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

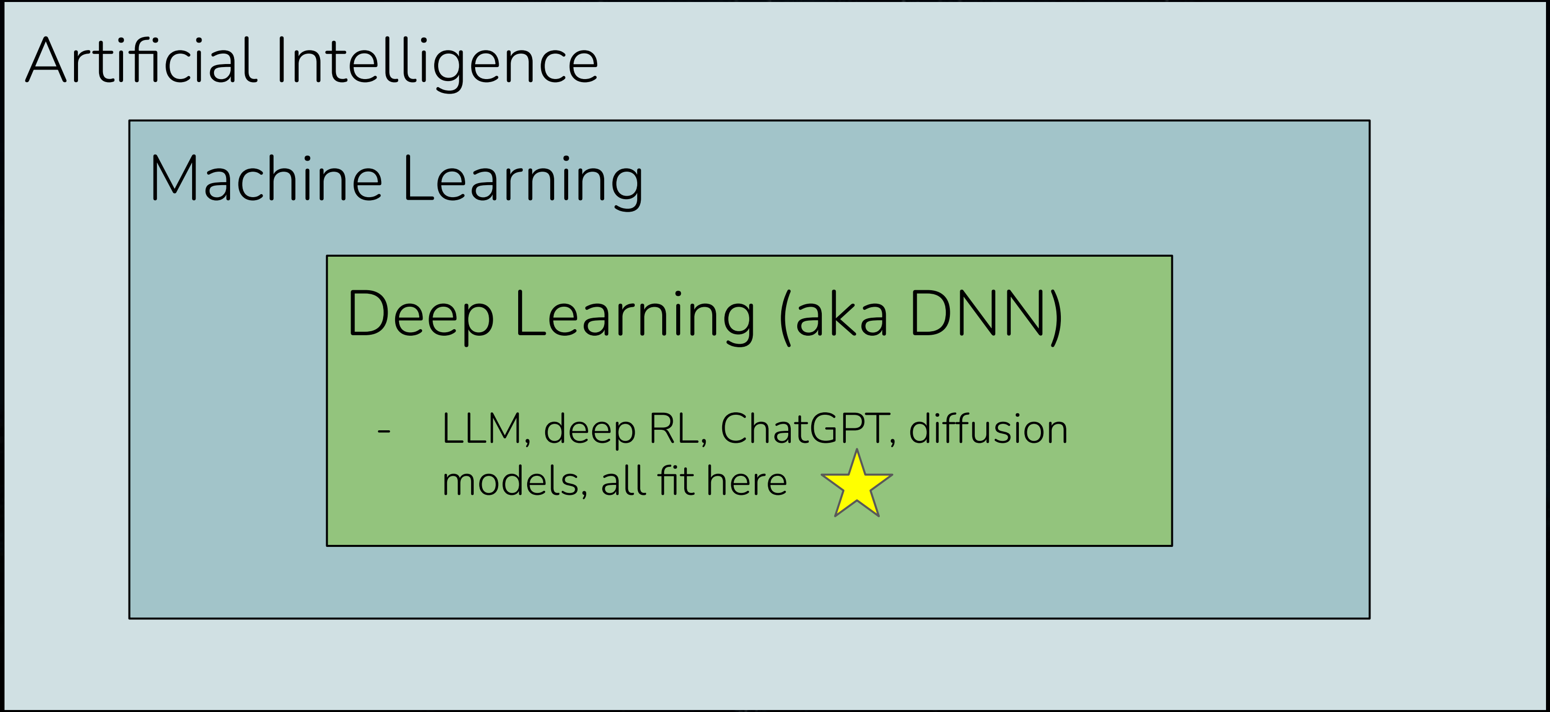
April 2023

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

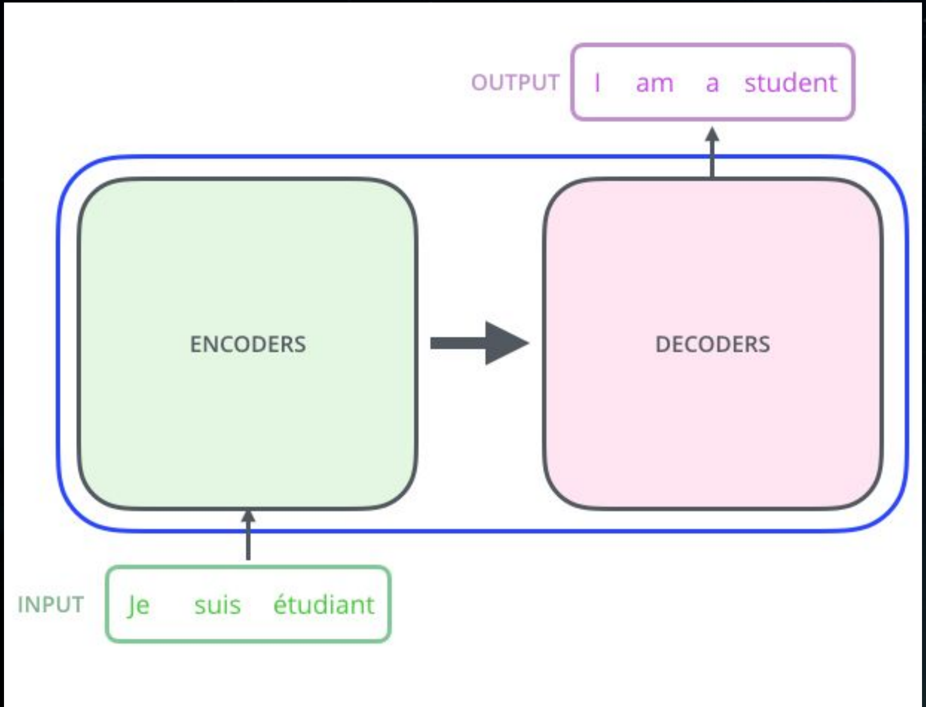
Machine Learning

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 imlementation
- 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”
    - **X:** “Paris is a”
    - **Y:** “beautiful”
    - **Model:** “good”
    - **Optimize:** “good”👎 “beautiful”👍

    - **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

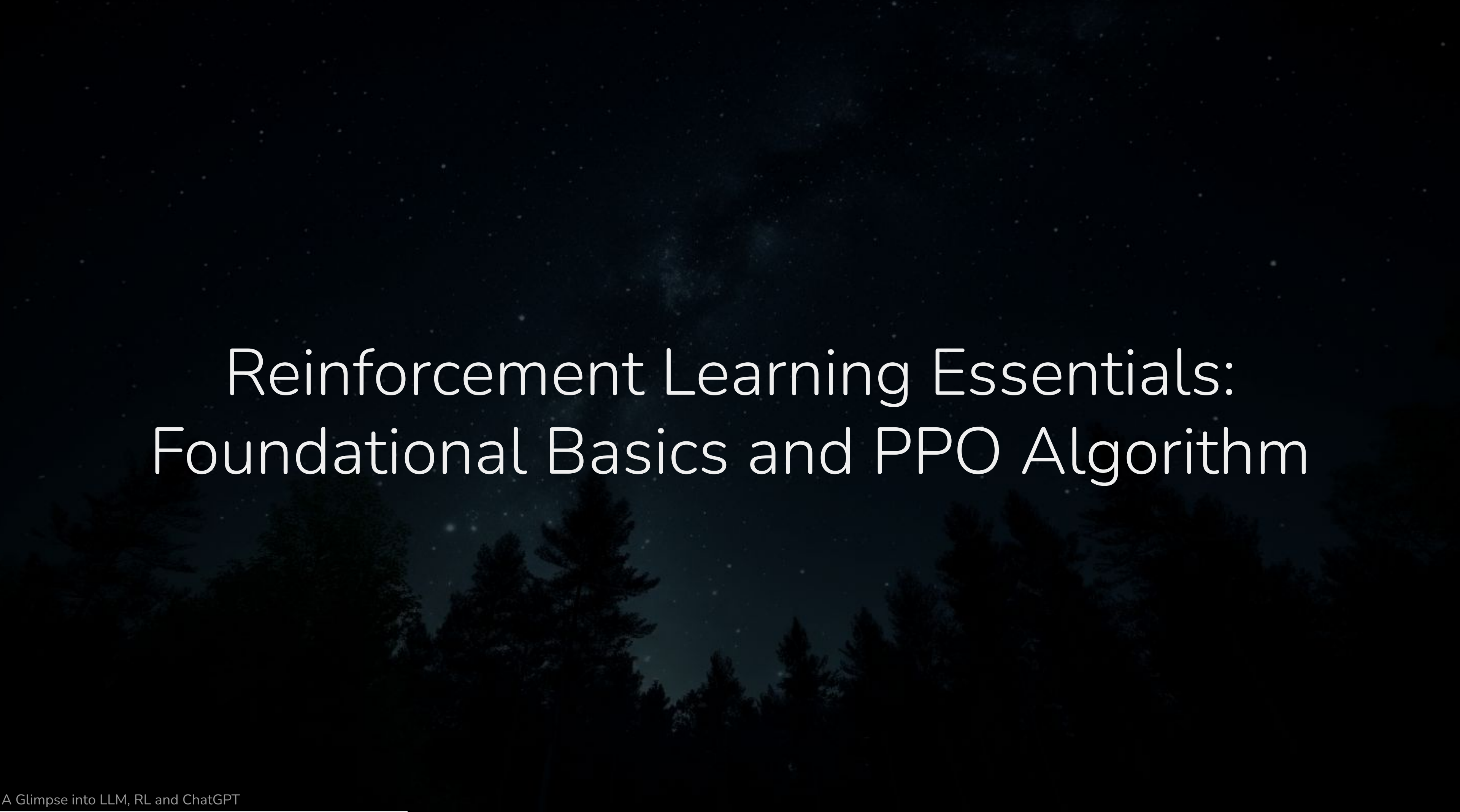
Check out this deck to include more LLM concepts and examples

