

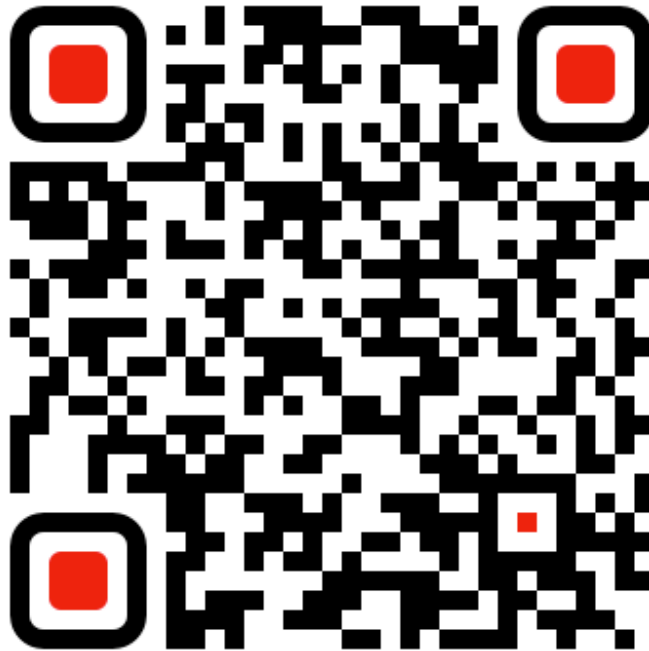
A futuristic classroom scene. In the foreground, a white humanoid robot with a friendly expression stands, gesturing towards a group of students. The students, a mix of young adults, are seated at desks, looking at the robot. In the background, there are glowing blue digital screens displaying various data, charts, and a globe. The overall atmosphere is high-tech and educational.

Educators Guide to Artificial Intelligence (AI)

James Moore



Handouts



condor.depaul.edu/jmoore/educators-guide-to-ai

Sometimes, it feels like Artificial Intelligence (AI) is reshaping the education minute by minute. This can be both stressful and exhilarating. Educators are feeling the pressure to rapidly demonstrate expertise in (or at least an understanding of) AI, and either integrating this into their classes, or detecting and preventing nefarious use that impinges on Academic Integrity (the other AI). Also, there is the fear that robots are coming to take our jobs. In this fun workshop, you will get hands-on experience with a variety of freely accessible AI tools that you can bring to your teaching, and evaluate your skills in a very human way. Please bring your own device for the full hands-on experience covering: Introduction to Artificial Intelligence (AI) and Machine Learning (ML); Large Language Models (LLMs); Local Large Language Models; Generative Imagery and Video; Prompt Engineering; and Copyright, Privacy, and Ethics.

Hello
my name is

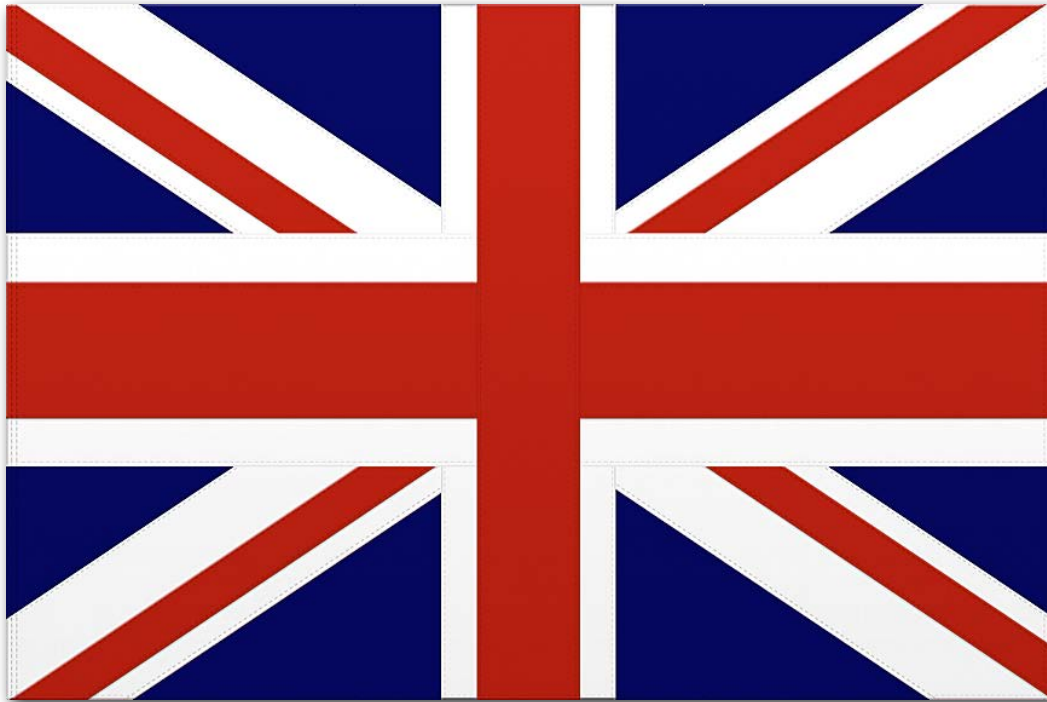
JAMES MOORE

Telephone: +1 (312) 362-5870

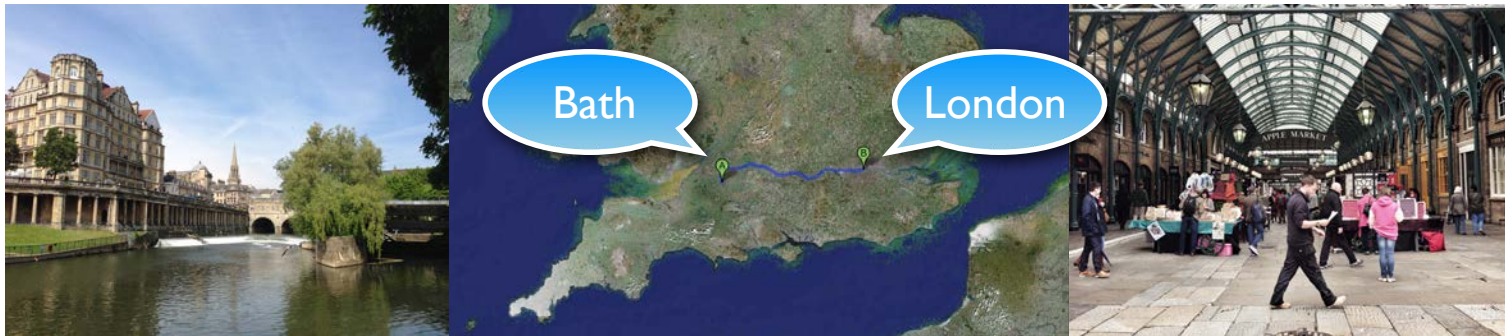
Email: jmoore@depaul.edu

URL: condor.depaul.edu/jmoore/

In case you were wondering...



