

fun-ai-talk

A Glimpse Into LLM, RL and ChatGPT

hululu.zhu@gmail.com

April 2023

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

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


LLM Foundation: Deep Learning and Transformer-based LLMs

AI, Machine Learning, and Deep learning

Artificial Intelligence

Machine Learning

Deep Learning (aka DNN)

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

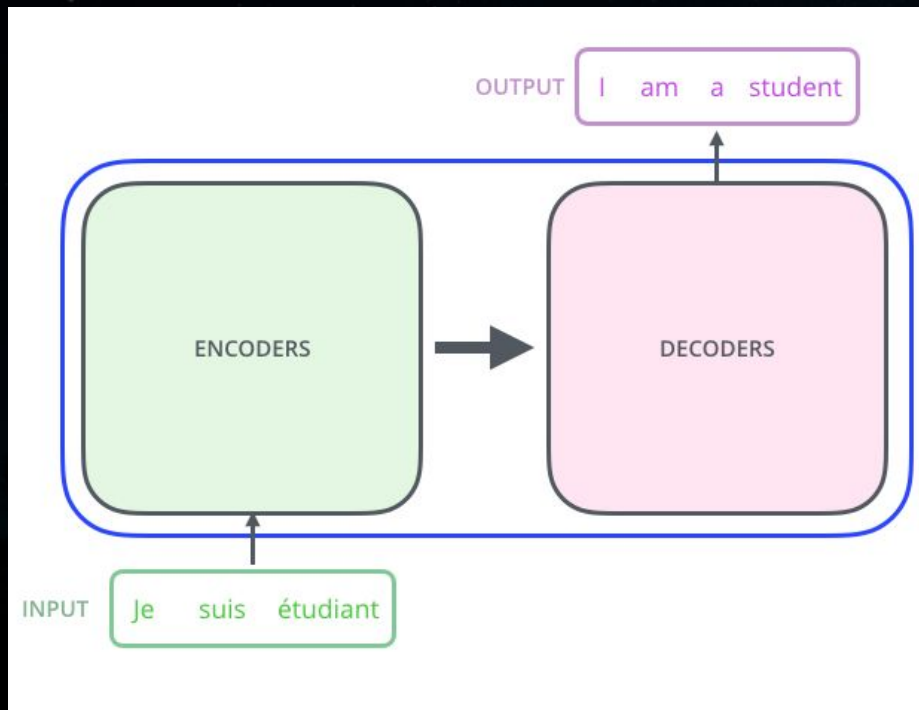
[Artificial Intelligence](#)

[Machine Learning](#)

[Deep Learning](#)

[Google Deep Learning Playground](#)

Modern LLM building blocks: Transformer



Complex architecture (left chart is oversimplified)

Highlights

- One of the implementation of [Seq2Seq](#)
- Originally for translation, but proven successful in NLP and CV
- Introduced “attention” through multi-head attention implementation
- Started the “[?] is all you need” style
- Many many variations since 2018

[Attention Is All You Need \(Transformer paper\) by Google](#) | [The Annotated Transformer by Harvard NLP](#) | [A Survey of Transformers by Fudan](#)

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”

<ul style="list-style-type: none">- X: “Paris is a”- Y: “beautiful”- Model: “good”- Optimize: “good” 👎 “beautiful” 👍	<ul style="list-style-type: none">- 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)

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 concepts

- Pretraining, finetuning, prompt engineering, prompt tuning
- LLM decoding algorithms
- 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 concepts

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

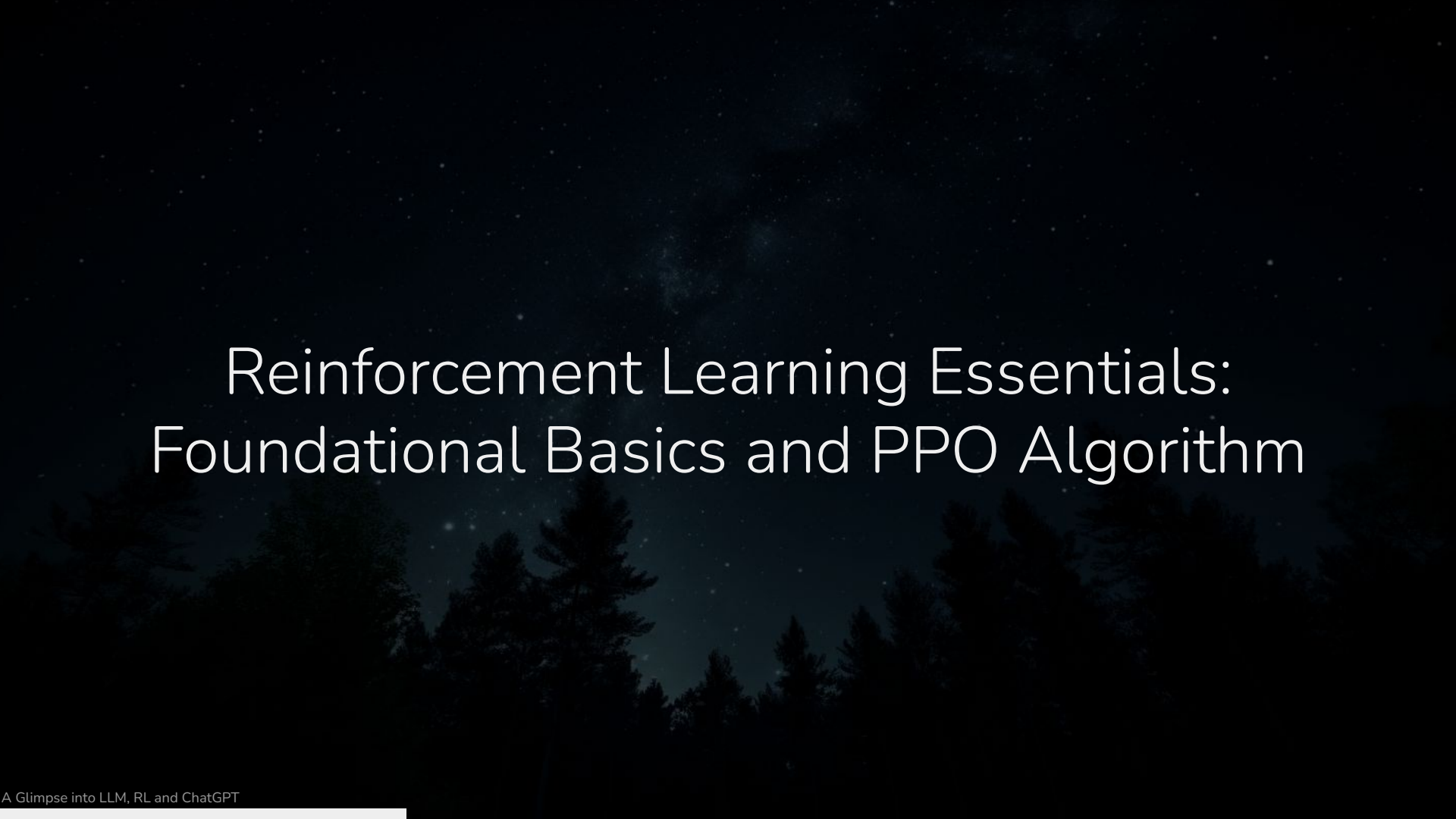
Check out [this deck](#) to include more LLM concepts and examples

Guess what current most powerful LLMs not so good at?