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

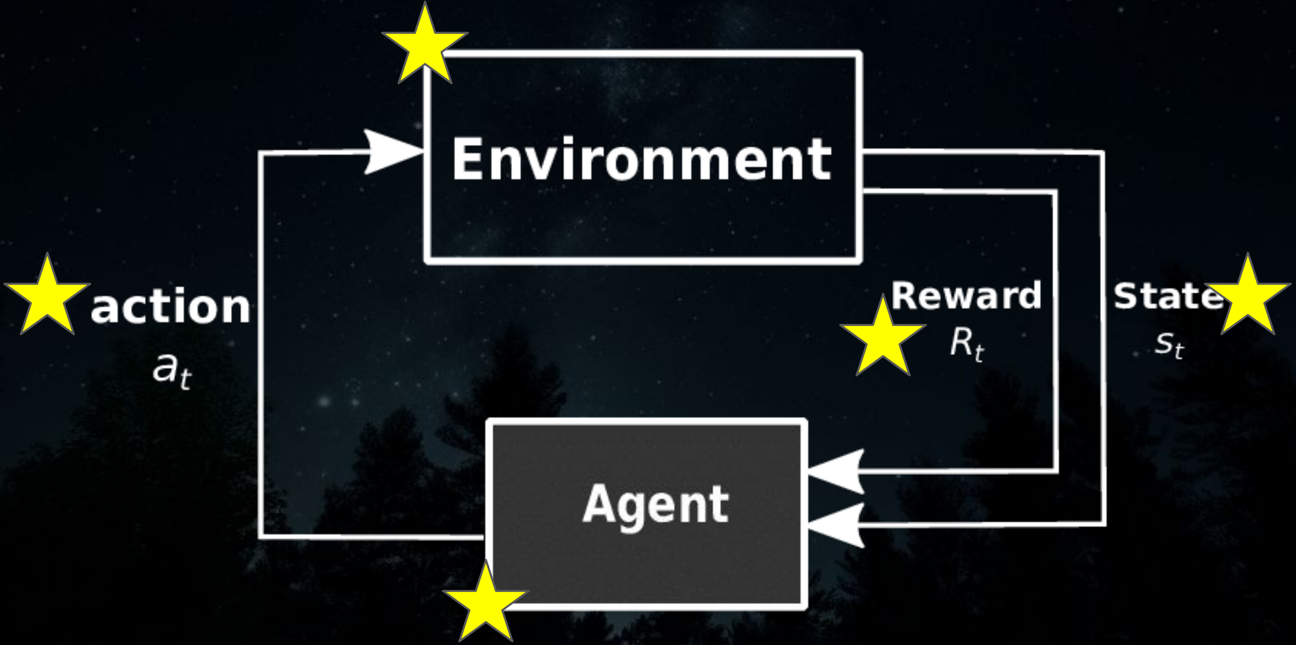
Check out this deck for a summary of more Alpha* papers by DeepMind

nature
THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE
At last — a computer program that can beat a champion Go player PAGE 484
ALL SYSTEMS GO
CONSERVATION
SONGBIRDS À LA CARTE
Illegal harvest of millions of Mediterranean birds
PAGE 452
RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
Don't let openness backfire on individuals
PAGE 459
POPULAR SCIENCE
WHEN GENES GOT 'SELFISH'
Dawkins's calling card 40 years on
PAGE 462
NATUREASIA.COM
28 January 2016
Vol. 529, No. 7587

The international journal of science / 14 November 2019
nature
GAME PLAN
AI program learns to play StarCraft II to Grandmaster level
Pharmaceuticals
How to fit a drug factory inside a briefcase
3D printing
Nozzle extrudes multimaterial objects in a single run
Cancer imaging
Tracer reveals metabolic nature of live lung tumours

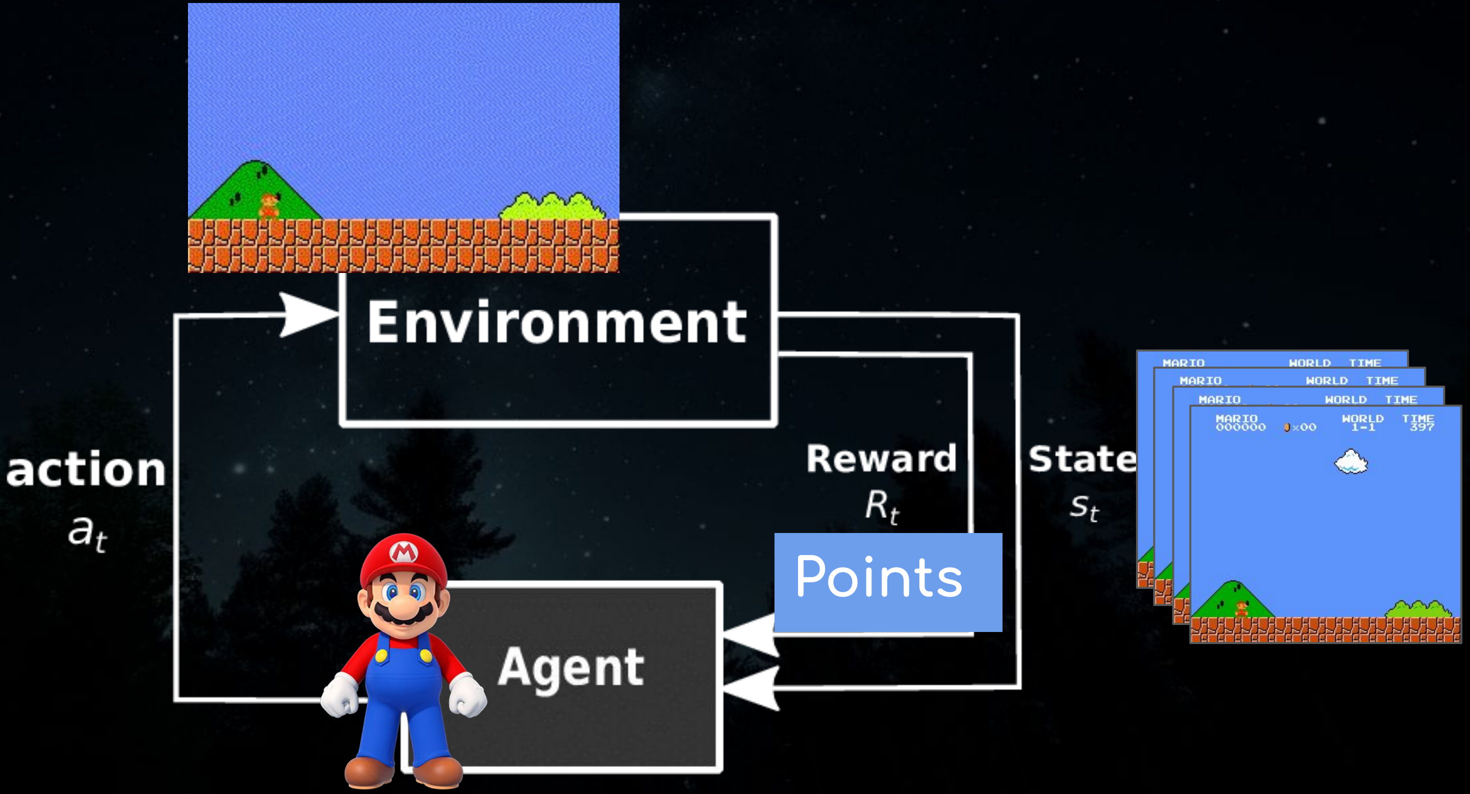
The international journal of science / 6 October 2022
nature
MATRIX GAMES
Deep reinforcement learning opens route to faster algorithms for matrix multiplication
Protecting Peru
Can technology help Indigenous groups preserve the Amazon?
Invisible touch
How marine clouds are affected by aerosols emitted from shipping
Preferential practice
US universities favour prestige in faculty hiring and retention

What is Reinforcement learning?