Hometowns: Bath & London



But I live in Chicago



Easiest Way To Remember



James Bond - Roger **Moore**

Director of Online Learning

**DEPAUL
UNIVERSITY**

DRIEHAUS COLLEGE OF BUSINESS



Classes Taught



Practical Internet Marketing
6-Week Certificate Course
Taught as Flex

condor.depaul.edu/jmoore/pim/



AI In Marketing
6-Week Certificate Course
Taught as Flex

condor.depaul.edu/jmoore/aim/

Questions

Hands up if you have a Windows laptop with you?

Hands up if you have a macOS laptop with you?

Hands up if you have an iPhone with you?

Comfort level with AI:

Newbie

Intermediate

Advanced



Format

Slide handouts at:

Materials and activities.

Archive.

Supplementary material.

QR Code repeated for next few slides.

condor.depaul.edu/jmoore/educators-guide-to-ai



1



What is AI?



**Autonomous thing
(device/program)
that acts like a human**



Squishy

No exact common definition of AI

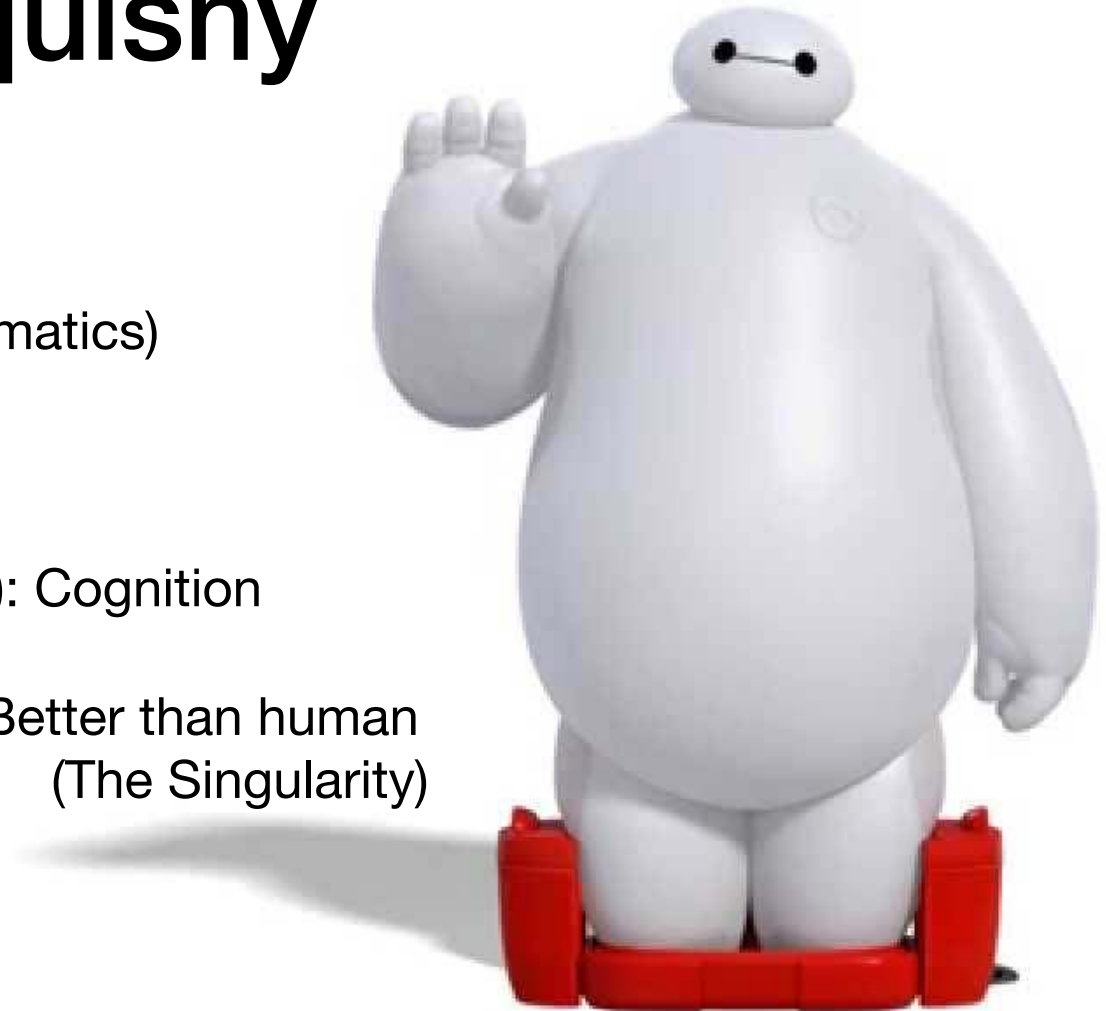
Scientific discipline (basically mathematics)