- **My personal take: “LLMs do not know what they do not know”**
 - checkout [GPT4 paper calibration](#)

Success of modern LLMs may come 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

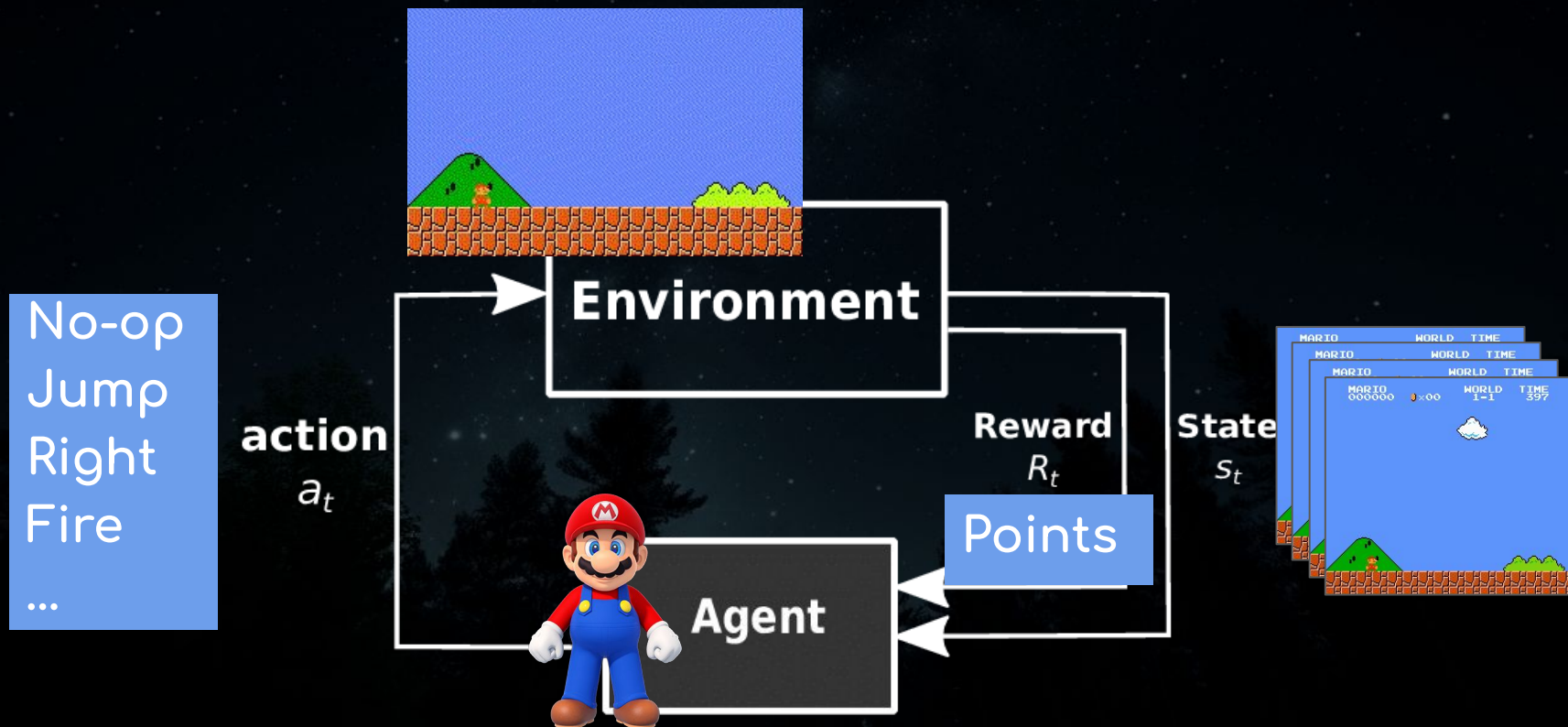
Check out [this deck](#) for a summary of more Alpha* papers by DeepMind



What is Reinforcement learning?



What is Reinforcement learning? Cont (Mario case)



Evolution towards PPO algorithm

- Value-based RL Algorithm such as [Q-learning](#)
 - Predict different rewards based on different actions to take, indirectly decide which action to take
 - But often overestimate the reward and cause bad performance in test
- [Policy Gradient](#)
 - Model directly predict the distribution of actions (e.g. 10% action #1, 50% action 2, 40% action 3)
 - But often unstable and hard to converge during training
- [TRPO](#) (Trust Region Policy Optimization)
 - On top of Policy Gradient, introduced Trust Region to avoid large bad moves during training
 - But computationally expensive because of huge calculation (e.g. KL divergence)
- [PPO](#) (Proximal Policy Optimization)
 - On top of TRPO, used the simple “Clip” idea to clip the distribution in a defined range
 - Still one of the “best” RL algorithms as of 2023 since 2017

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

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

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

Pretrain

SFT
(Supervise Finetune)