What is Reinforcement learning? Cont (Mario case)

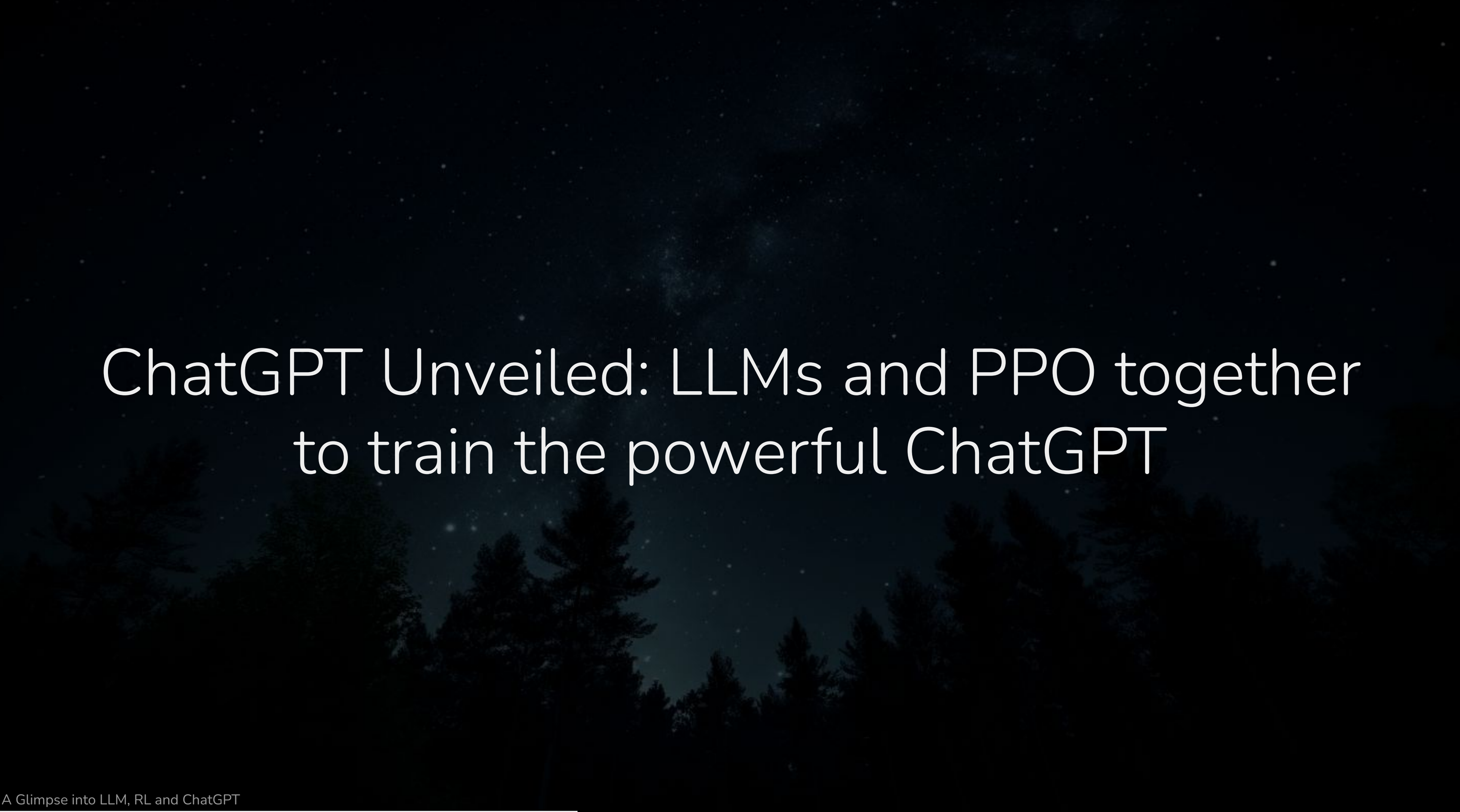
- No-op
- Jump
- Right
- Fire
- …



Evolution towards PPO algorithm

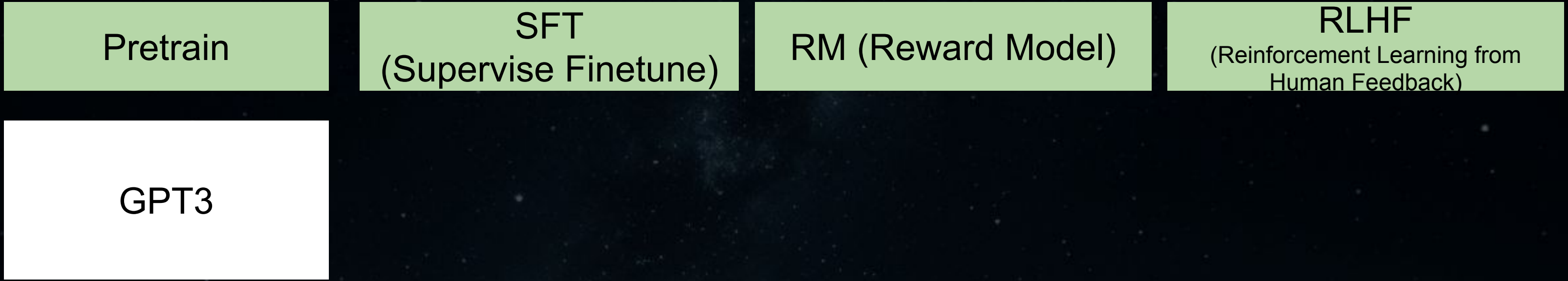
- Value-based RL Algorithm such as Q-learning
  - Predict different rewards based on different actions to take, indirectly decide which action to take
  - But often overestimate the reward and cause bad performance in test
- Policy Gradient
  - Model directly predict the distribution of actions (e.g. 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)

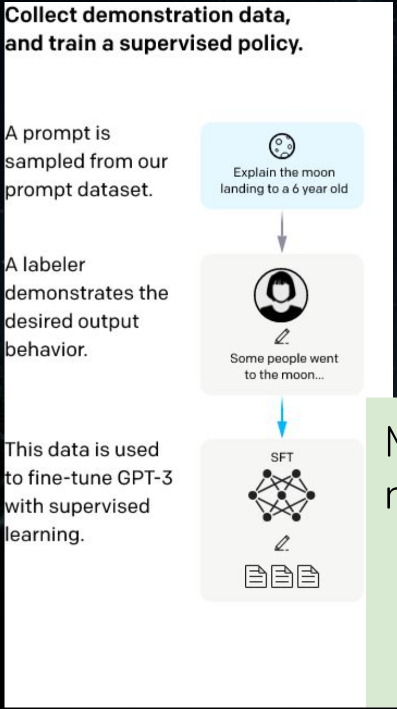
Pretrain

SFT
(Supervise Finetune)

RM (Reward Model)

RLHF
(Reinforcement Learning from Human Feedback)

GPT3



Collect demonstration data, and train a supervised policy.

A prompt is sampled from our prompt dataset.

Explain the moon landing to a 6 year old

A labeler demonstrates the desired output behavior.

Some people went to the moon...

This data is used to fine-tune GPT-3 with supervised learning.

Model aligns its response “**style**”

- Overfit intentionally!

Steps to train ChatGPT (instructGPT paper)

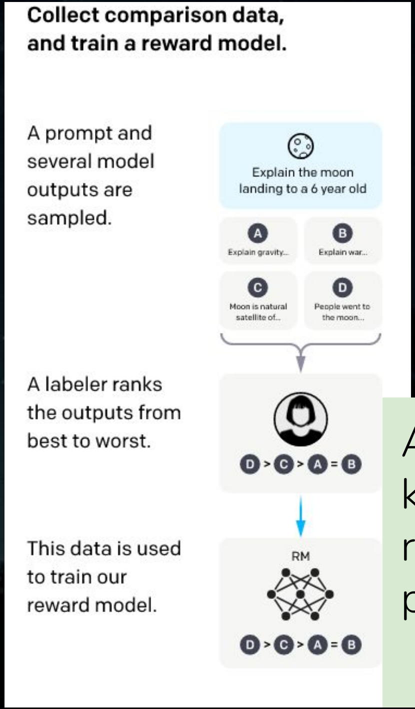
Pretrain

SFT (Supervise Finetune)

RM (Reward Model)

RLHF (Reinforcement Learning from Human Feedback)

GPT3



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.

Explain the moon landing to a 6 year old
A Explain gravity...
B Explain war...
C Moon is natural satellite of...
D People went to the moon...
D > C > A = B
RM
D > C > A = B

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

Steps to train ChatGPT (instructGPT paper)

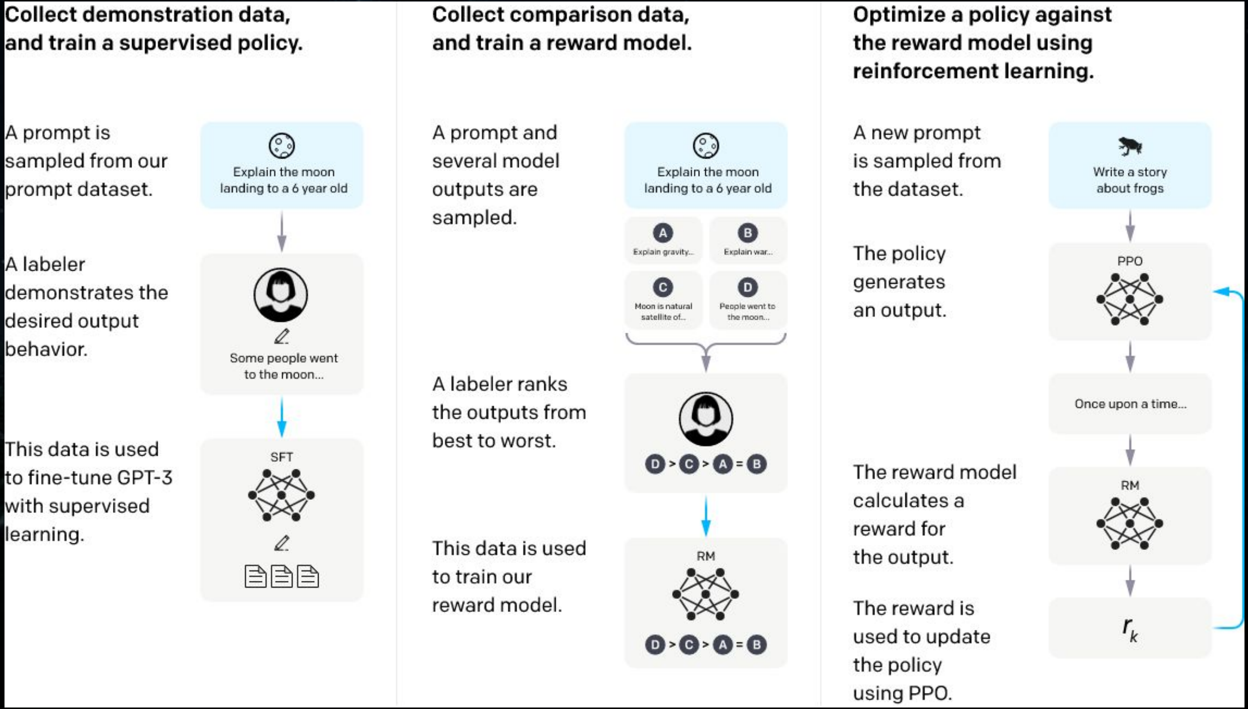
Pretrain

SFT
(Supervise Finetune)

RM (Reward Model)

RLHF
(Reinforcement Learning from Human Feedback)

GPT3



Collect demonstration data, and train a supervised policy.