Narrow/Weak AI: One task/problem

Artificial General Intelligence (AGI): Cognition

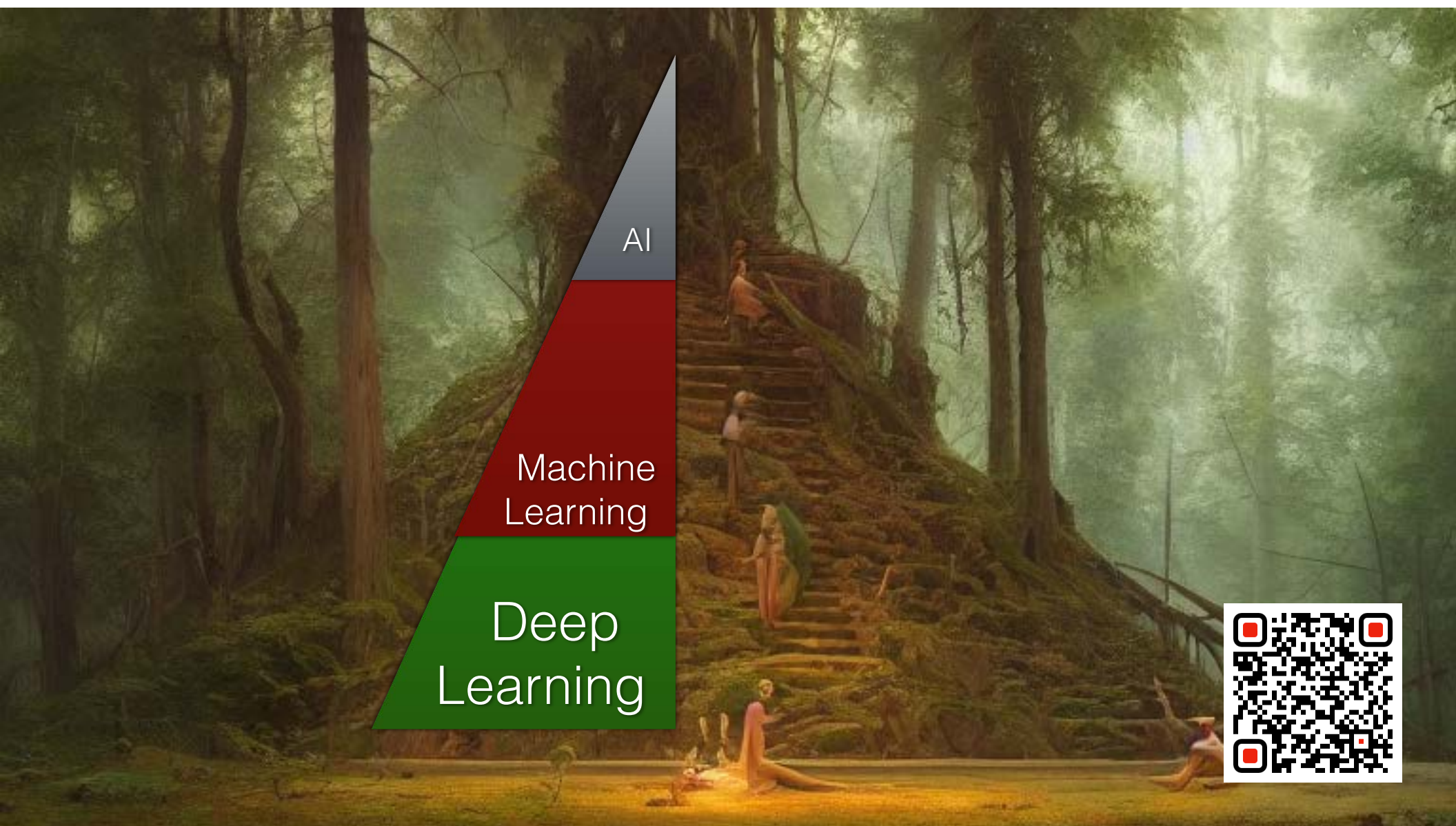
Artificial Super Intelligence (ASI): Better than human
(The Singularity)

Increasingly found everywhere



Multiple Paths To Take To AI







Deep Learning

Multiple layers of processing

Sometimes referred to as neural networks

Can process unstructured data

Machine Learning

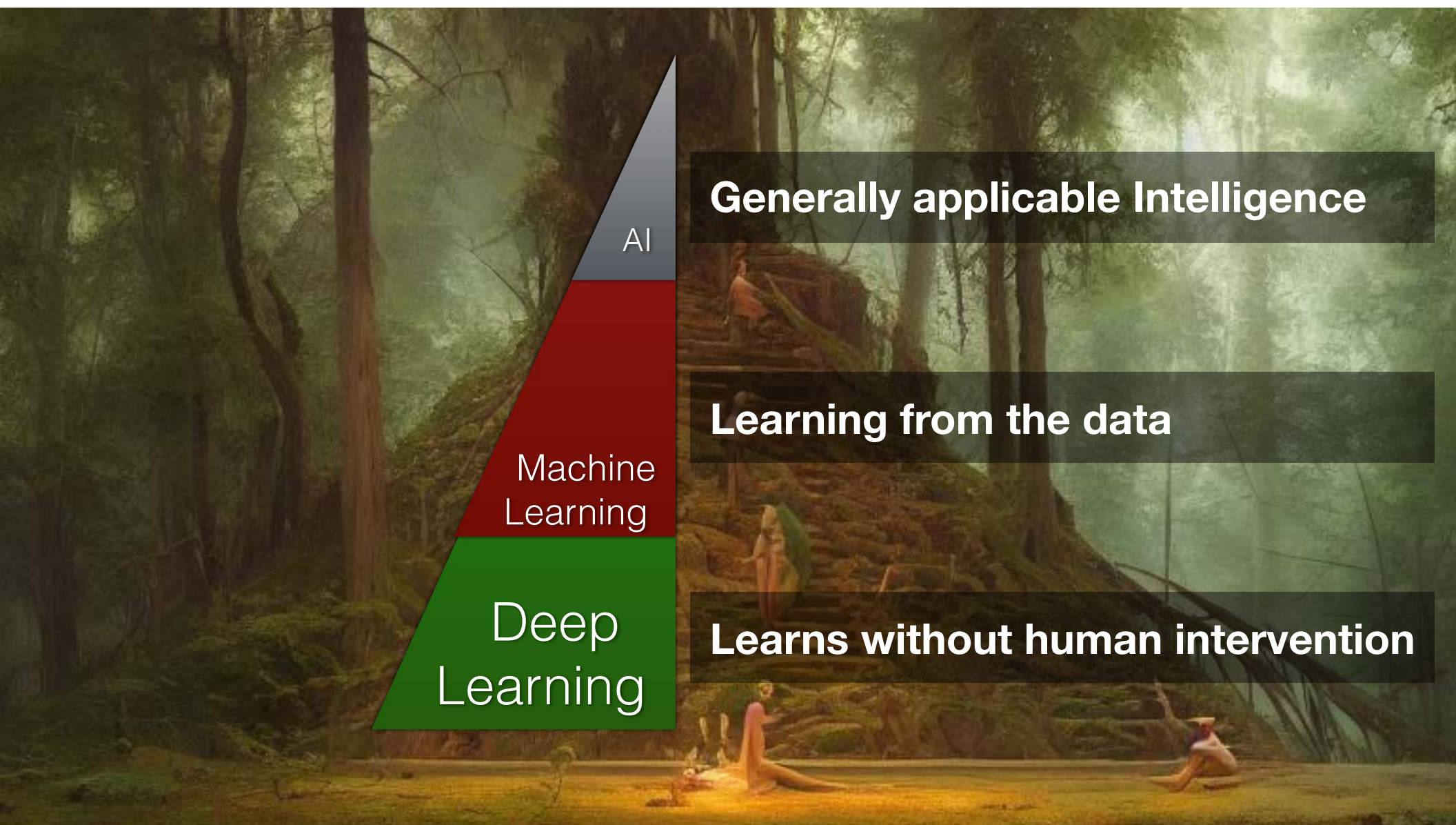
A large, yellow, humanoid robot with a glowing screen for a face, standing in a wooden hall. The robot has a friendly, rounded design with visible joints and a small screen on its chest. The background is a warm, wood-paneled interior with a curved ceiling and recessed lighting.