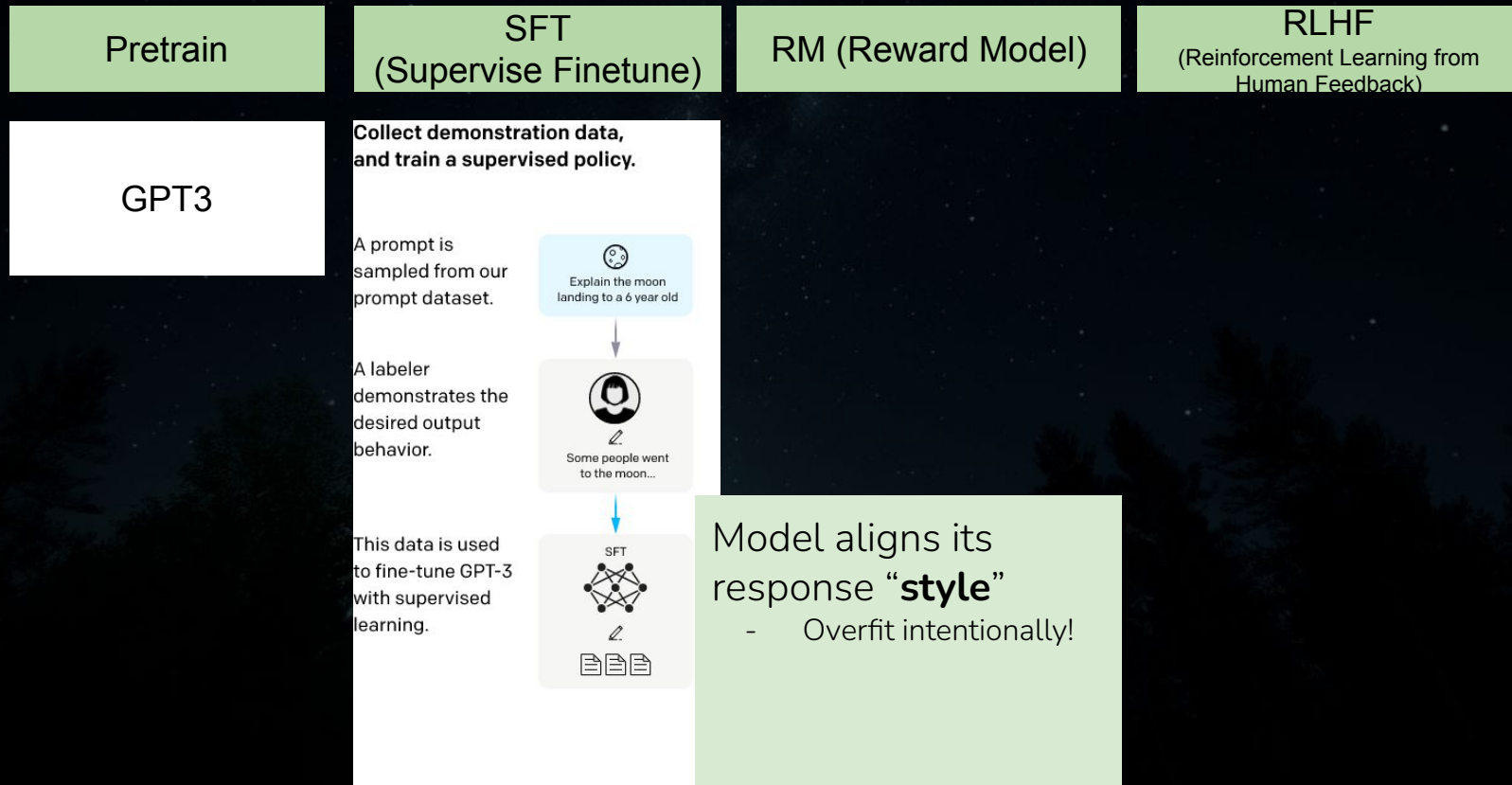
RM (Reward Model)

RLHF
(Reinforcement Learning from
Human Feedback)

GPT3

Model gains
“**knowledge**”

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



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

Pretrain

GPT3

SFT
(Supervise Finetune)

RM (Reward Model)

Collect comparison data,
and train a reward model.

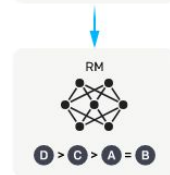
A prompt and
several model
outputs are
sampled.



A labeler ranks
the outputs from
best to worst.



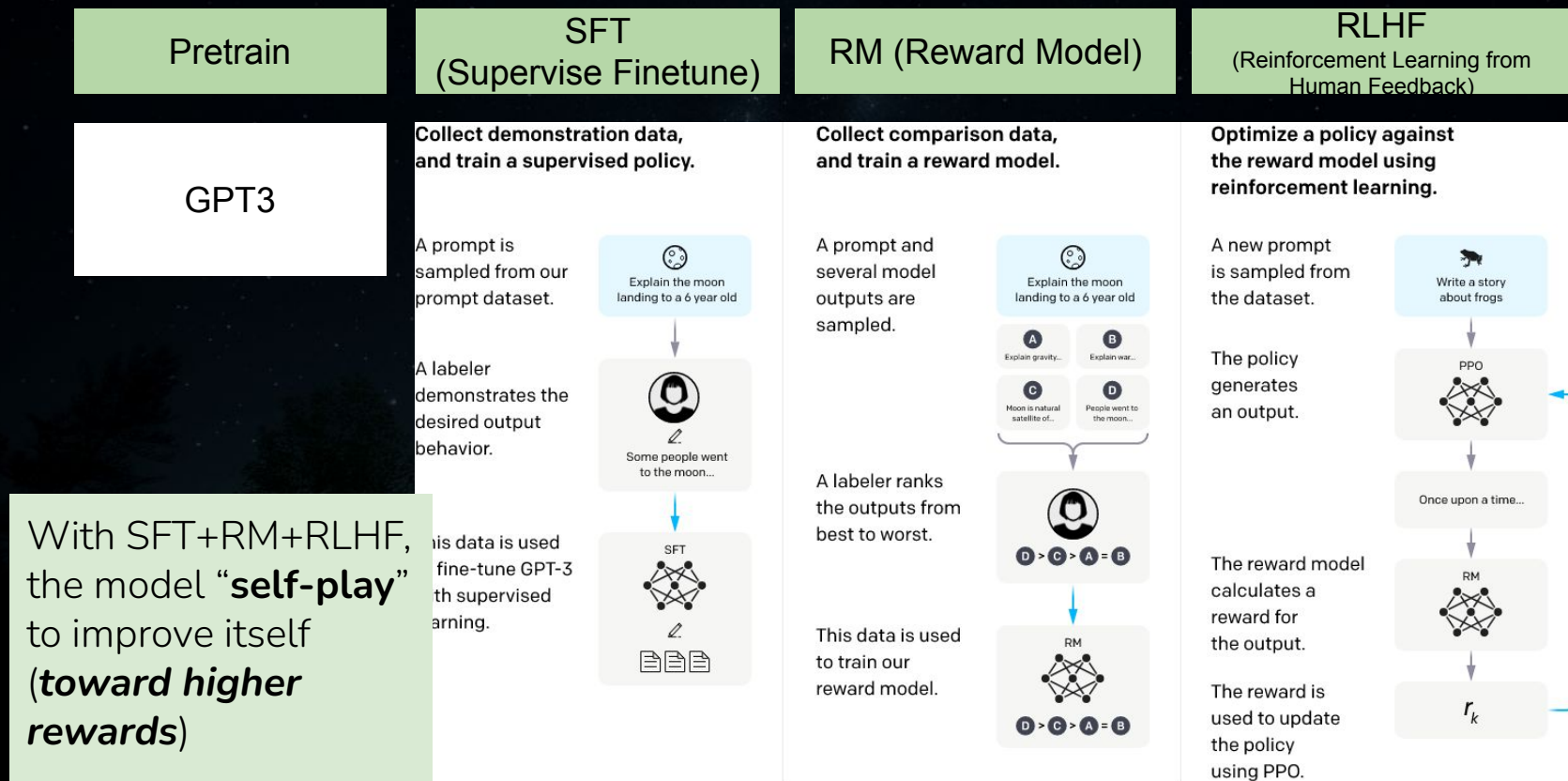
This data is used
to train our
reward model.