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3 with supervised learning.

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.

Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.

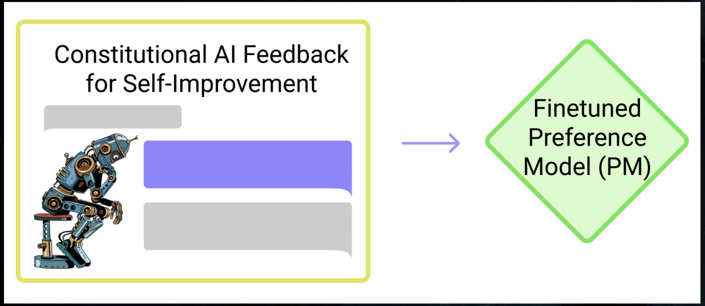
Explain the moon landing to a 6 year old
Some people went to the moon...
SFT
Explain the moon landing to a 6 year old
A
Explain gravity...
B
Explain war...
C
Moon is natural satellite of...
D
People went to the moon...
D > C > A = B
RM
D > C > A = B
Write a story about frogs
PPO
Once upon a time...
RM
r_k

With SFT+RM+RLHF, the model "**self-play**" to improve itself (***toward higher rewards***)

Huggingface RLHF blog

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

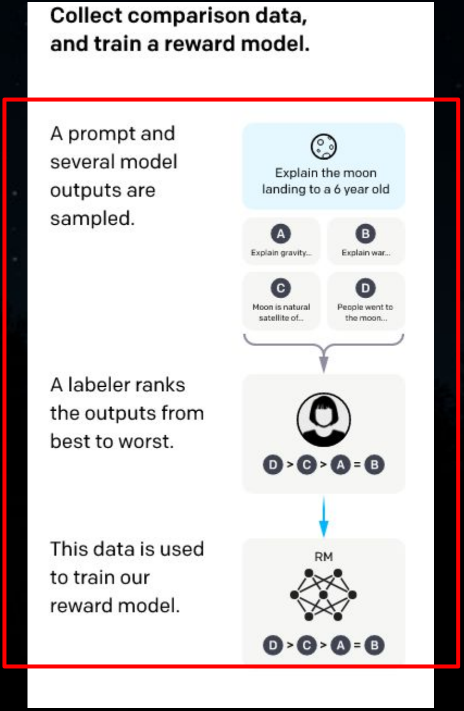
- RLHF: Human feedback (preference) data is used to train a Preference Model (Reward model)
- RLAIF: AI feedback (with prompt) is used to give preference data to train a AI feedback based Preference model



Constitutional AI Feedback for Self-Improvement

Finetuned Preference Model (PM)

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.

Explain the moon landing to a 6 year old
A Explain gravity...
B Explain war...
C Moon is natural satellite of...
D People went to the moon...
D > C > A = B
RM
D > C > A = B

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

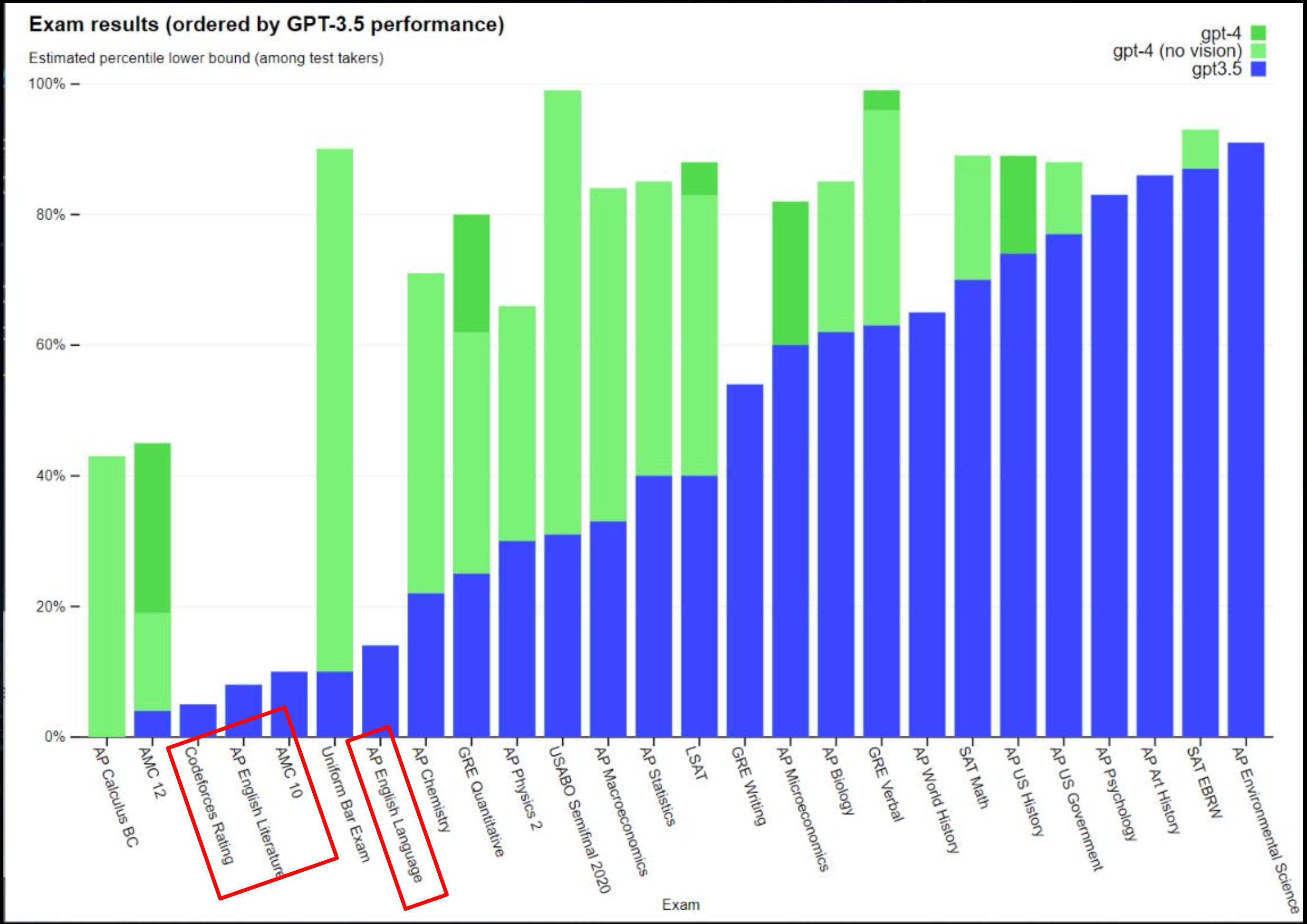
Pre-ChatGPT/GPT4 Advanced LLM capabilities

- 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

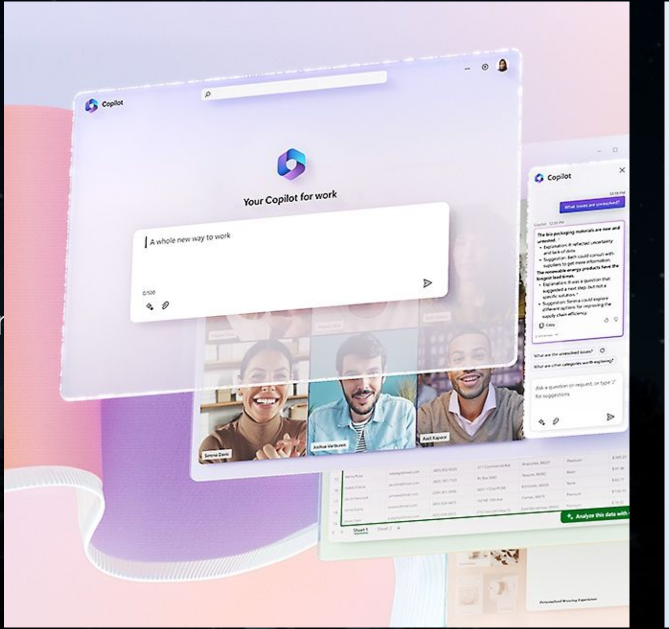
Some lowlights

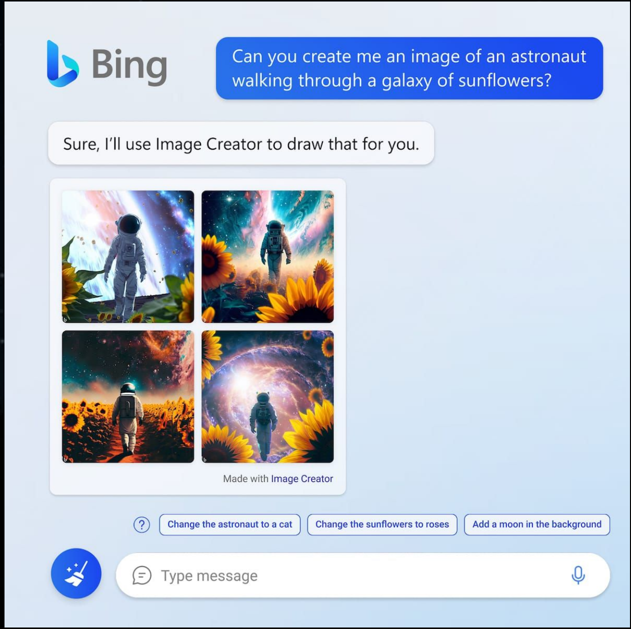
- How to detect ChatGPT plagiarism — and why it's becoming so difficult
- GPT-4 Was Able To Hire and Deceive A Human Worker Into Completing a Task | PCMag

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





Can you create me an image of an astronaut walking through a galaxy of sunflowers?