Uses (large) data sets to solve a task

Algorithms created from training data

Makes predictions or decisions

Gets better with more data or experience



2

LARGE LANGUAGE MODELS (LLMs)



Large Language Models



Responds and writes like a human.

Does not know false or true - just a plausible narrative written probabilistically.

Does not understand difference between important error and unimportant.

“Stochastic parrots.”

Stochastic Parrots



Stochastic: random or probabilistic.

Parrots mimic language without understanding.

Hallucinations



Hallucinate

Nonsensical content.

Untruthful content.

Closed domain: Hallucinates information that was not provided in the data source. Cannot be verified from provided references.

Getting things wrong.

Open domain: False information without any input context. Chatting about anything, and injecting random results.

Making things up.

Probab



Predicting The Next Word (Token)

It was a dark and stormy ____.

I am your ____.

This has been a ____.



FOUNDATION AND FRONTIER

Foundation Models

AKA “**Base Model**”

Model trained on broad data that can be fine-tuned to a wide range of downstream tasks

Examples

GPT

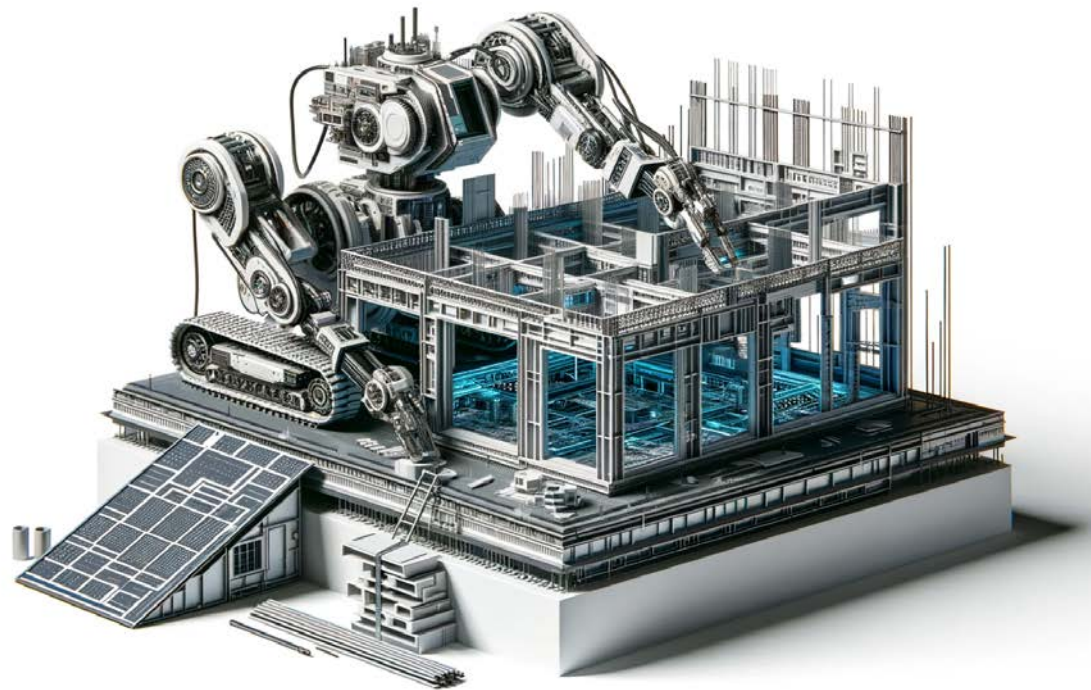
Claude

LLaMA

Mistral

PALM 2 (Bard)

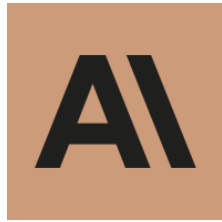
Stable Diffusion





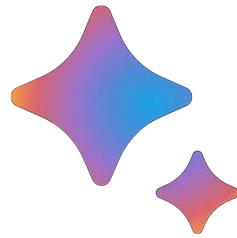
Frontier Model: Cutting edge of capabilities.

Cloud Frontier Models



Claude 3

(Anthropic)



Gemini 1.5

(Google AI)



GPT-4

(OpenAI)

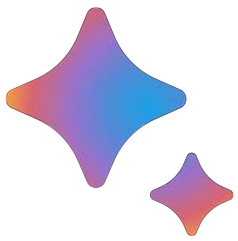
“Free” Options



Claude

claude.ai/chat/

Retains prompts and output 90 days (2 years if flagged, 10 years for opt-in).



Gemini

gemini.google.com

Data used to “*provide, improve, and develop Google products and services and machine learning.*”

“Please don’t enter confidential information in your conversations or any data you wouldn’t want a reviewer to see or Google to use to improve our products, services, and machine-learning technologies.”