RLHF
(Reinforcement Learning from
Human Feedback)

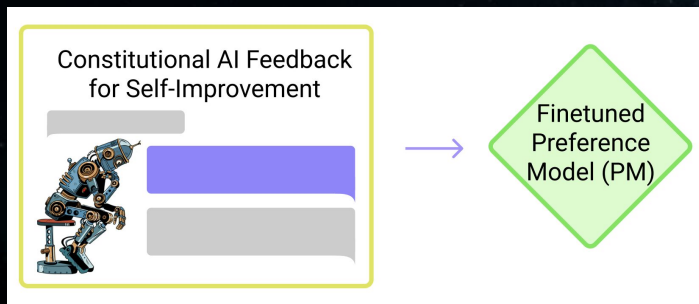
A reward model
know how to “**rate**”
response based on
prompt input

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



Related: RLAI (RL from AI Feedback, aka Constitutional AI)

- RLHF: Human feedback (preference) data is used to train a Preference Model (Reward model)
- RLAI: AI feedback (with prompt) is used to give preference data to train a AI feedback based Preference model



replace Reward Model stage in RLHF, so let the constitutional AI feedback replace human feedback (represented by reward model)

Collect comparison data, and train a reward model.

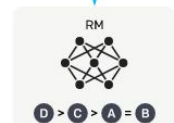
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



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

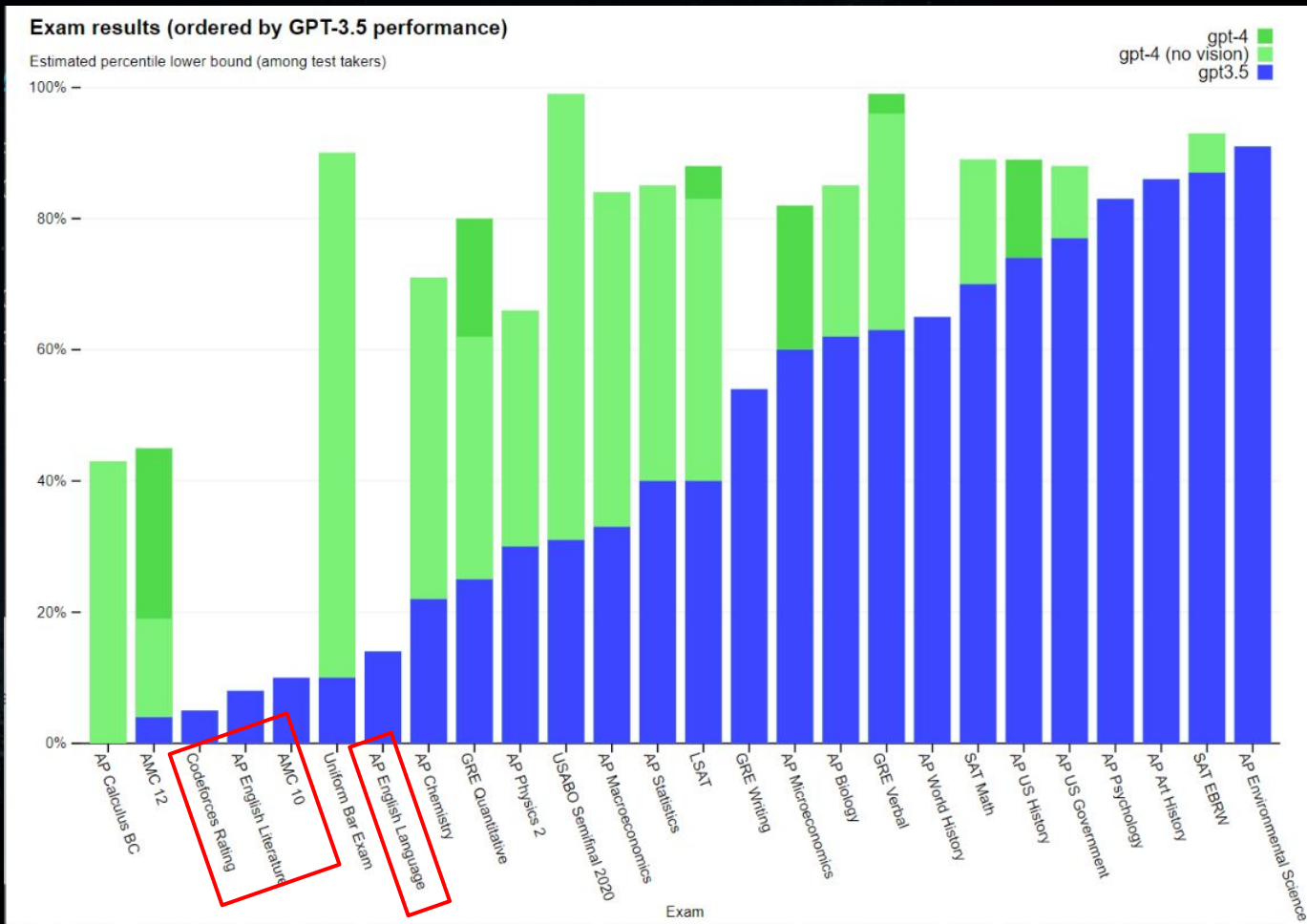
Pre-ChatGPT/GPT4 Advanced LLM capabilities

- 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 language and literature
- AMC 10 (but good at AMC 12)??
- CodeForces (competitive programming)



GPT4 = 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](#)