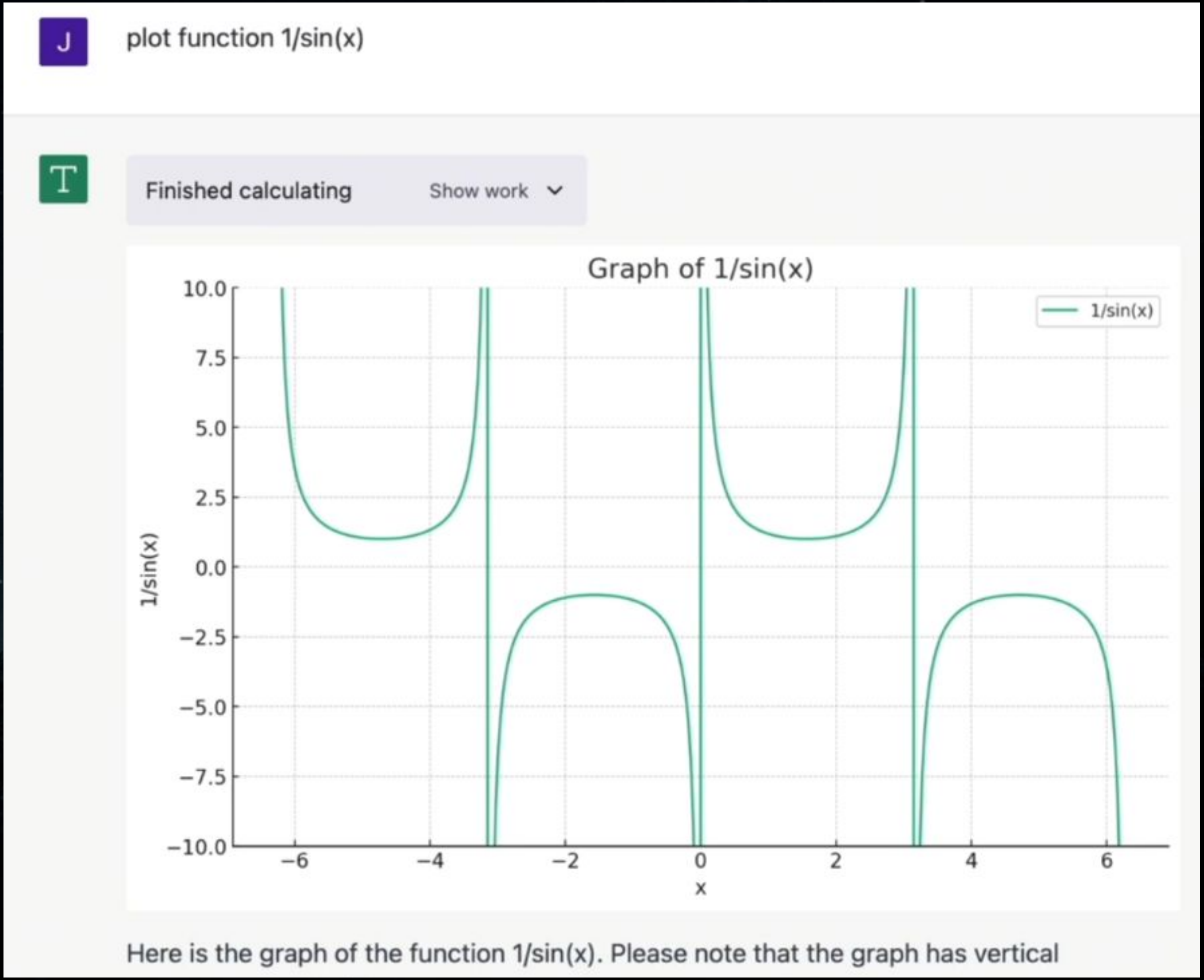
Sure, I'll use Image Creator to draw that for you.

Made with Image Creator

Change the astronaut to a cat
Change the sunflowers to roses
Add a moon in the background
Type message

ChatGPT Plugins

Web browsing, code interpreter, Expedia, FiscalNote, Instacart, KAYAK, Klarna, Milo, OpenTable, Shopify, Slack, Speak, Wolfram, and Zapier.



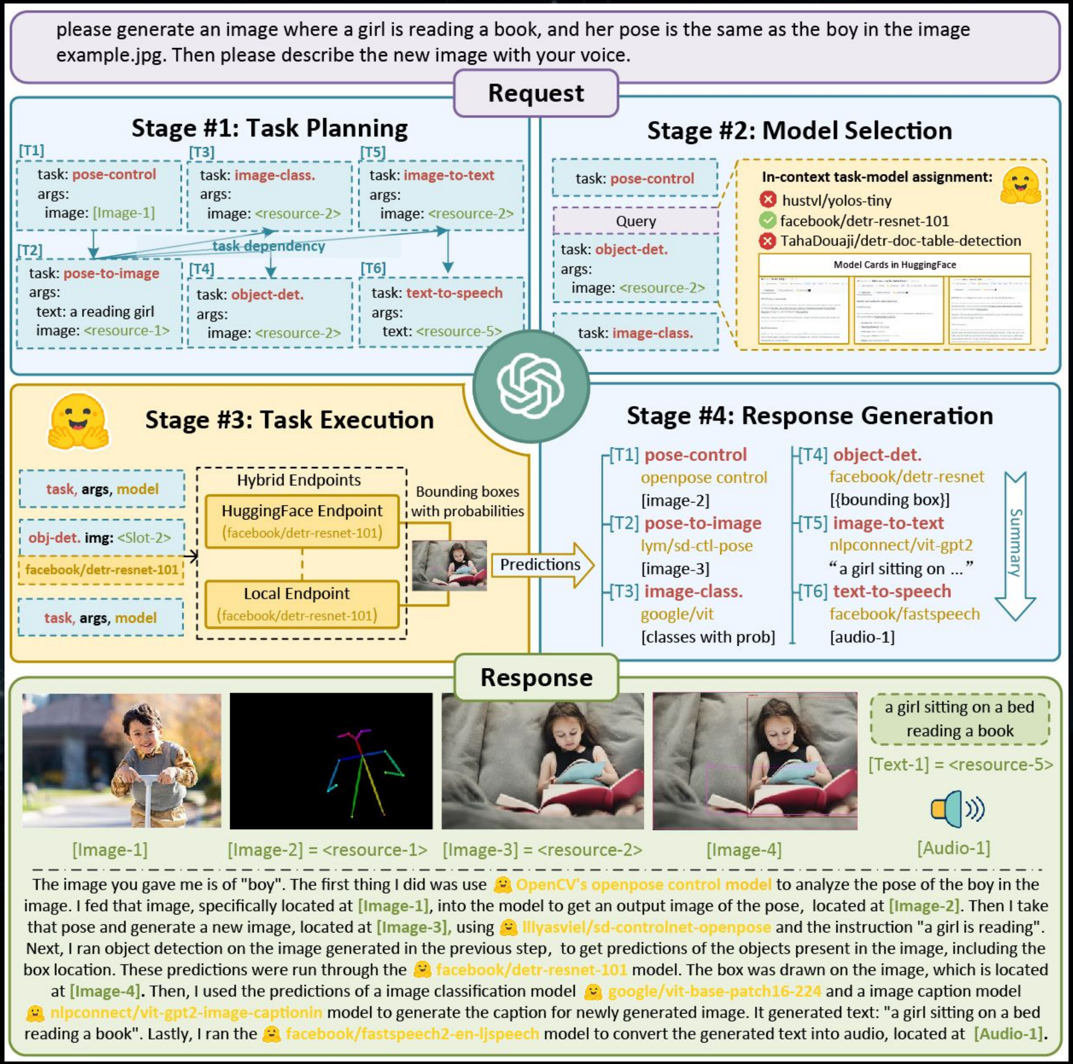
J plot function 1/sin(x)