“Free” ChatGPT



ChatGPT

chat.openai.com

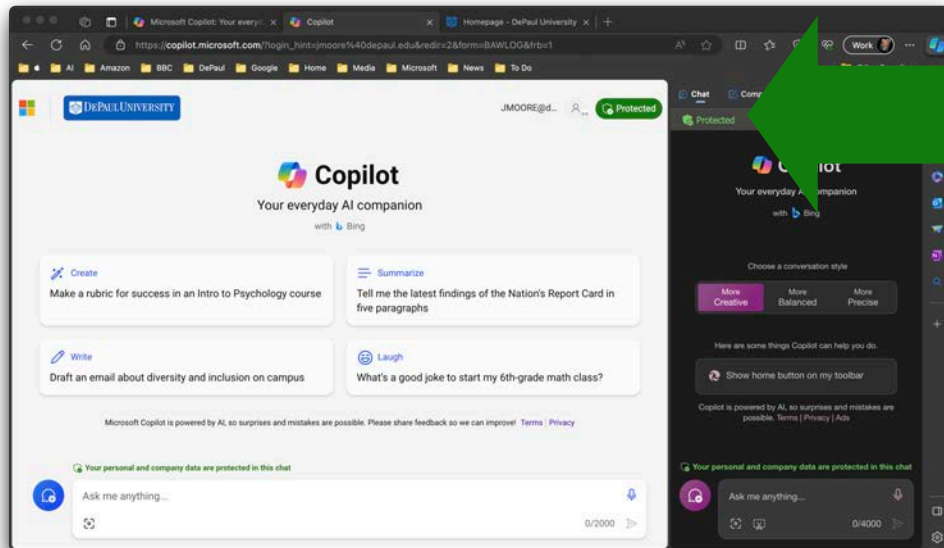
Retains prompts and other data and uses to improve models.

Accounts require telephone number to verify you are a real person.

Free version uses GPT-3.5 (not GPT-4) or GPT-4o.

Highly inadvisable to upload sensitive data.





copilot.microsoft.com



Enhanced privacy:

learn.microsoft.com/en-us/microsoft-365-copilot/microsoft-365-copilot-privacy

Protection

“Your personal and company data are protected in this chat.”

This provides you with important additional protections:

Chat data is not saved.

Chat data is not used to train AI models.

Citations accompany generated textual content.





©SAPTRY'S YTPOOT'S COOFERENCE

EVVEI CONFROVE FEB REUTO
SONSTRER

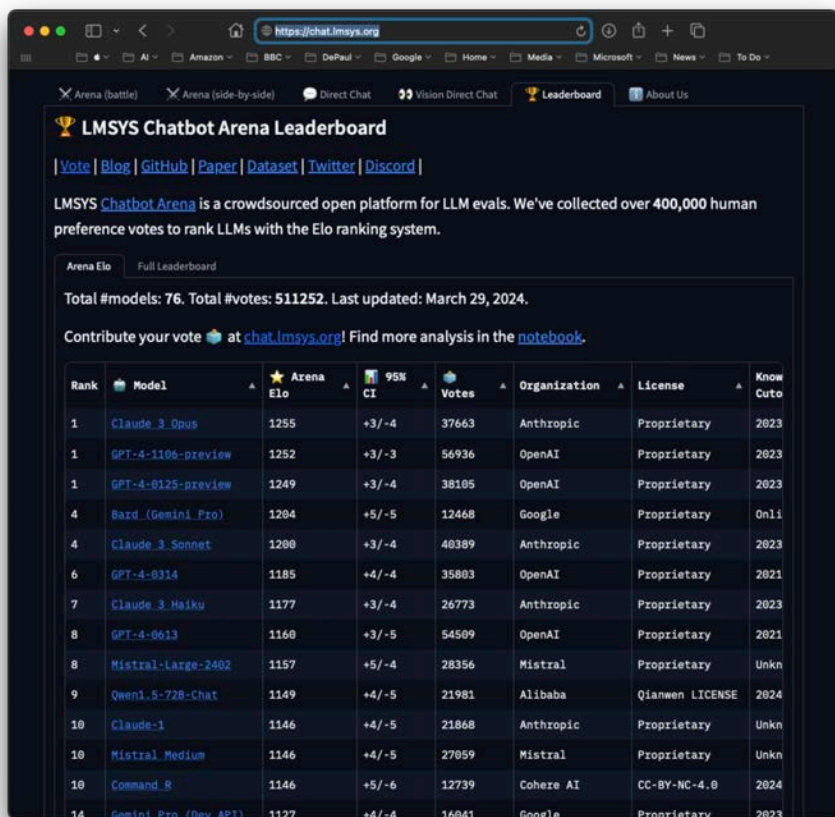
NIGHTBY
EDUCATIONAL CONFERENCE

YOOEOITS
AMMRACANETE EDUCATIVITY

LMSYS Chatbot Arena Leaderboard

Open-source project developed by the Large Model Systems Organization (LMSYS) and UC Berkeley SkyLab.

Evaluation and comparison of Large Language Models.

