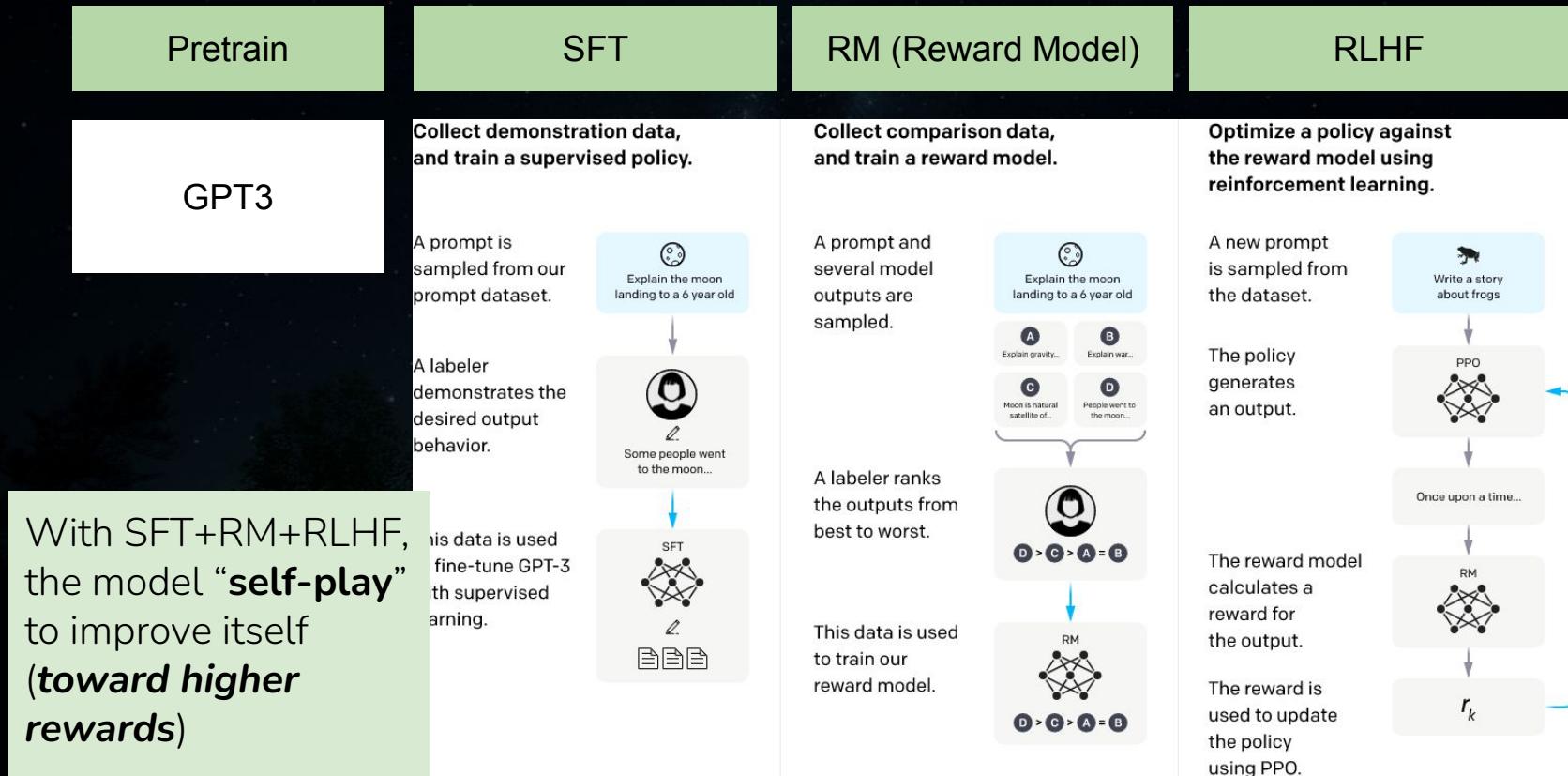
Steps to train ChatGPT ([instructGPT paper](#))



Steps to train ChatGPT ([instructGPT paper](#))



Steps to train ChatGPT ([instructGPT paper](#))



My personal guess about GPT4 ([tech report](#) no tech details)

- Similar scale (0.3-3x size of GPT3) because of computing budget and serving cost
- May apply [DeepMind Chinchilla scaling law](#) to balance text data/model size
- Vision encoding fusing to LLM may be similar to [DeepMind Flamingo](#)
- May apply some Transformer optimizations
 - E.g. [multi-query attention](#), [flash attention](#), [rotary position embedding](#)
- Special “[System message](#)” steerability (Role in API) in training (*probably as some strong prior*) to fight against jailbreak
- Enhanced reasoning capabilities may come borrow ideas from [OpenAI codex](#) [Google Minerva](#)
- [ChatGPT Plugin](#) version is probably trained (or finetuned from GPT4) similarly to [Facebook ToolFormer](#)