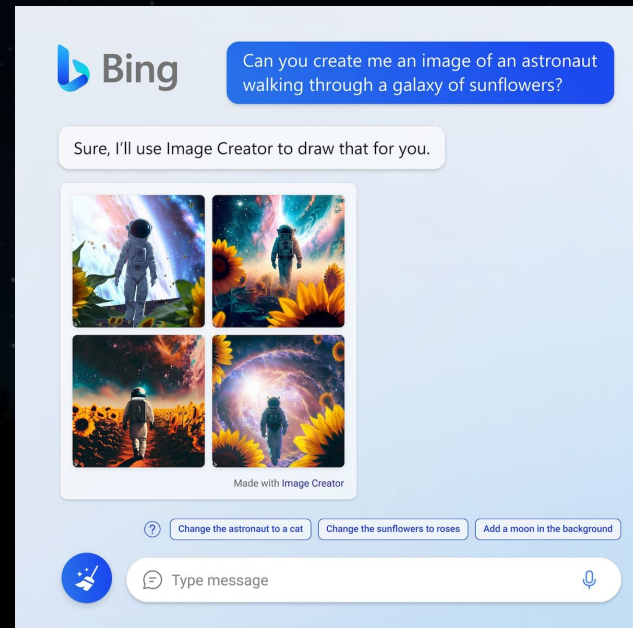
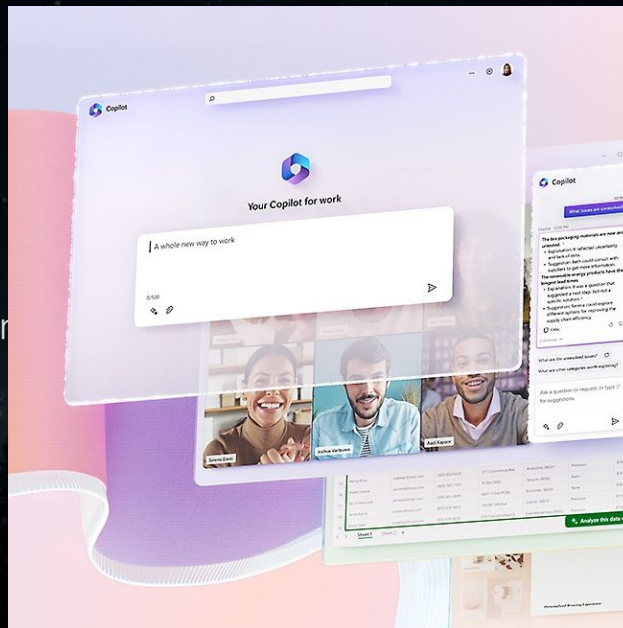
Some lowlights

- [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](#).