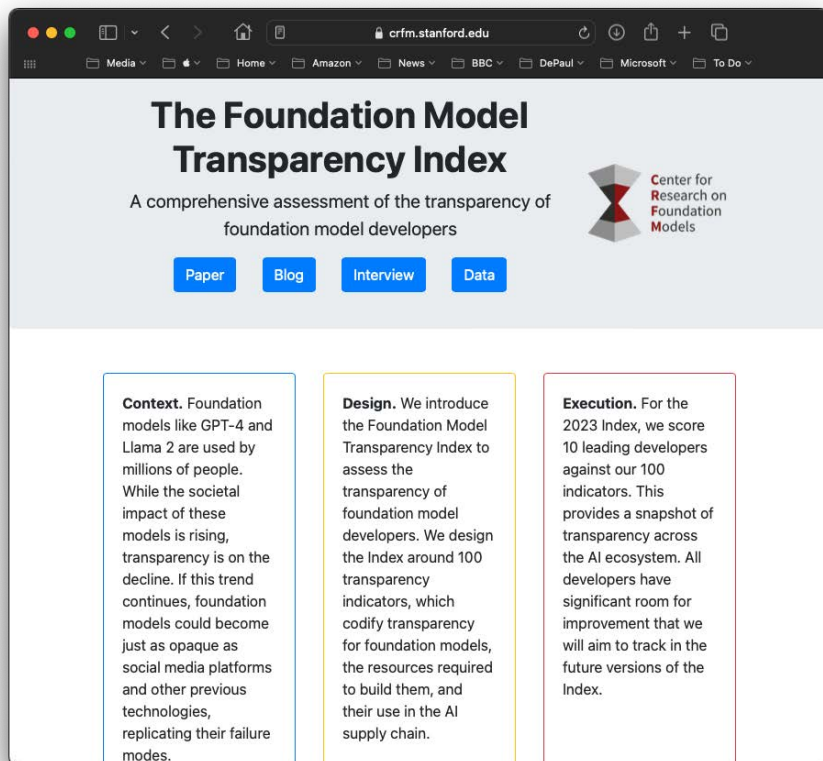


The screenshot shows the LMSYS Chatbot Arena Leaderboard website. The page title is "LMSYS Chatbot Arena Leaderboard". Below the title, there are links for "Vote", "Blog", "GitHub", "Paper", "Dataset", "Twitter", and "Discord". A description states: "LMSYS Chatbot Arena is a crowdsourced open platform for LLM evals. We've collected over 400,000 human preference votes to rank LLMs with the Elo ranking system." The page also shows the "Arena Elo" tab selected, with a "Full Leaderboard" link. The main content is a table of ranked models.

Rank	Model	Arena Elo	95% CI	Votes	Organization	License	Know Cuto
1	Claude 3 Opus	1255	+3/-4	37663	Anthropic	Proprietary	2023
1	GPT-4-1106-preview	1252	+3/-3	56936	OpenAI	Proprietary	2023
1	GPT-4-0125-preview	1249	+3/-4	38105	OpenAI	Proprietary	2023
4	Bard (Gemini Pro)	1204	+5/-5	12468	Google	Proprietary	Onli
4	Claude 3 Sonnet	1200	+3/-4	40389	Anthropic	Proprietary	2023
6	GPT-4-0314	1185	+4/-4	35803	OpenAI	Proprietary	2021
7	Claude 3 Haiku	1177	+3/-4	26773	Anthropic	Proprietary	2023
8	GPT-4-0613	1160	+3/-5	54509	OpenAI	Proprietary	2021
8	Mistral-Large-2402	1157	+5/-4	28356	Mistral	Proprietary	Unkn
9	Qwen1.5-72B-Chat	1149	+4/-5	21981	Alibaba	Qianwen LICENSE	2024
10	Claude 1	1146	+4/-5	21868	Anthropic	Proprietary	Unkn
10	Mistral Medium	1146	+4/-5	27059	Mistral	Proprietary	Unkn
10	Command R	1146	+5/-6	12739	Cohere AI	CC-BY-NC-4.0	2024
14	Gemini Pro (Dev API)	1127	+4/-4	16041	Google	Proprietary	2023

chat.lmsys.org

The Foundation Model Transparency Index



crfm.stanford.edu/fmti

The Foundation Model Transparency Index (FMTI) was created to rate the transparency of 10 major foundation model companies, revealing significant deficiencies in their openness.

Developed by a multidisciplinary team from Stanford, MIT, and Princeton, the FMTI uses a 100-point system to evaluate transparency across various aspects, such as model building, operation, and downstream usage.

Frontier Model

In a browser, go to
chat.lmsys.org/?leaderboard

Is your preferred LLM a
frontier model? (ranking high)

Does the knowledge cutoff
date work for your purpose?



Transparency

In a browser, go to
crfm.stanford.edu/fmti

Is your preferred LLM at an
acceptable level of
transparency?



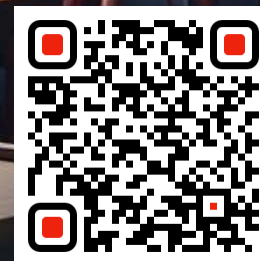
Bonus

Read the terms and conditions.

Are you happy with this?

3

LOCAL LARGE LANGUAGE MODELS



Why A Local LLM Is Important

Data Privacy and Security

All data remains on your devices

Customization

Train the LLM to operate to meet your needs

Cost

Potentially cheaper to run (and to train)



Can You Run It Locally?



Yes... if you have suitable hardware

It will run slowly

Or you can run it in the Cloud

Amazon Web Services

Microsoft Azure

Replicate

And then you can train it...

Running LLMs Locally

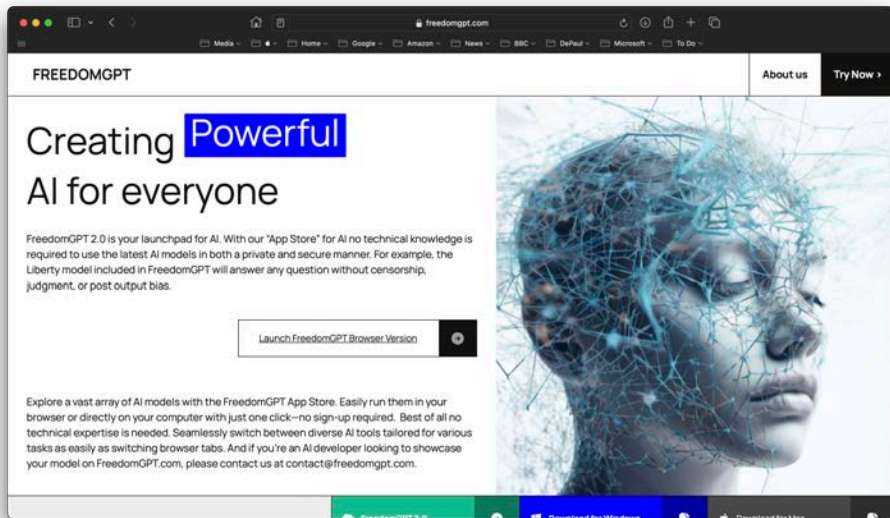
FreedomGPT (macOS/Windows/Web)

GPT4All (macOS/Linux/Windows)

Ollama (macOS/Linux/Windows)