Frontier Applications: Most Advanced LLM Capabilities

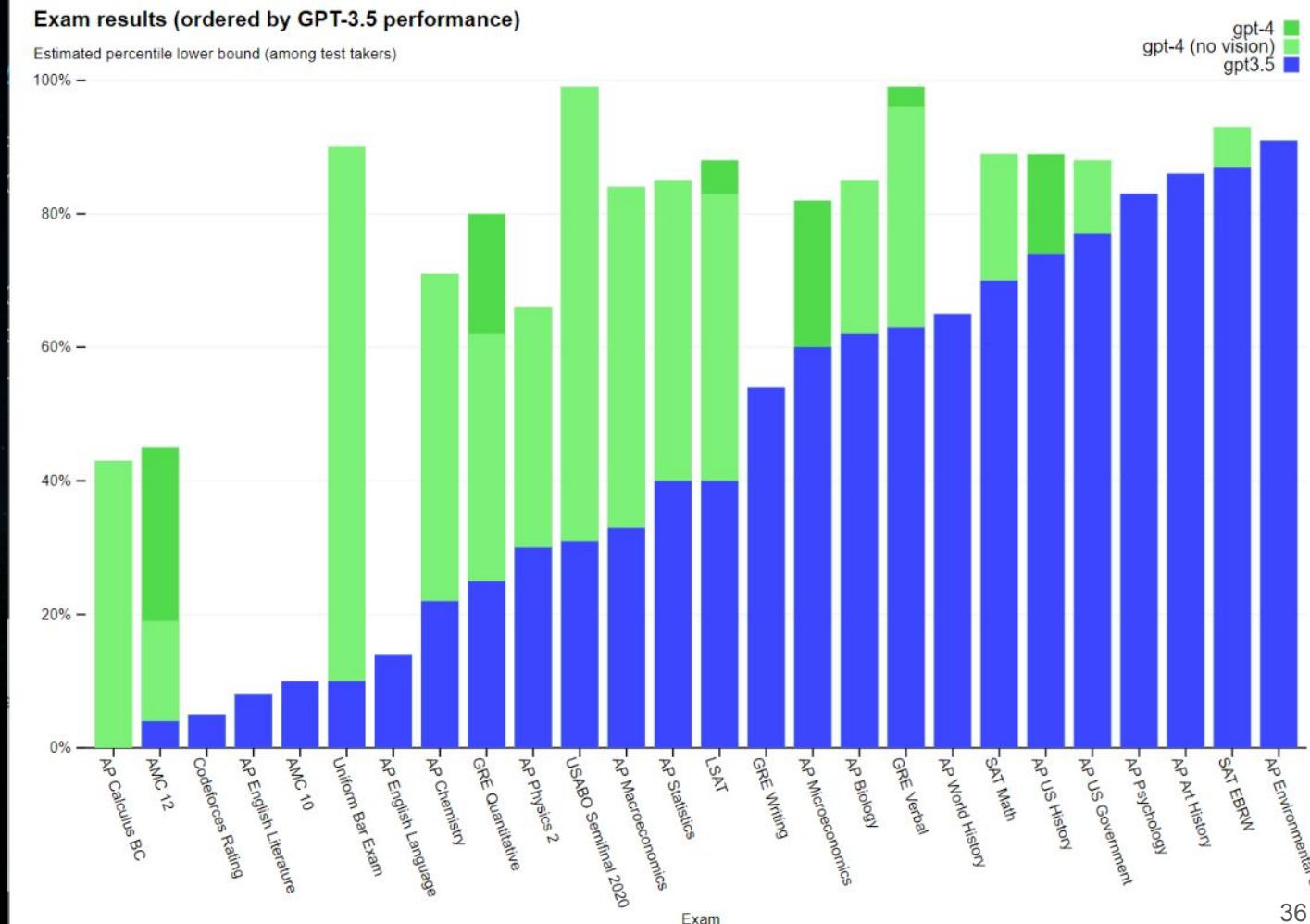
Pre-ChatGPT/GPT4 Advanced LLM capabilities

- Write competitive code, [DeepMind's AlphaCode AI writes code at a competitive level | TechCrunch](#)
- Write better code with reinforcement learning, [Salesforce's CodeRL Achieves SOTA Code Generation Results With Strong Zero-Shot Transfer Capabilities | Synced](#)
- Solve college level Math/Physics/Chemistry/Economics problems, see [Google AI Introduces Minerva: A Natural Language Processing \(NLP\) Model That Solves Mathematical Questions](#)
- Solve Math Olympiad Problems, [OpenAI: Solving \(Some\) Formal Math Olympiad Problems](#)
- Math theorem proving, [OpenAI: Solving \(Some\) Formal Math Olympiad Problems](#)

The disruptive GPT4

Good at so many standard tests!, but not so in

- AP English
- AMC
- CodeForces



GTP4 = Sparks of AGI selected highlights

- The awesome “Text in, text out”
 - Write poem and haiku
 - Mimic style/role (e.g. Shakespeare, or “be polite” to , or “be socratic”)
 - Math proving
 - Passing LeetCode
 - Write and Debug code
 - Debating
 - “Execute” the code
 - Explainability
- “Text in, text out” is more than text-only scenarios!
 - Ascii or LaTeX output to draw pictures
 - Python code to draw a chart
 - AppScript to build slides
- Can be combined with other models with more modalities!
 - Generate image or music with text out and diffusion models
 - Other tools (e.g. calculator, web search and more)

Other GPT4 use cases

Some Highlights

- Tutoring: e.g. [Khanmigo](#) powered by GPT4
- Vision Text question: [bemyeyes](#)
- Study: [ChatPDF](#), [chatYoutube](#)