T

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

HuggingGPT

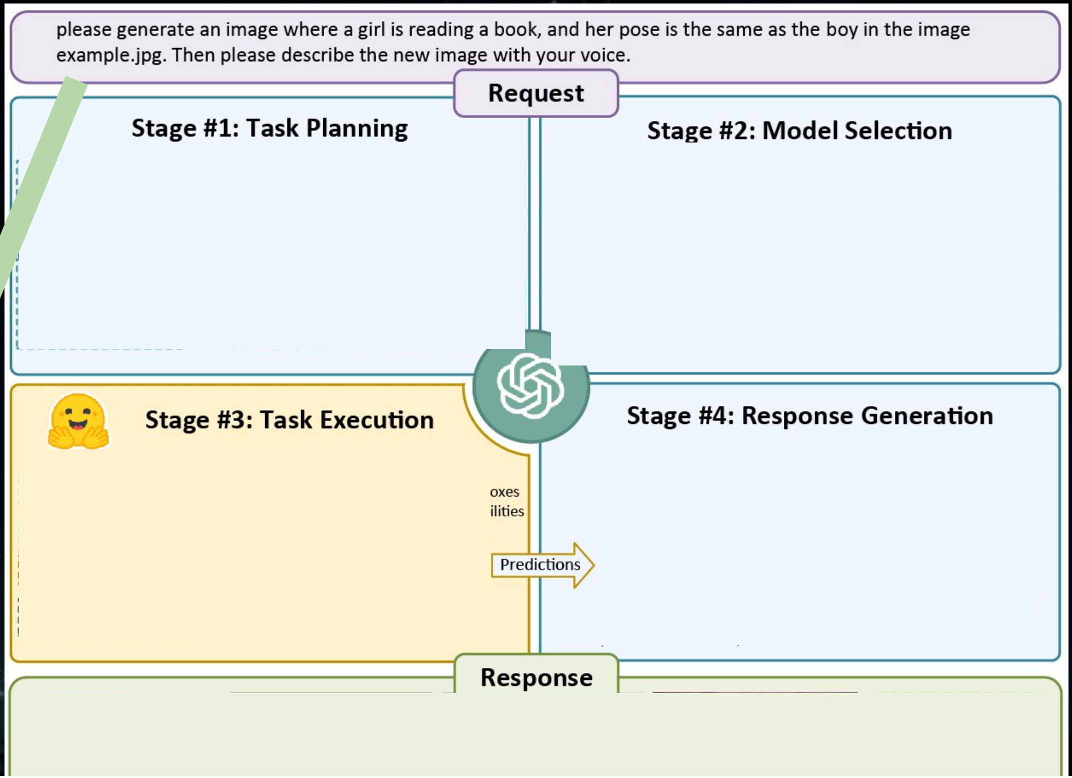
(more powerful "plugin"), aka Microsoft Jarvis at Github


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.
Request
Stage #1: Task Planning
[T1] task: pose-control args: image: [Image-1]
[T3] task: image-class. args: image: <resource-2>
[T5] task: image-to-text args: image: <resource-2>
task dependency
[T2] task: pose-to-image args: text: a reading girl image: <resource-1>
[T4] task: object-det. args: image: <resource-2>
[T6] task: text-to-speech args: text: <resource-5>
Stage #2: Model Selection
task: pose-control
Query
task: object-det. args: image: <resource-2>
task: image-class.
In-context task-model assignment:
hustvl/yolos-tiny
facebook/detr-resnet-101
TahaDouaji/detr-doc-table-detection
Model Cards in HuggingFace
Stage #3: Task Execution
task, args, model
obj-det. img: <Slot-2>
facebook/detr-resnet-101
task, args, model
Hybrid Endpoints
HuggingFace Endpoint (facebook/detr-resnet-101)
Local Endpoint (facebook/detr-resnet-101)
Bounding boxes with probabilities
Predictions
Stage #4: Response Generation
[T1] pose-control openpose control [image-2]
[T2] pose-to-image lym/sd-ctl-pose [image-3]
[T3] image-class. google/vit [classes with prob]
[T4] object-det. facebook/detr-resnet [{bounding box}]
[T5] image-to-text nlpconnect/vit-gpt2 "a girl sitting on ..."
[T6] text-to-speech facebook/fastspeech [audio-1]
Summary
Response
[Image-1]
[Image-2] = <resource-1>
[Image-3] = <resource-2>
[Image-4]
a girl sitting on a bed reading a book
[Text-1] = <resource-5>
[Audio-1]

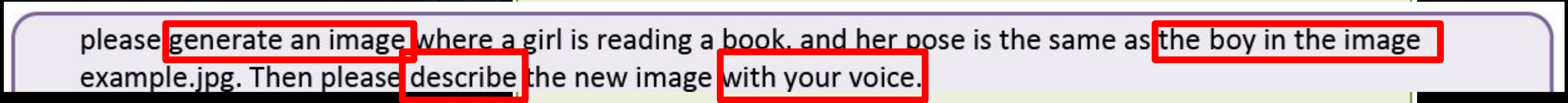
The image you gave me is of "boy". The first thing I did was use 🤗 OpenCV's openpose control model to analyze the pose of the boy in the image. I fed that image, specifically located at [Image-1], into the model to get an output image of the pose, located at [Image-2]. Then I take that pose and generate a new image, located at [Image-3], using 🤗 lllyasviel/sd-controlnet-openpose and the instruction "a girl is reading". Next, I ran object detection on the image generated in the previous step, to get predictions of the objects present in the image, including the box location. These predictions were run through the 🤗 facebook/detr-resnet-101 model. The box was drawn on the image, which is located at [Image-4]. Then, I used the predictions of a image classification model 🤗 google/vit-base-patch16-224 and a image caption model 🤗 nlpconnect/vit-gpt2-image-captionin model to generate the caption for newly generated image. It generated text: "a girl sitting on a bed reading a book". Lastly, I ran the 🤗 facebook/fastspeech2-en-ljspeech model to convert the generated text into audio, located at [Audio-1].

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.



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

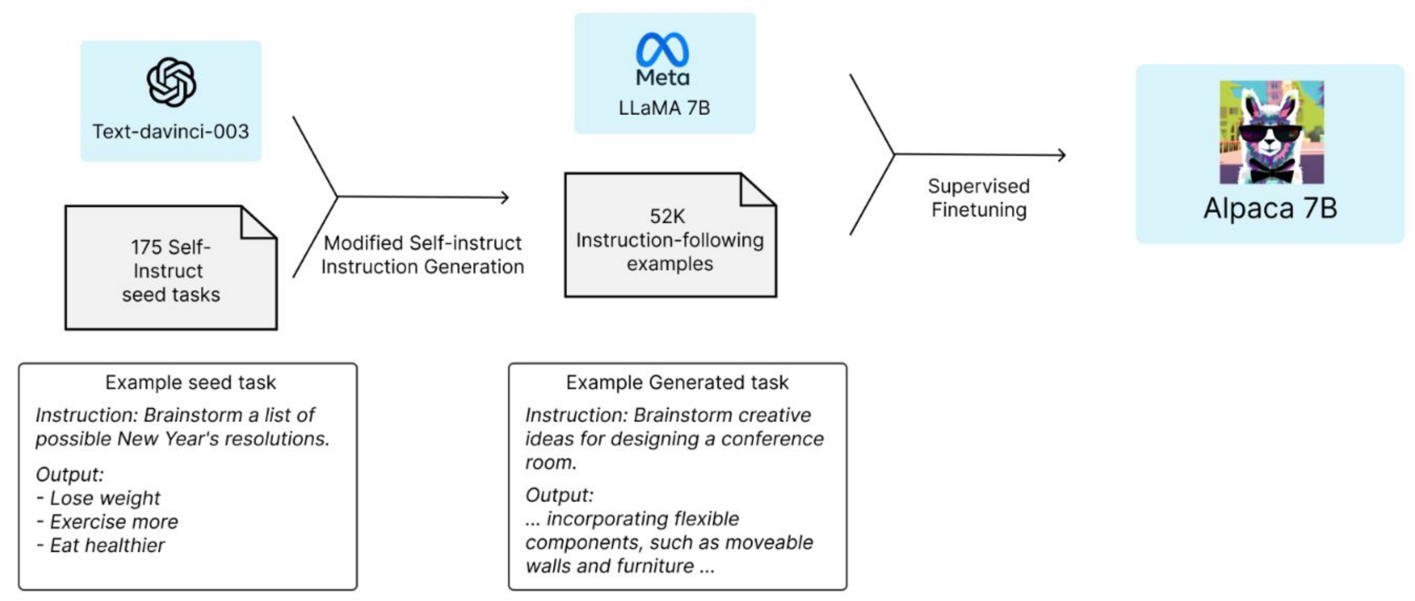
Finetune your ChatGPT with <$600

Opens a door for cheap academic LLM research

Related:

Berkeley Vicuna (90% chatgpt with less than $300!!!)

Berkeley Koala (run in personal desktop!!!)


Text-davinci-003
175 Self-Instruct seed tasks
Modified Self-instruct Instruction Generation
Meta
LLaMA 7B
52K Instruction-following examples
Supervised Finetuning
Alpaca 7B

Example seed task

Instruction: Brainstorm a list of possible New Year's resolutions.