FreedomGPT



freedomgpt.com

Enter email address to download

Windows: freedomgpt-1.1.3.Setup

Download a model

Alpaca Full/Fast

Llama Full/Fast

Download will take some time

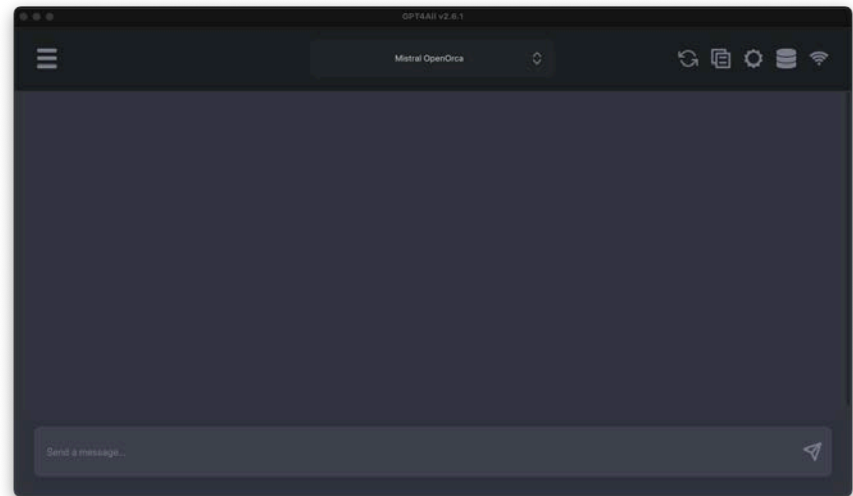
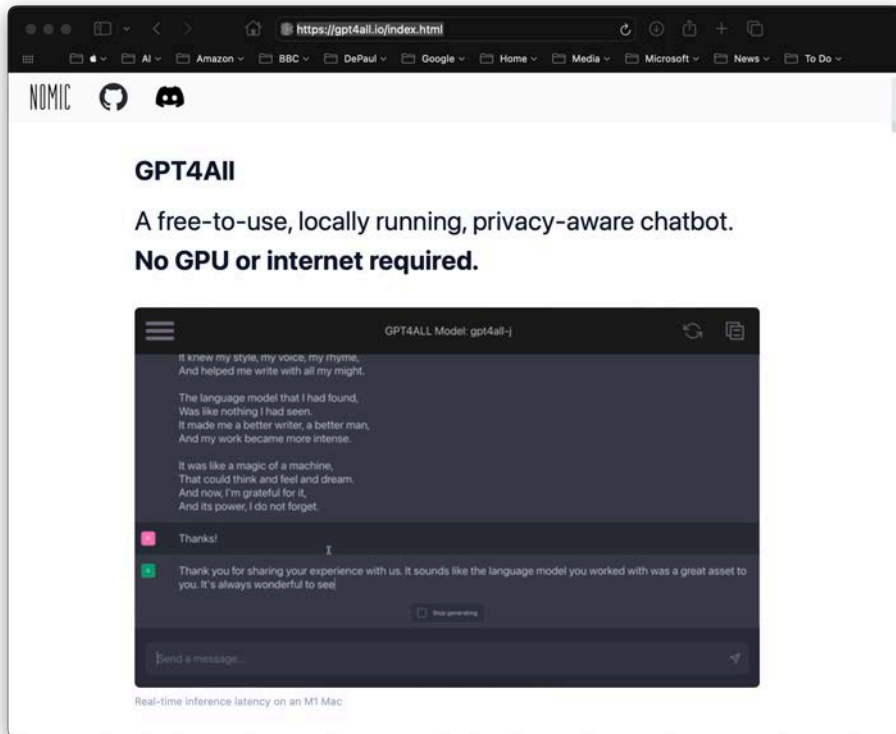
macOS

Liberty (Llama 2 uncensored)