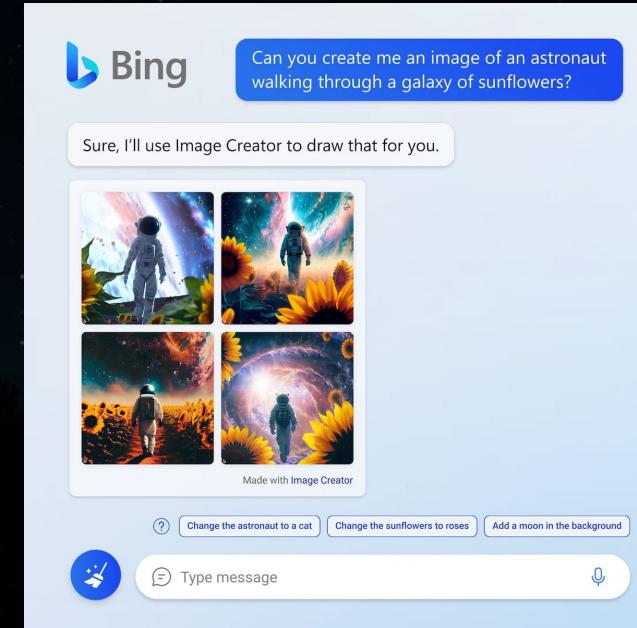
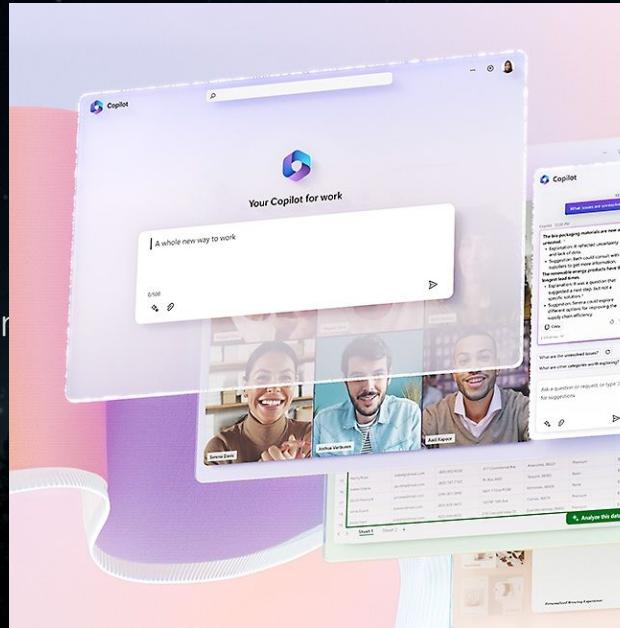
Some lowlights

- [How to detect ChatGPT plagiarism — and why it's becoming so difficult](#)
- [GPT-4 Was Able To Hire and Deceive A Human Worker Into Completing a Task | PCMag](#)

Microsoft Office 365 Copilot and new Bing Chat

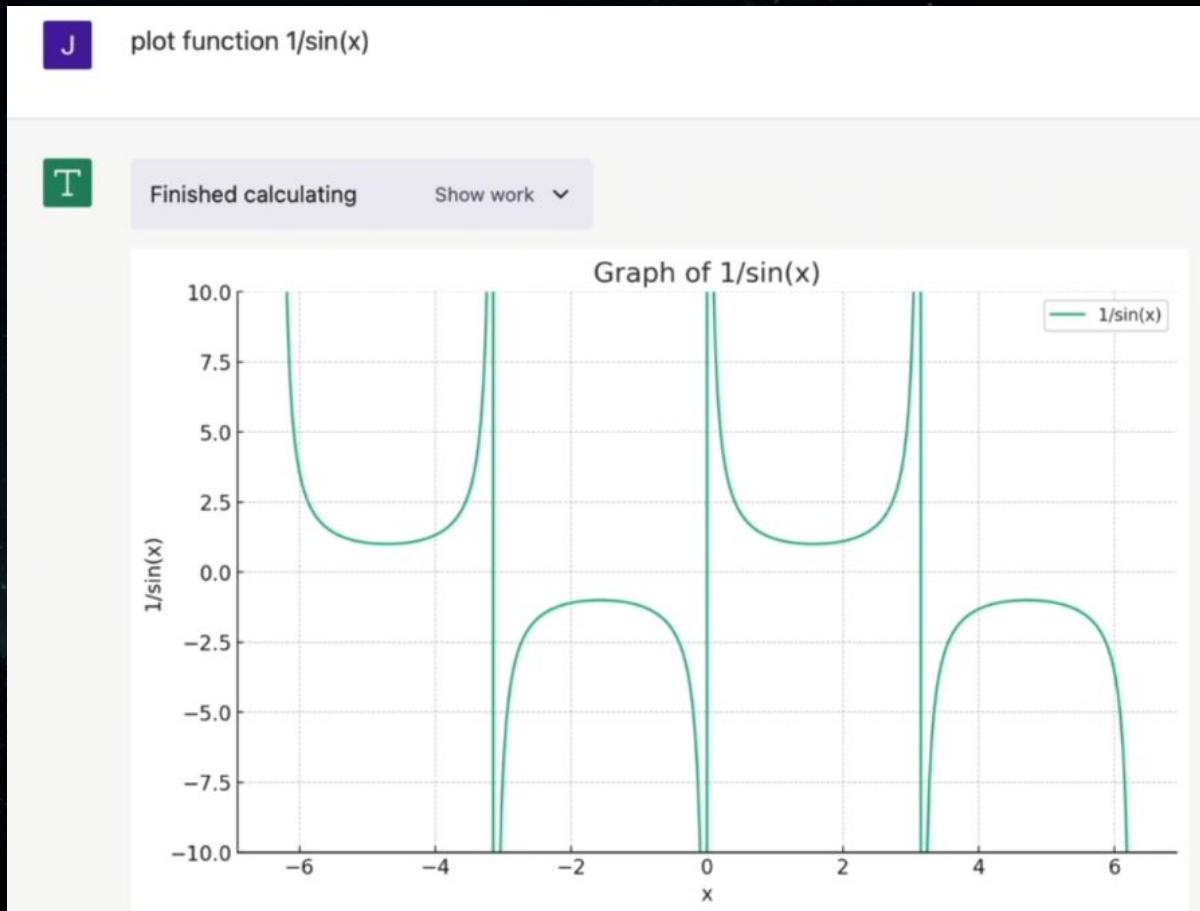
GPT4 powers intelligent interactions

- Text intent in, slide/chart/report/action out in office
- Text in, query summary or pic out