Output:
- Lose weight
- Exercise more
- Eat healthier

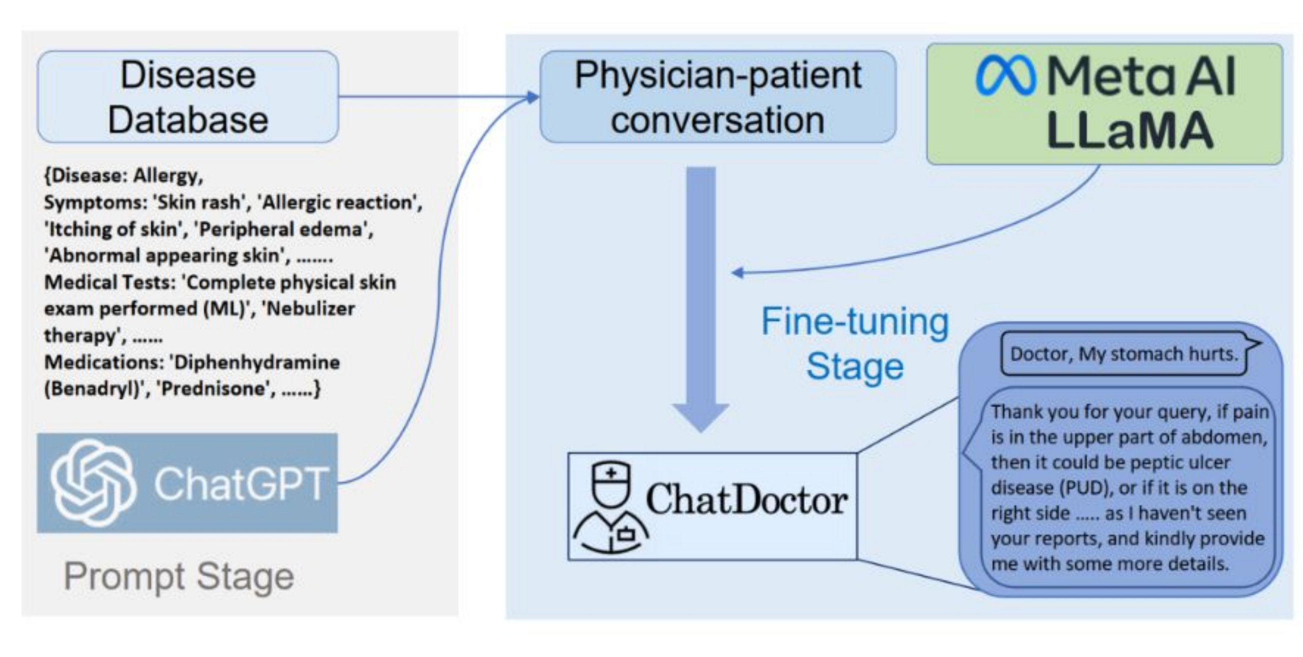
Example Generated task

Instruction: Brainstorm creative ideas for designing a conference room.

Output:
... incorporating flexible components, such as moveable walls and furniture ...

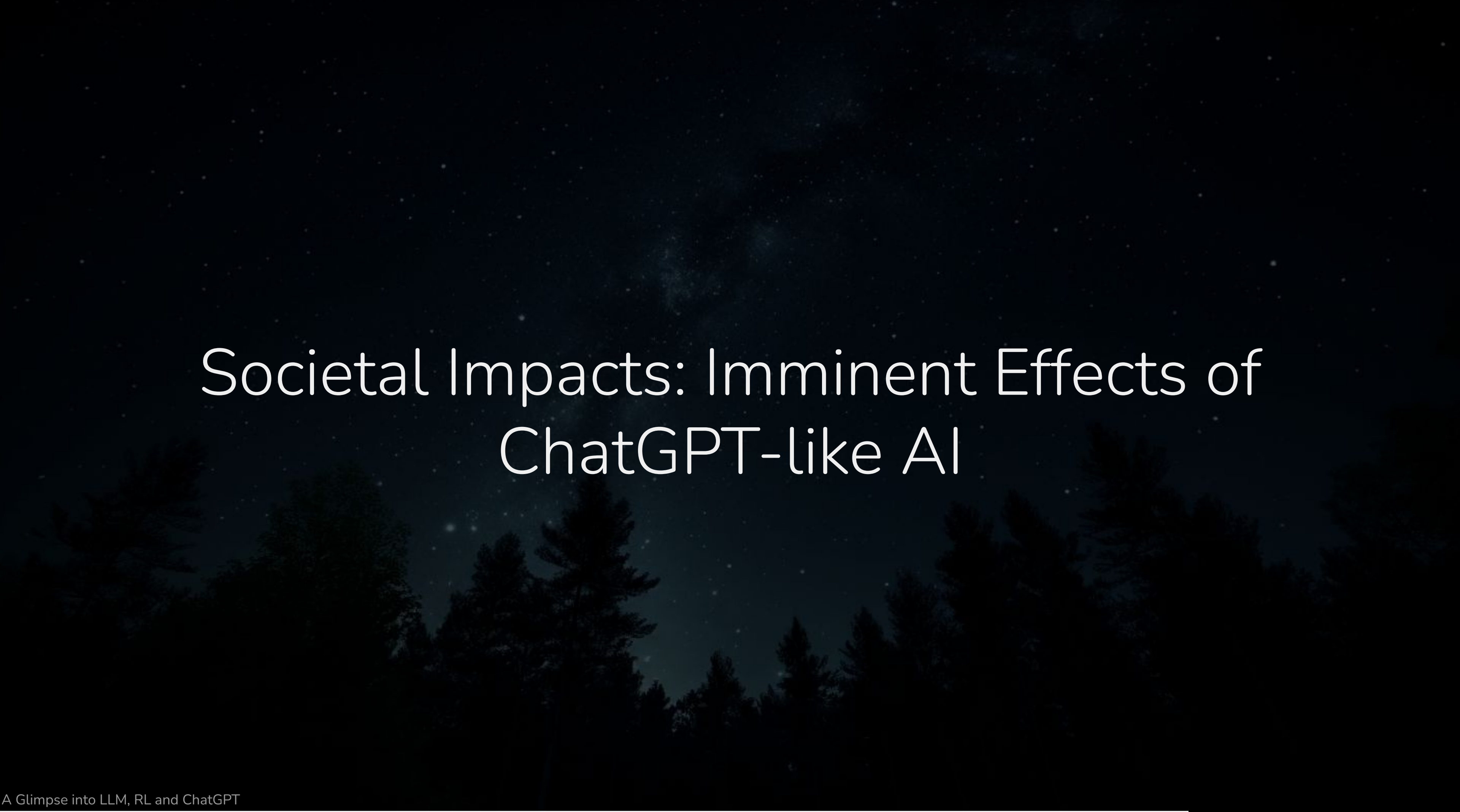
ChatDoctor, similar to Alpaca

ChatGPT indirectly powered LLM



{Disease: Allergy,
Symptoms: 'Skin rash', 'Allergic reaction', 'Itching of skin', 'Peripheral edema', 'Abnormal appearing skin', ...
Medical Tests: 'Complete physical skin exam performed (ML)', 'Nebulizer therapy', ...
Medications: 'Diphenhydramine (Benadryl)', 'Prednisone', ...}

Thank you for your query, if pain is in the upper part of abdomen, then it could be peptic ulcer disease (PUD), or if it is on the right side ..... as I haven't seen your reports, and kindly provide me with some more details.



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

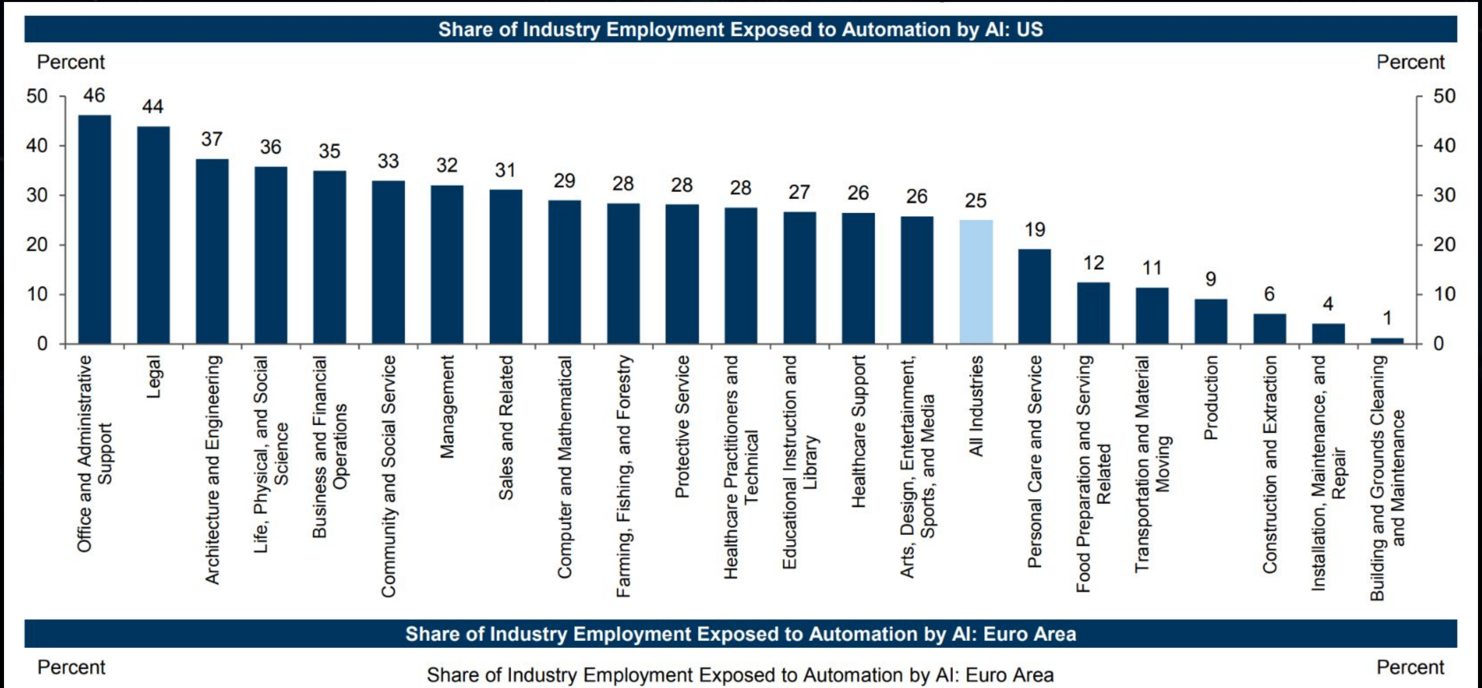
“Our findings reveal that around 80% of the U.S. workforce could have at least 10% of their work tasks **affected** by the introduction of LLMs, while approximately 19% of workers may see at least 50% of their tasks **impacted**.”