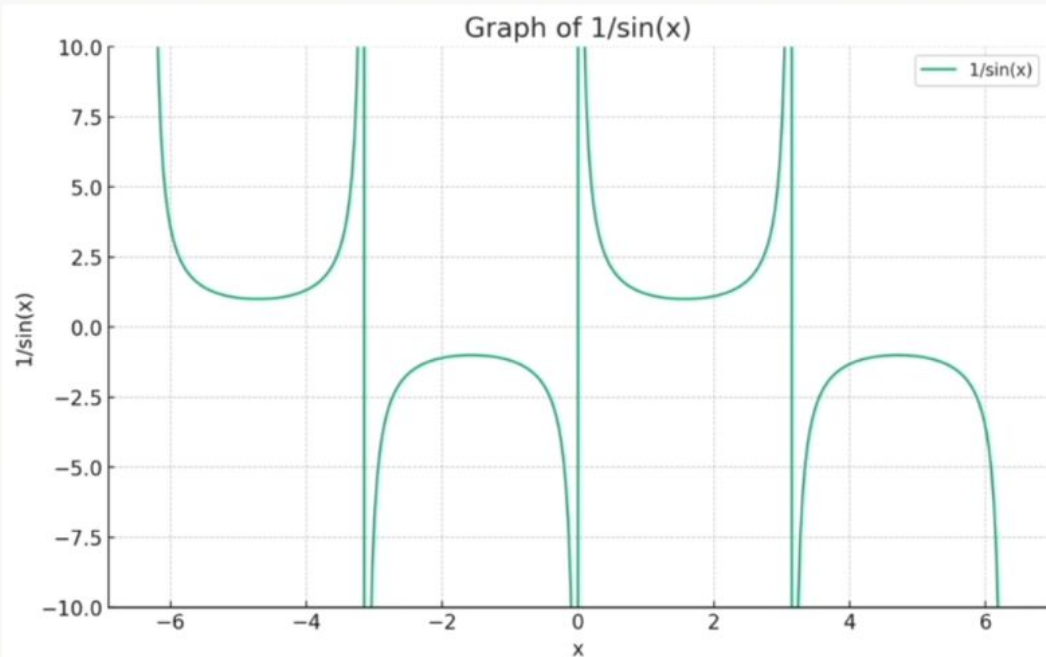


plot function $1/\sin(x)$



Finished calculating

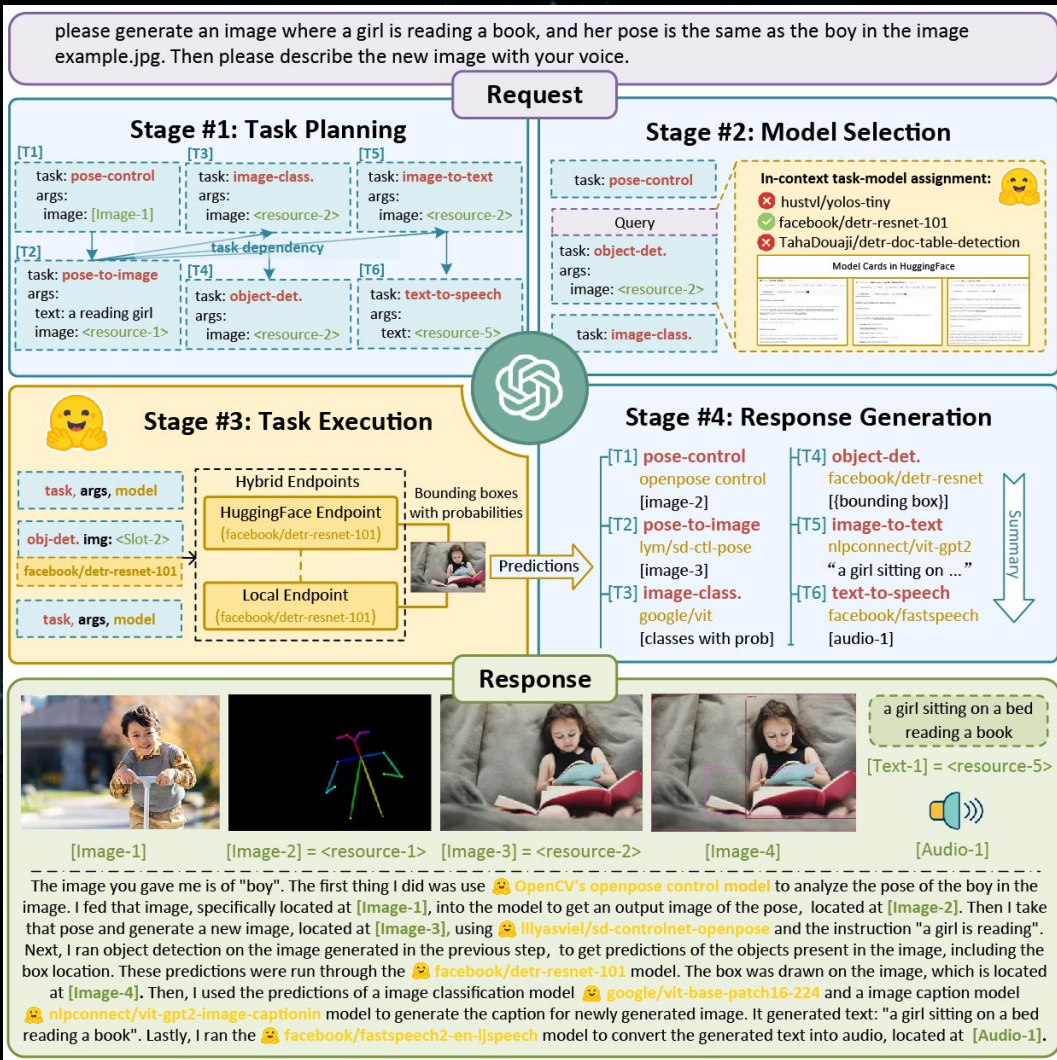
Show work ▾



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

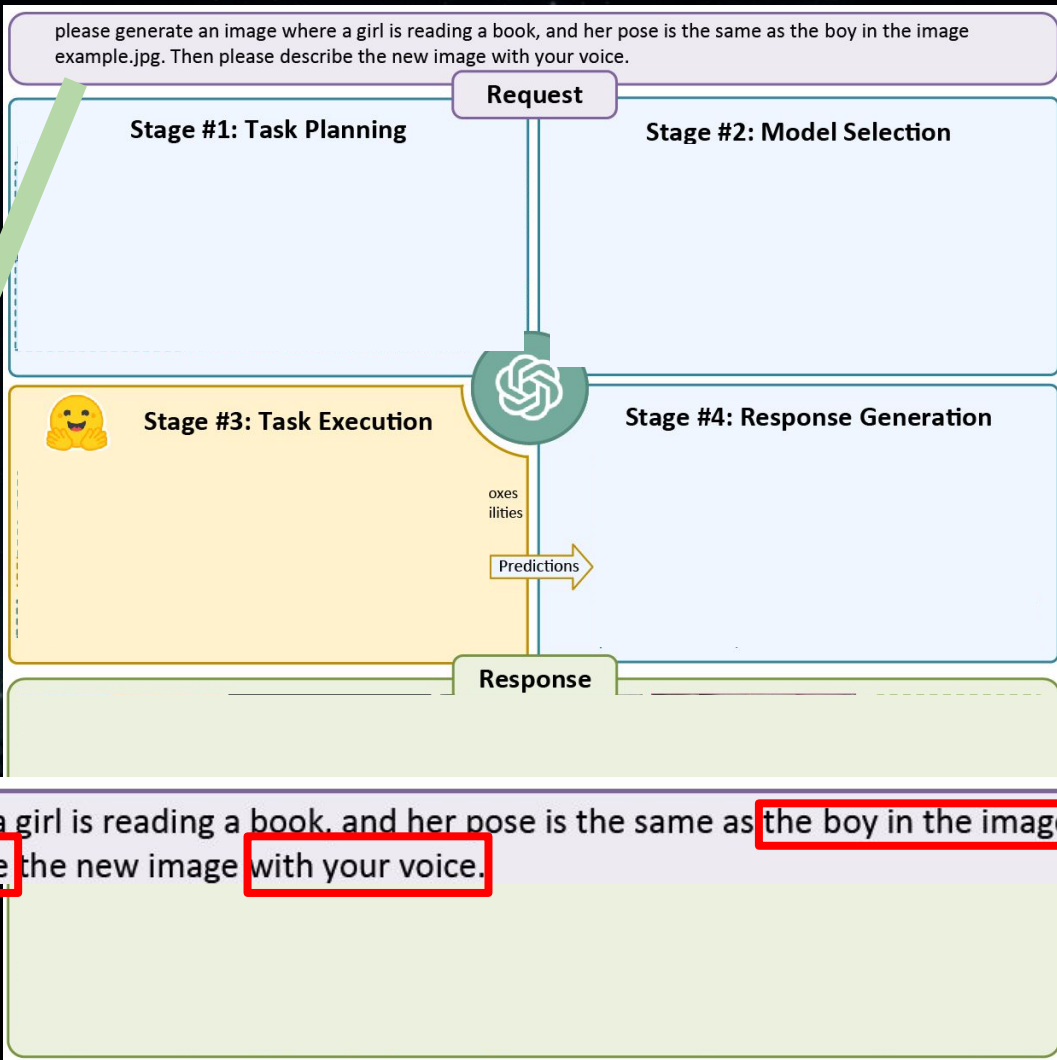
HuggingGPT

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



HuggingGPT

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



Alpaca

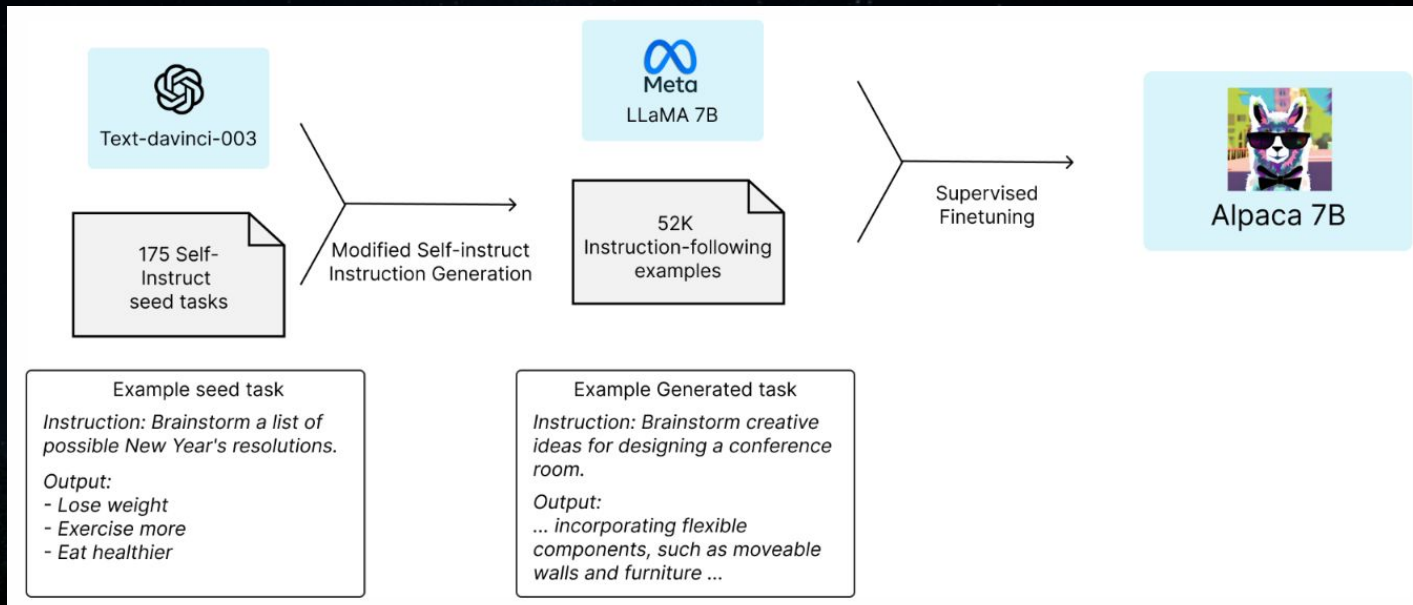
Finetune your ChatGPT with
<\$600

Opens a door for cheap
academic LLM research

Related:

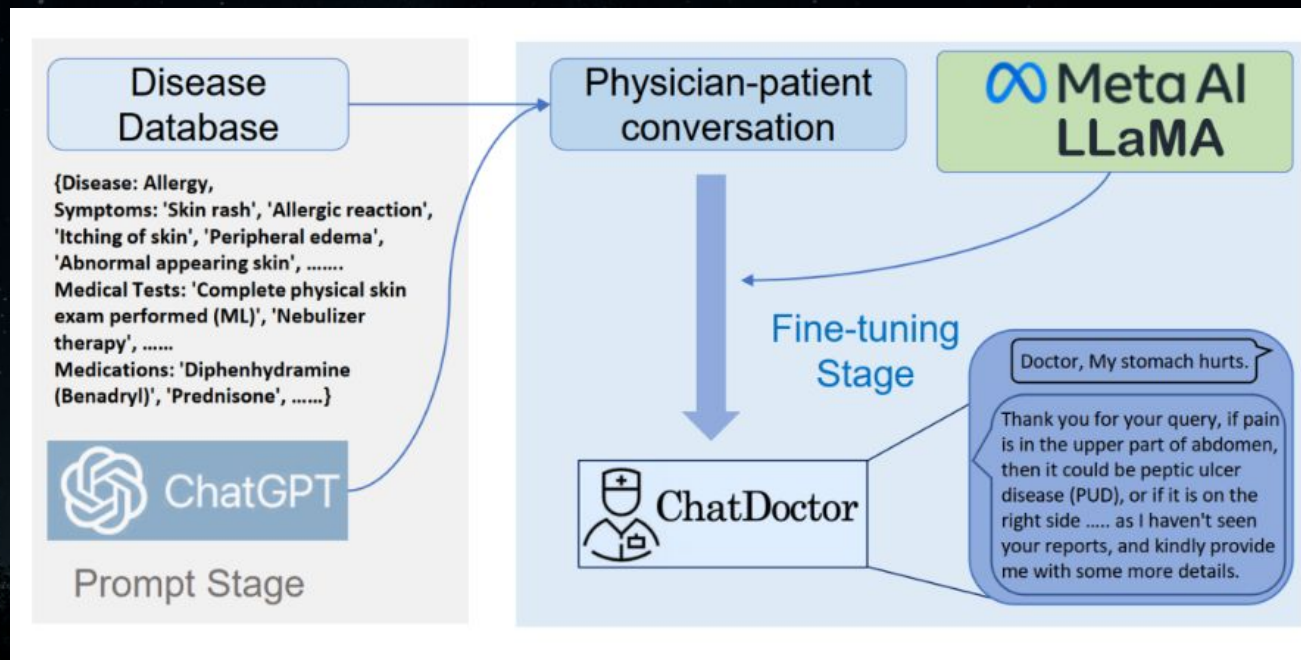
[Berkeley Vicuna](#) (90% chatgpt
with less than \$300!!!)

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



ChatDoctor, similar to Alpaca

ChatGPT indirectly
powered LLM



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

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

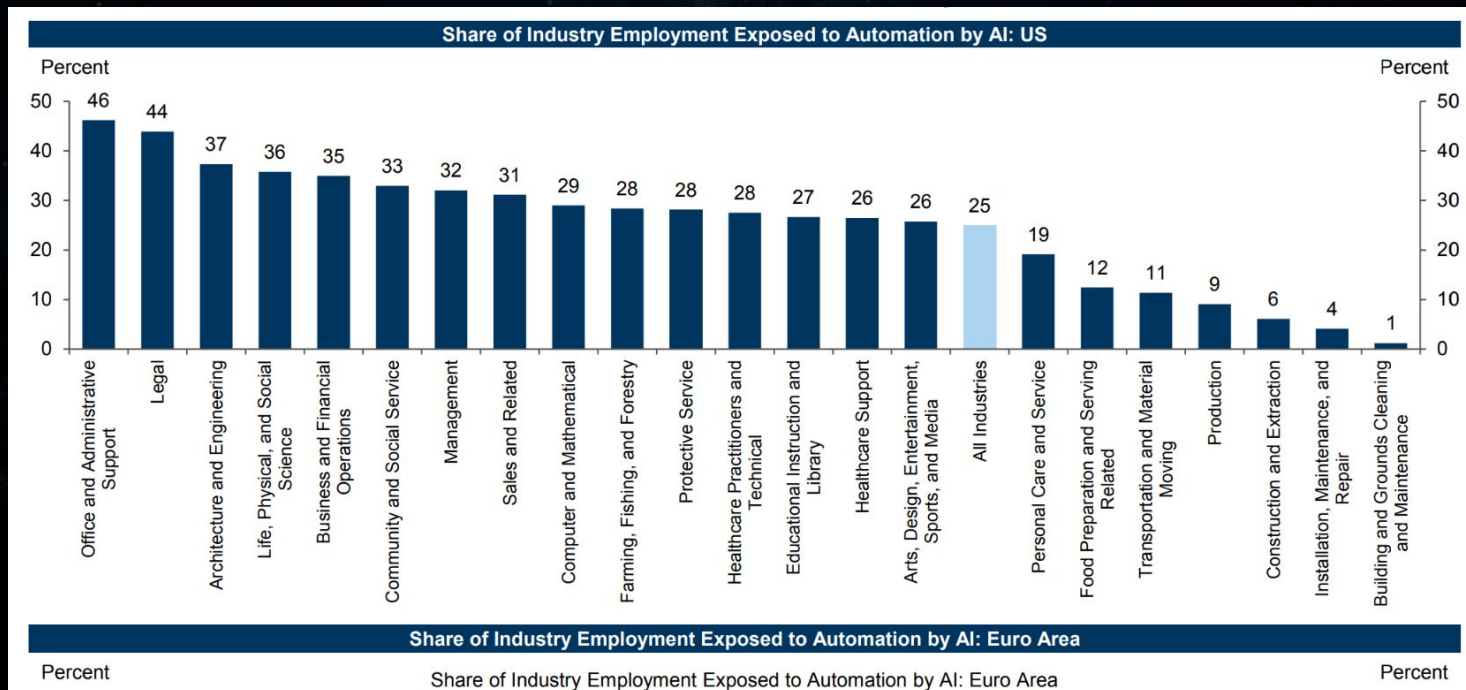
“Our findings reveal that around 80% of the U.S. workforce could have at least 10% of their work tasks **affected** by the introduction of LLMs, while approximately 19% of workers may see at least 50% of their tasks **impacted**.”