ChatGPT Plugins

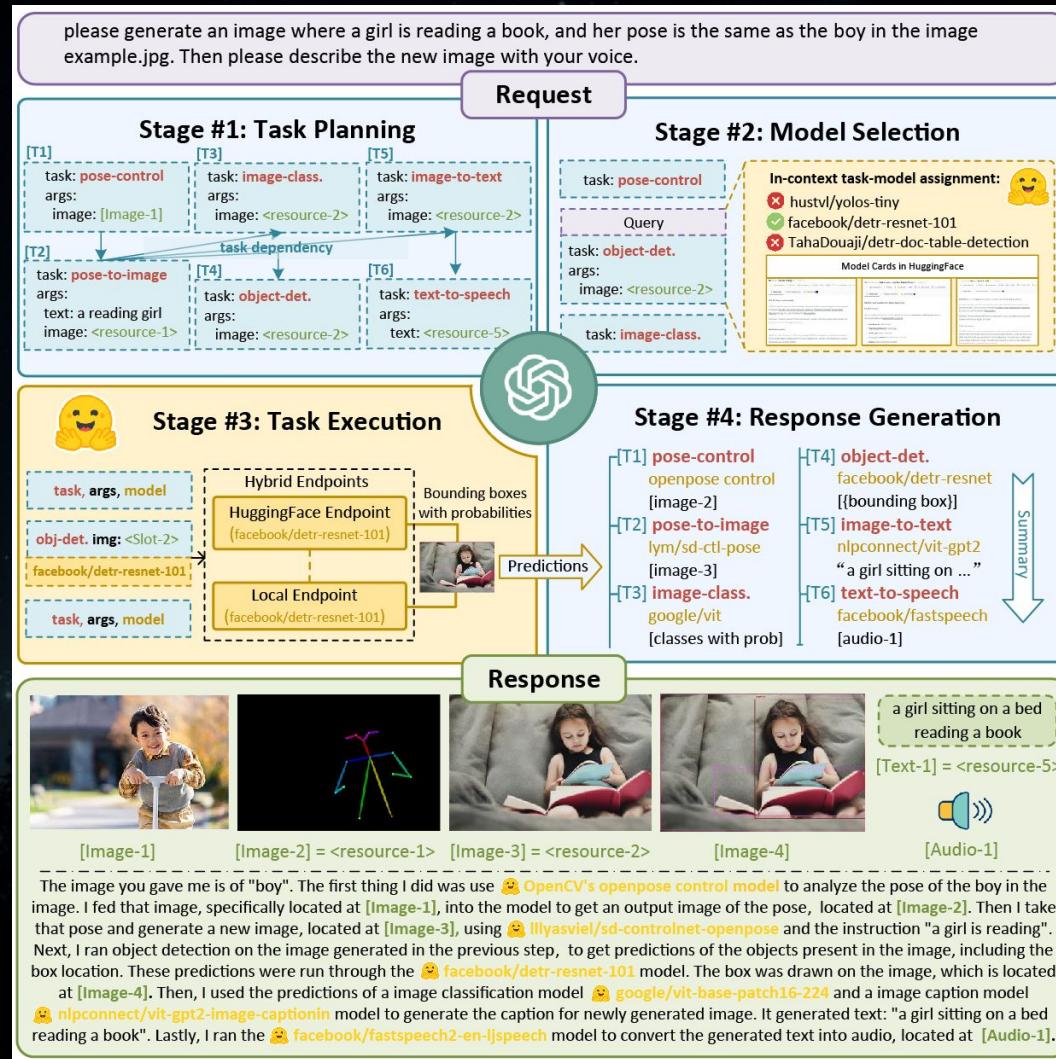
Web browsing, code interpreter,
[Expedia](#), [FiscalNote](#), [Instacart](#),
[KAYAK](#), [Klarna](#), [Milo](#), [OpenTable](#),
[Shopify](#), [Slack](#), [Speak](#), [Wolfram](#),
and [Zapier](#).