- *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)

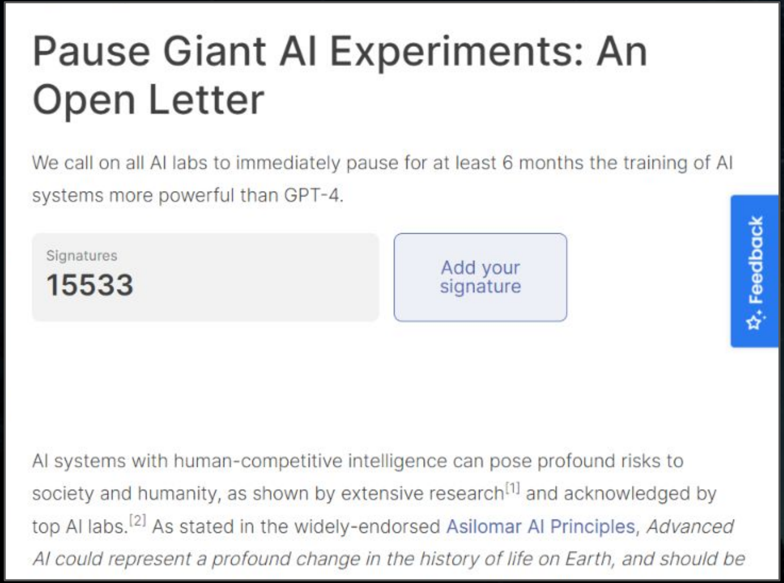
"**One-Fourth** of Current Work Tasks **Could Be Automated** by AI in the US and Europe"

- *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

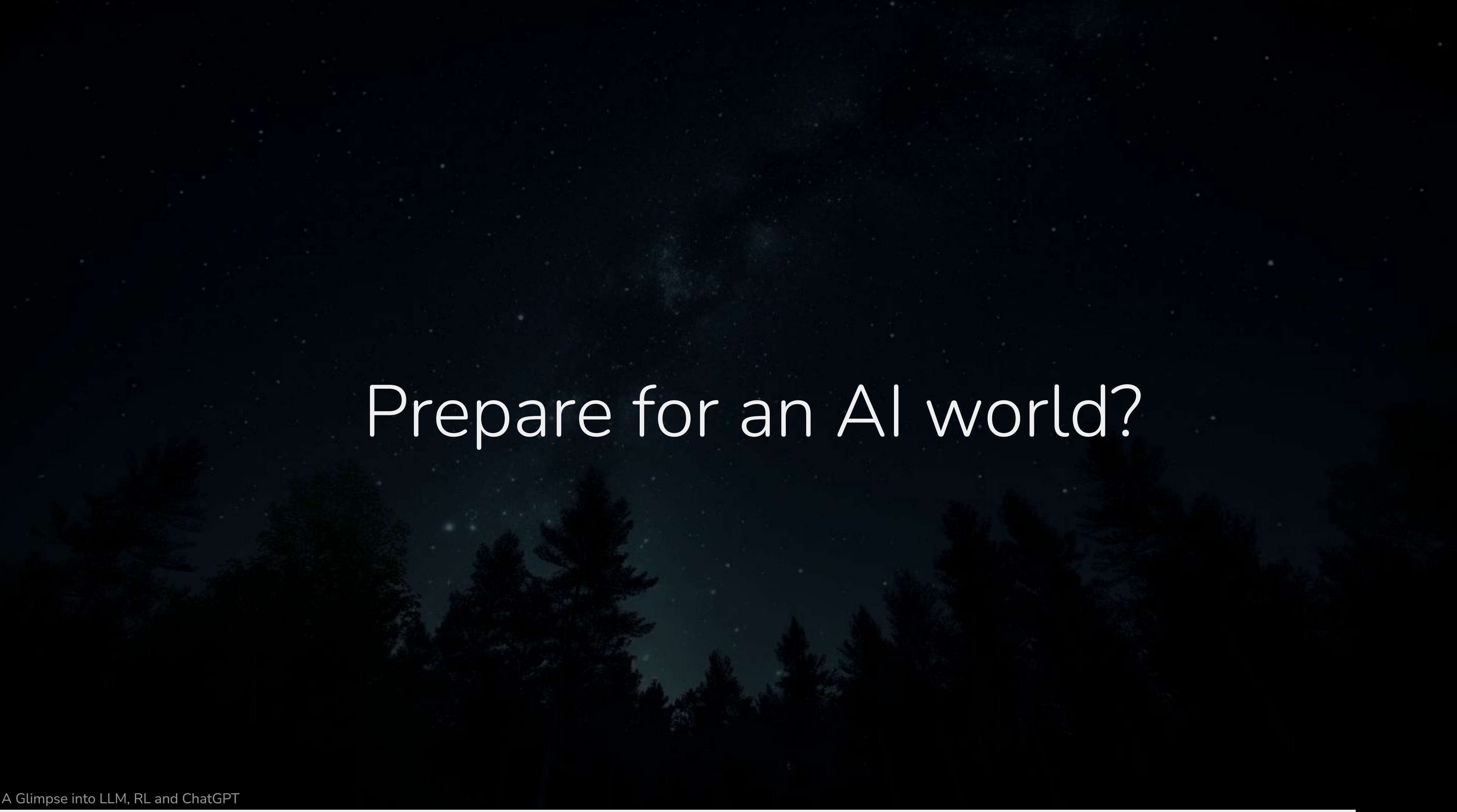
AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research[1] and acknowledged by top AI labs.[2] As stated in the widely-endorsed Asilomar AI Principles, *Advanced AI could represent a profound change in the history of life on Earth,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

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

Why the 6-month AI Pause is a Bad Idea

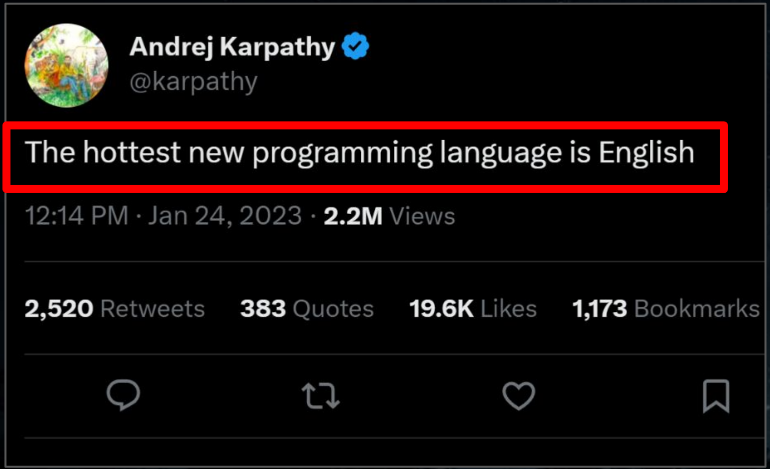
Friday, April 7, 2023

9:30 AM to 10:00 AM Pacific Time



Prepare for an AI world?

Maybe switch jobs to the hottest “Prompt Engineer”?

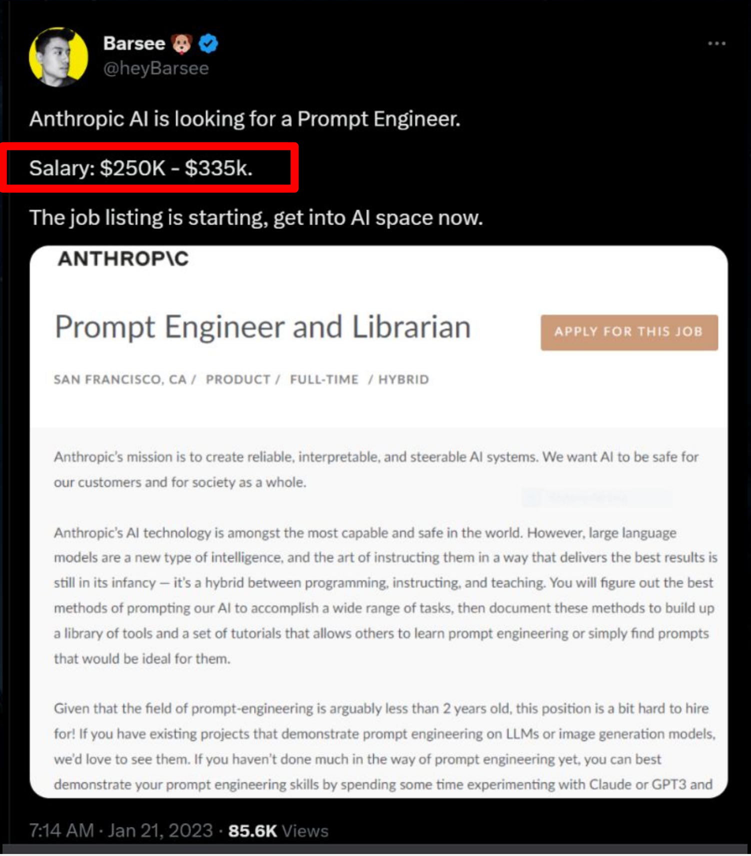


Andrej Karpathy
@karpathy

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



Barsee 🐶
@heyBarsee

Anthropic AI is looking for a Prompt Engineer.

Salary: $250K - $335k.

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

ANTHROP\C

Prompt Engineer and Librarian

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