- *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

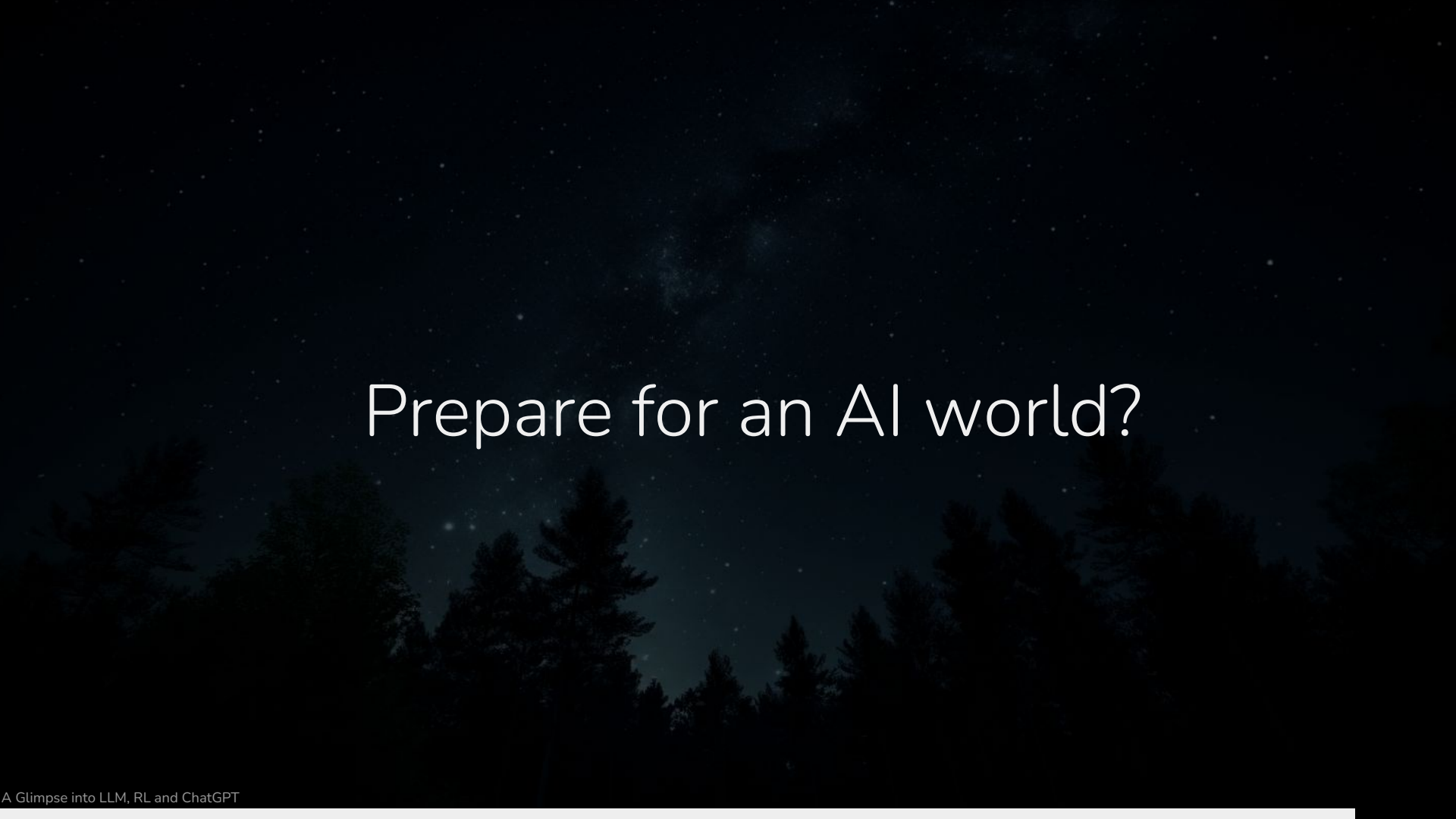
9:30 AM to 10:00 AM Pacific Time



Yann LeCun
VP and Chief AI Scientist
Meta

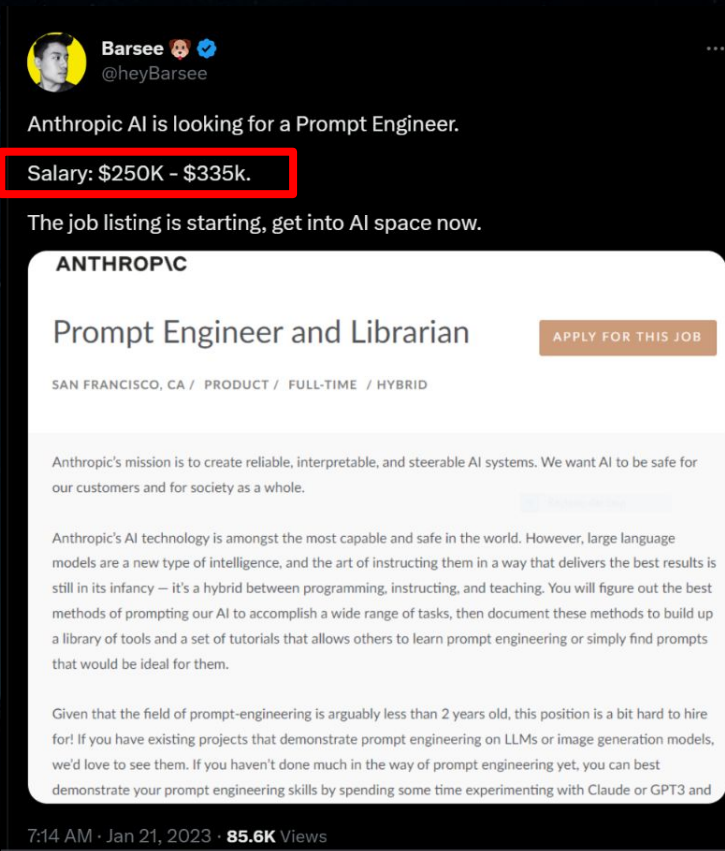


Andrew Ng
Founder
DeepLearning.AI



Prepare for an AI world?

Maybe switch jobs to the hottest “Prompt Engineer”?



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?

fun-ai-talk

Thank you! Questions?

A Glimpse Into LLM, RL and ChatGPT

hululu.zhu@gmail.com

April 2023