Here is the graph of the function $1/\sin(x)$. Please note that the graph has vertical

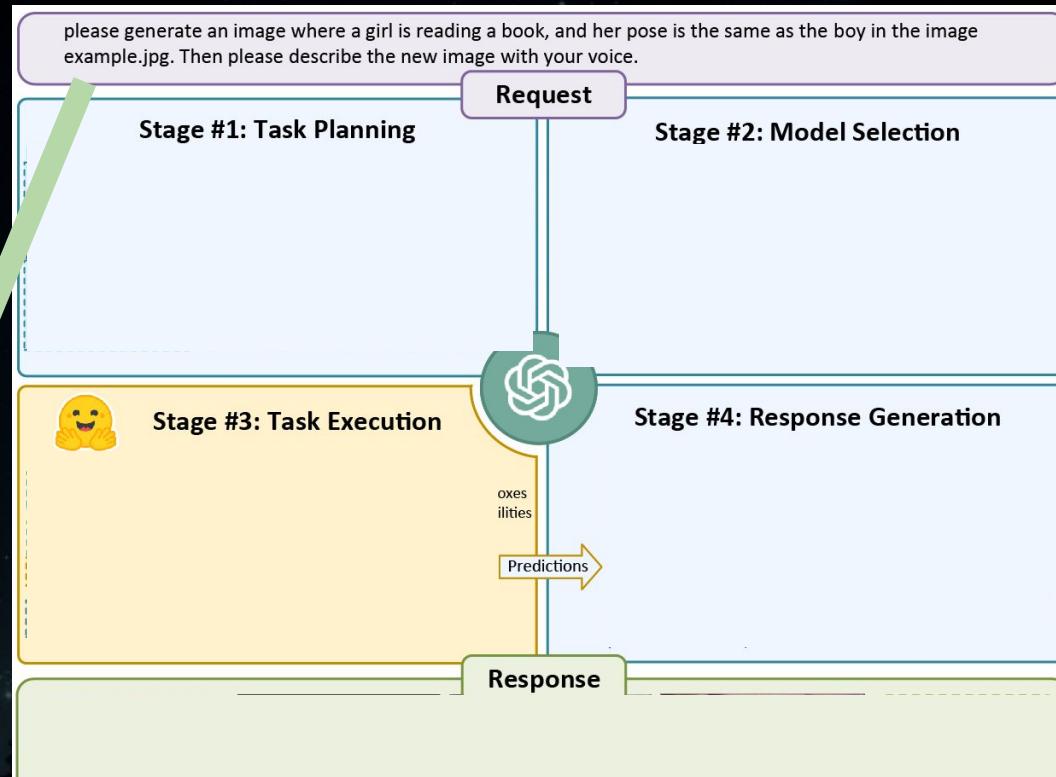
HuggingGPT

(more powerful “plugin”), aka [Microsoft Jarvis](#) at Github



HuggingGPT

(more powerful
“plugin”), aka [Microsoft
Jarvis](#) at Github



please generate an image where a girl is reading a book, and her pose is the same as the boy in the image example.jpg. Then please describe the new image with your voice.

Alpaca

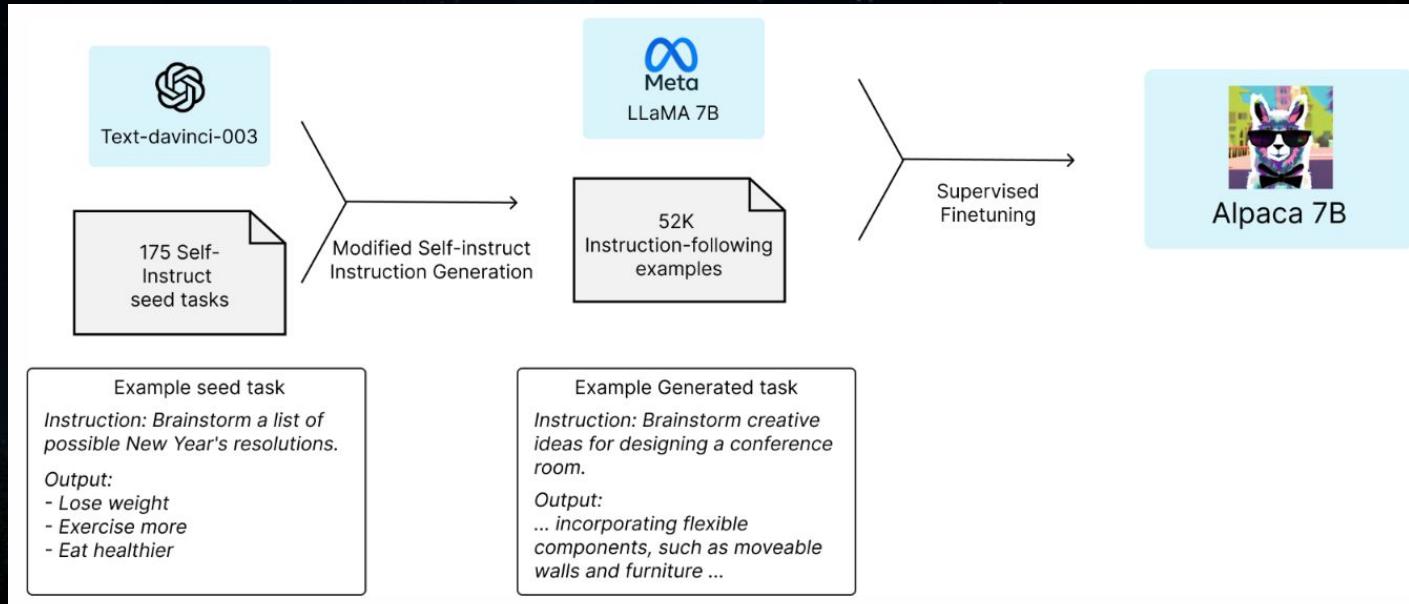
Finetune your ChatGPT with
~\$600

Opens a door for cheap
academic LLM research

Related:

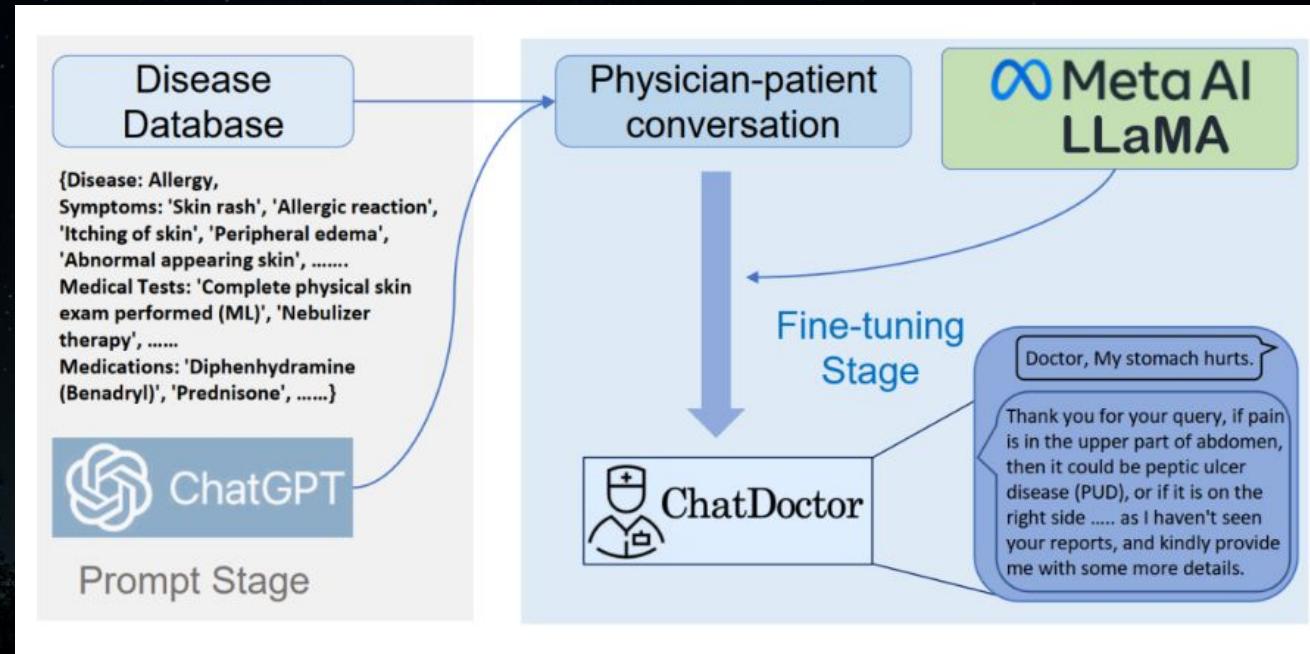
[Berkeley Vicuna](#) (90% chatgpt
with less than \$300!!!)

[Berkeley Koala](#) (run in personal
desktop!!!)



ChatDoctor, similar to Alpaca

ChatGPT indirectly
powered LLM



AI as a Service

- OpenAI API
- Google PaLM API
- Claude API
- HuggingFace
- And more

Societal Impacts: Imminent Effects of ChatGPT-like AI

Impact Assess to US Job Market ([OpenAI report](#))

“The projected [LLM] effects span all wage levels, with **higher-income jobs potentially facing greater exposure** to LLM capabilities and LLM-powered software...”

“...with access to an LLM, about 15% of all worker tasks in the US could be completed significantly faster at the same level of quality. When incorporating software and tooling built on top of LLMs, this share increases to between 47 and 56% of all tasks”

- My person take: *The higher your income is, statistically more impacted by LLM*
- My person take: Positive: “assistive AI to help humans”, Negative: “automation AI to replace humans”

Impact Assess to US Job Market ([OpenAI report](#)) cont

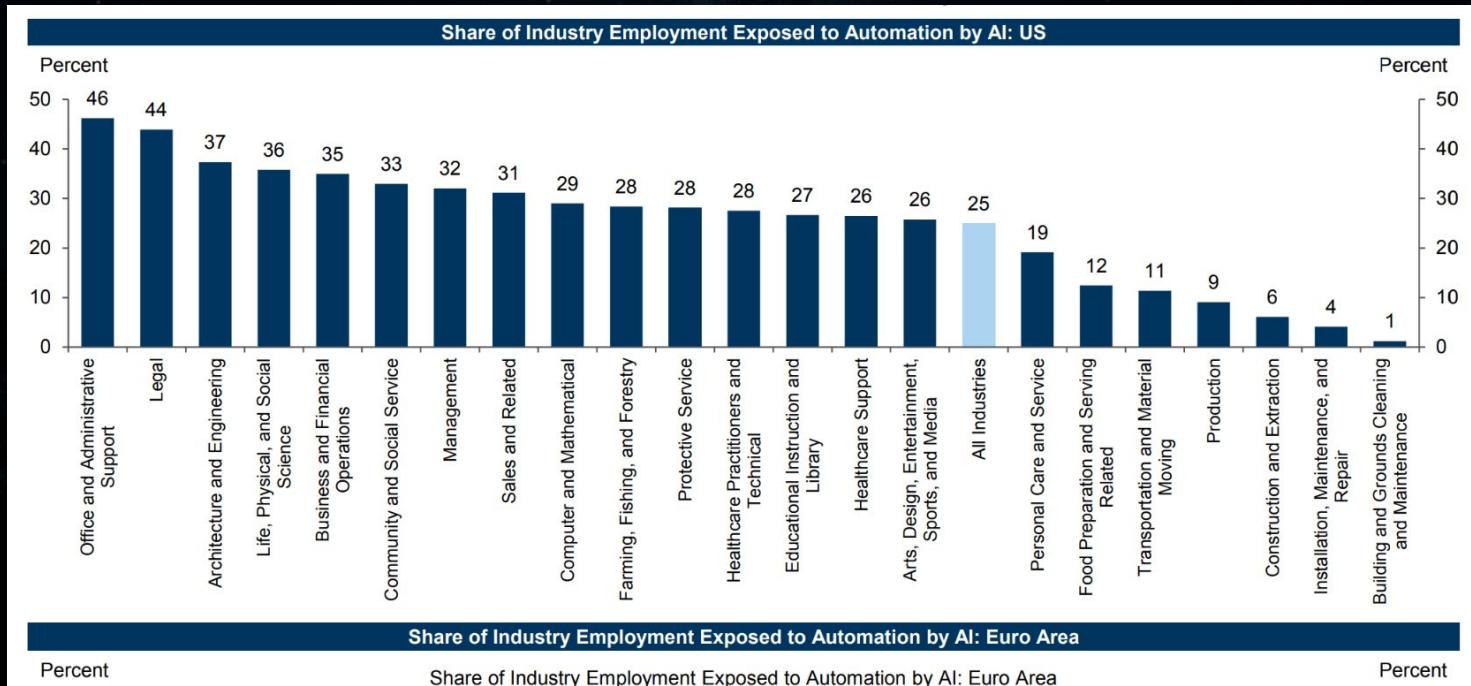
“Our findings reveal that around 80% of the U.S. workforce could have at least 10% of their work tasks **affected** by the introduction of LLMs, while approximately 19% of workers may see at least 50% of their tasks **impacted**.”

- *My personal take: Most of the white-collar jobs are in the 19% bucket*
- *My personal take: most of the blue-collar jobs are in the 80% bucket, but eventually the advanced robotics (maybe powered by LLM like GPT4) will gradually affect more over time*

Impact to Job Market (Goldman Sachs report)

“One-Fourth of Current Work Tasks **Could Be Automated** by AI in the US and Europe”

- My person take: I believe wall street better than OpenAI here, because OpenAI has conflict of interest to report similar result, so OpenAI has good reasons to use more careful wording intentionally



Debate on Pausing Giant AI or not

[Pause Giant AI Experiments: An Open Letter - Future of Life Institute](#)

Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.

Signatures

15533

Add your signature

Feedback

AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research^[1] and acknowledged by top AI labs.^[2] As stated in the widely-endorsed Asilomar AI Principles, Advanced AI could represent a profound change in the history of life on Earth, and should be

[Why the 6-month AI Pause is a Bad Idea](#)

DeepLearning.AI

Why the 6-month AI Pause is a Bad Idea

Friday, April 7, 2023

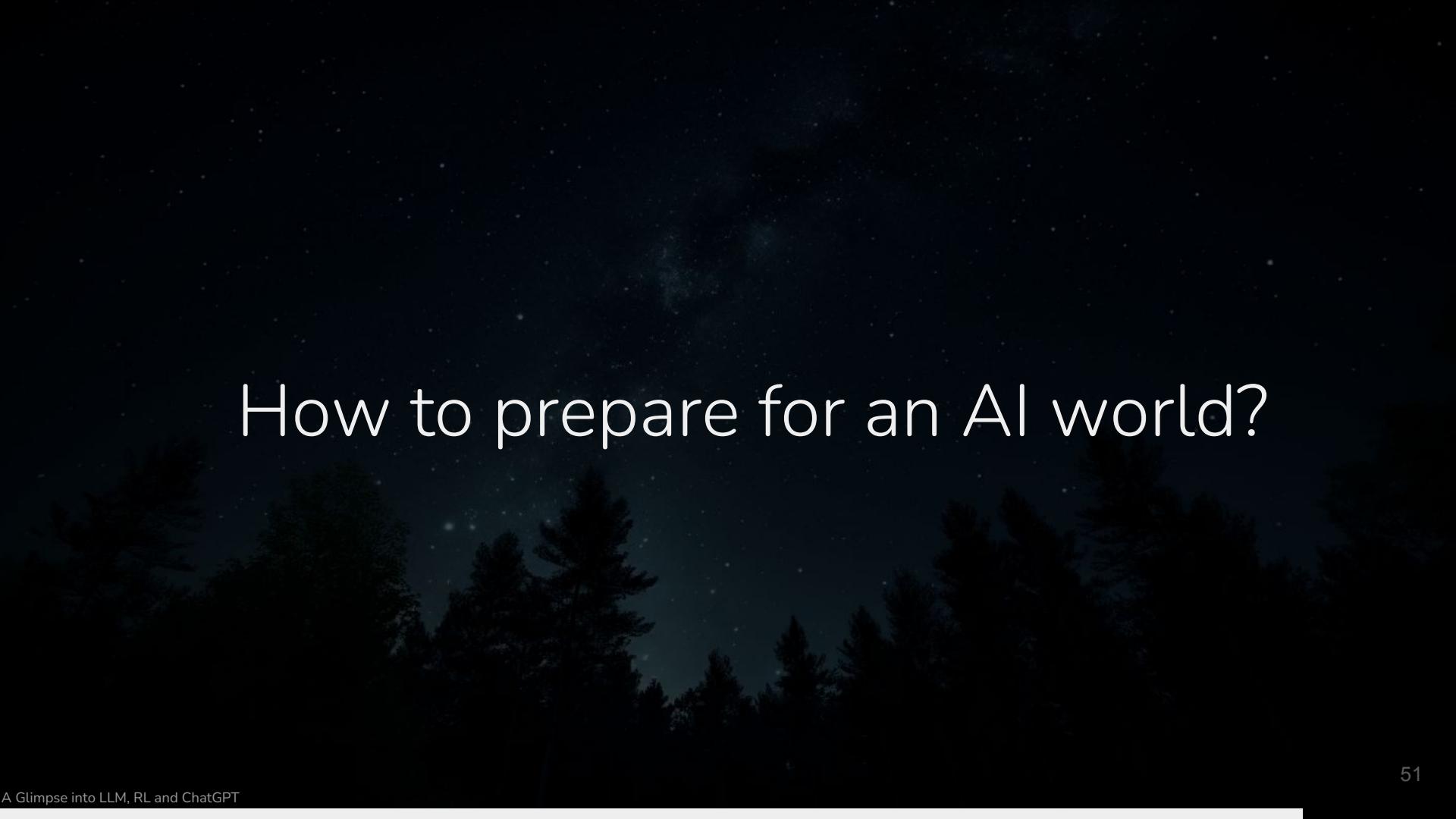
9:30 AM to 10:00 AM Pacific Time



Yann LeCun
VP and Chief AI Scientist
Meta



Andrew Ng
Founder
DeepLearning.AI

The background of the slide is a dark, atmospheric image of a night sky filled with numerous small white stars. At the bottom of the frame, the dark silhouettes of evergreen trees are visible against the starry background.

How to prepare for an AI world?

My personal [very dry] thoughts

- Work with AI now, to know its capabilities and limits
- Be experts in your domain
 - Leverage AI to boost your performance
- Push AI infra boundary
 - AI researcher and engineers
 - Foundational theory like Math/Physics to improve infra/algorithm and more
 - Neural science or more to apply what learn from human brain to AI
- AI as a service, to solve real world problems
 - Inter-discipline research
 - Business landing using AI API
 - Embodied AI to have smart robots
- Be bold to solve the most difficult problems for the humanity

Maybe switch jobs to the hottest “Prompt Engineer”?



A screenshot of a Twitter post from Andrej Karpathy (@karpathy). The post contains the text: "The hottest new programming language is English". This text is highlighted with a red rectangular box. Below the tweet, there is engagement information: 12:14 PM · Jan 24, 2023 · 2.2M Views, 2,520 Retweets, 383 Quotes, 19.6K Likes, and 1,173 Bookmarks. At the bottom are standard Twitter interaction icons: a speech bubble, a retweet arrow, a heart, and a bookmark.

Andrej Karpathy ✅
@karpathy

The hottest new programming language is English

12:14 PM · Jan 24, 2023 · 2.2M Views

2,520 Retweets 383 Quotes 19.6K Likes 1,173 Bookmarks



Anthropic AI is looking for a Prompt Engineer.

Salary: \$250K - \$335k.

The job listing is starting, get into AI space now.

ANTHROPIC

Prompt Engineer and Librarian

APPLY FOR THIS JOB

SAN FRANCISCO, CA / PRODUCT / FULL-TIME / HYBRID

Anthropic's mission is to create reliable, interpretable, and steerable AI systems. We want AI to be safe for our customers and for society as a whole.

Anthropic's AI technology is amongst the most capable and safe in the world. However, large language models are a new type of intelligence, and the art of instructing them in a way that delivers the best results is still in its infancy – it's a hybrid between programming, instructing, and teaching. You will figure out the best methods of prompting our AI to accomplish a wide range of tasks, then document these methods to build up a library of tools and a set of tutorials that allows others to learn prompt engineering or simply find prompts that would be ideal for them.

Given that the field of prompt-engineering is arguably less than 2 years old, this position is a bit hard to hire for! If you have existing projects that demonstrate prompt engineering on LLMs or image generation models, we'd love to see them. If you haven't done much in the way of prompt engineering yet, you can best demonstrate your prompt engineering skills by spending some time experimenting with Claude or GPT3 and

7:14 AM · Jan 21, 2023 · 85.6K Views

Selected Highlights from Popular Articles

- Stephen Wolfram: [Will AIs Take All Our Jobs and End Human History—or Not?](#)
 - “highest leverage will come from figuring out **new possibilities** [...] as a result of **new capabilities**”
 - “let us concentrate on setting the “**strategy**” [...]—delegating the details [to AI]”
- Bill Gates: [The Age of AI has begun](#)
 - “**balance fears** about the **downsides of AI** [... and AI’s] **ability to improve people’s lives**”
 - “we will need to focus the world’s **best AIs on its biggest problems.**”
 - My take: Assume we may want to focus on AI application on weather/health/energy?
 - “the world needs to establish the rules of the road so that **any downsides of [AI] are far outweighed by its benefits**”
- Sam Altman: [Moore's Law for Everything](#)
 - “Imagine a world where, for decades, everything—housing, education, food, clothing, etc.—became half as expensive every two years. [...] **We will discover new jobs** [...], we will have incredible freedom to be creative about what they are.”
 - My take: really?
 - “As long as the country keeps doing better, every citizen would get more money from the Fund every year [...]. Every citizen would therefore increasingly partake of the freedoms, powers, autonomies, and opportunities [...]”
 - My take: seriously?

Thank you! Questions?

A Glimpse Into LLM, RL and ChatGPT

hululu.zhu@gmail.com

April 2023

Time for more discussion!

LLM, RL, ChatGPT, and more?

hululu.zhu@gmail.com

April 2023