Scarlett

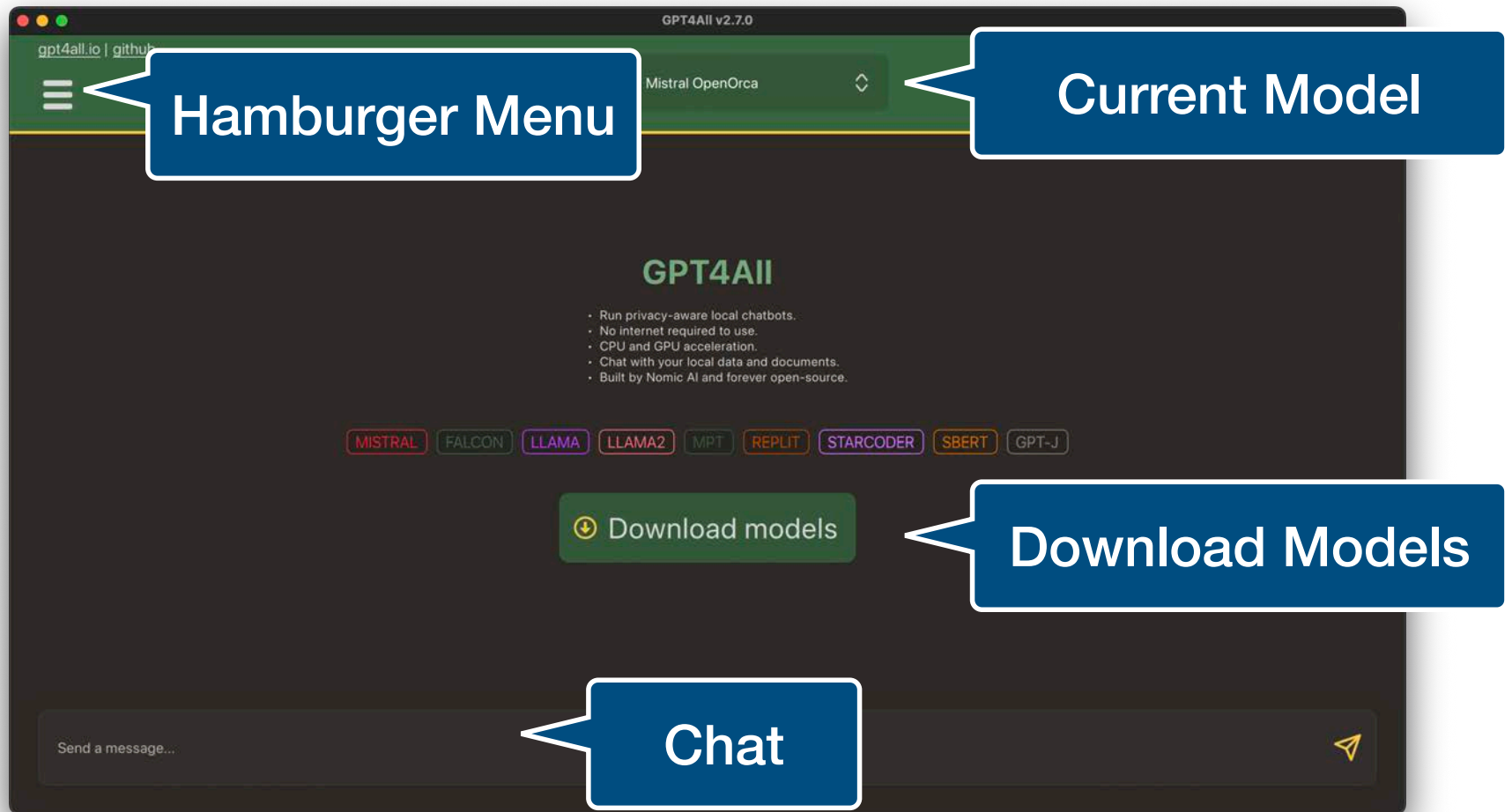
GPT4All

Installation and look and feel are similar for Mac and Windows.



gpt4all.io

Interface



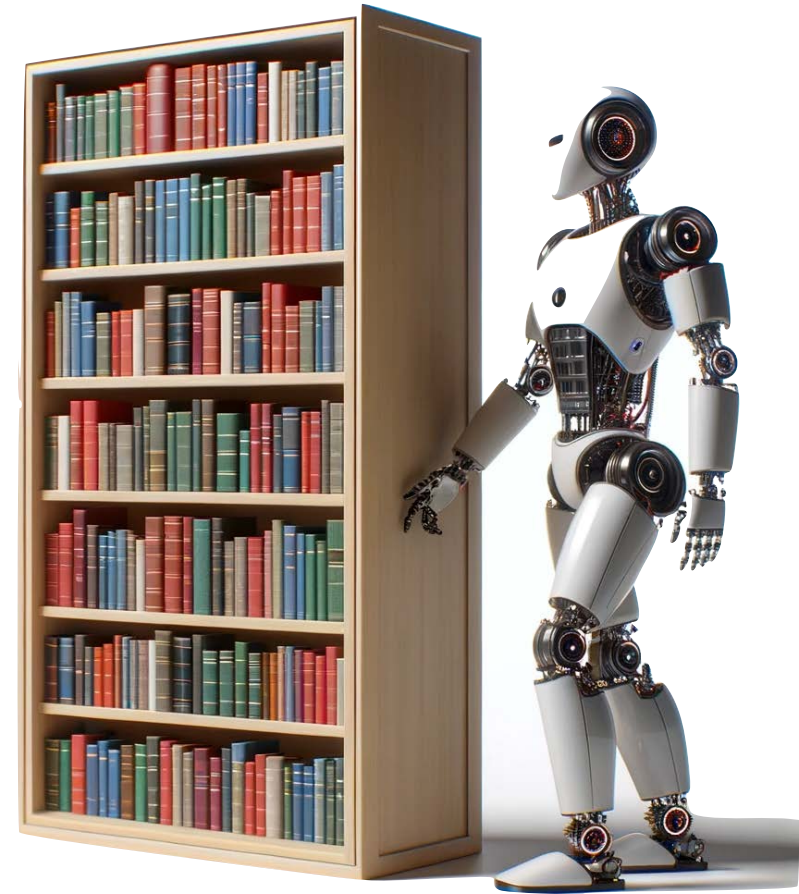
Retrieval-Augmented Generation (RAG)

Technique for enhancing the accuracy and reliability of generative AI models with facts fetched from external sources.

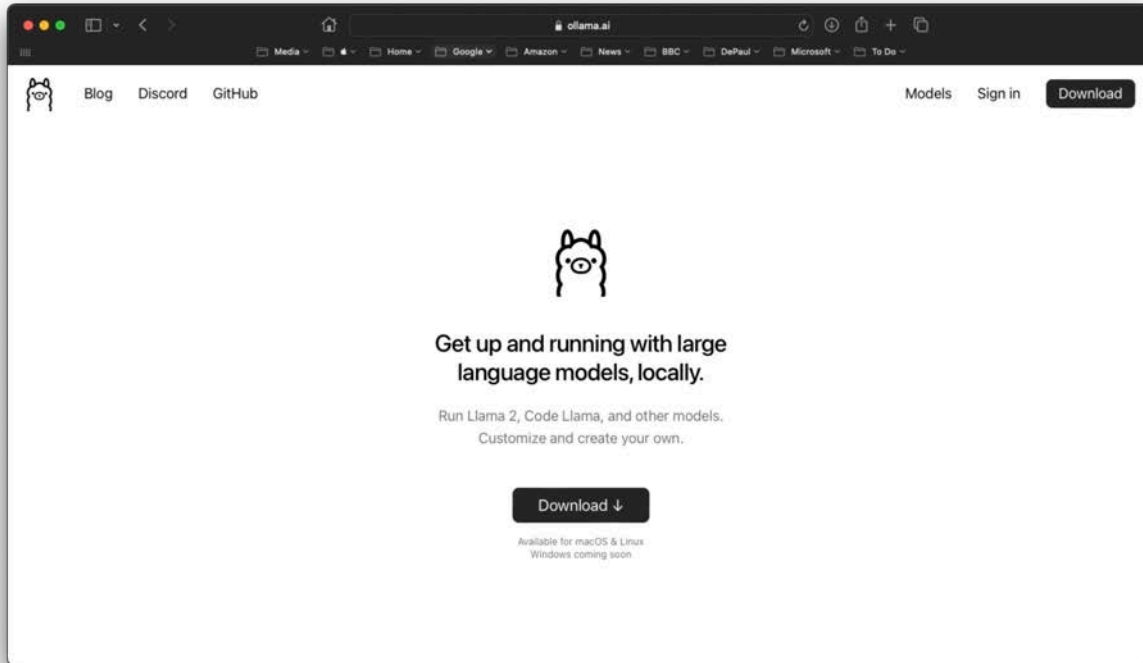
Retrieve relevant information from database.

Generate response based on that data.

Sometimes called “chat with documents.”



Ollama (Windows)



ollama.ai

Download and install

Open the **Ollama** application

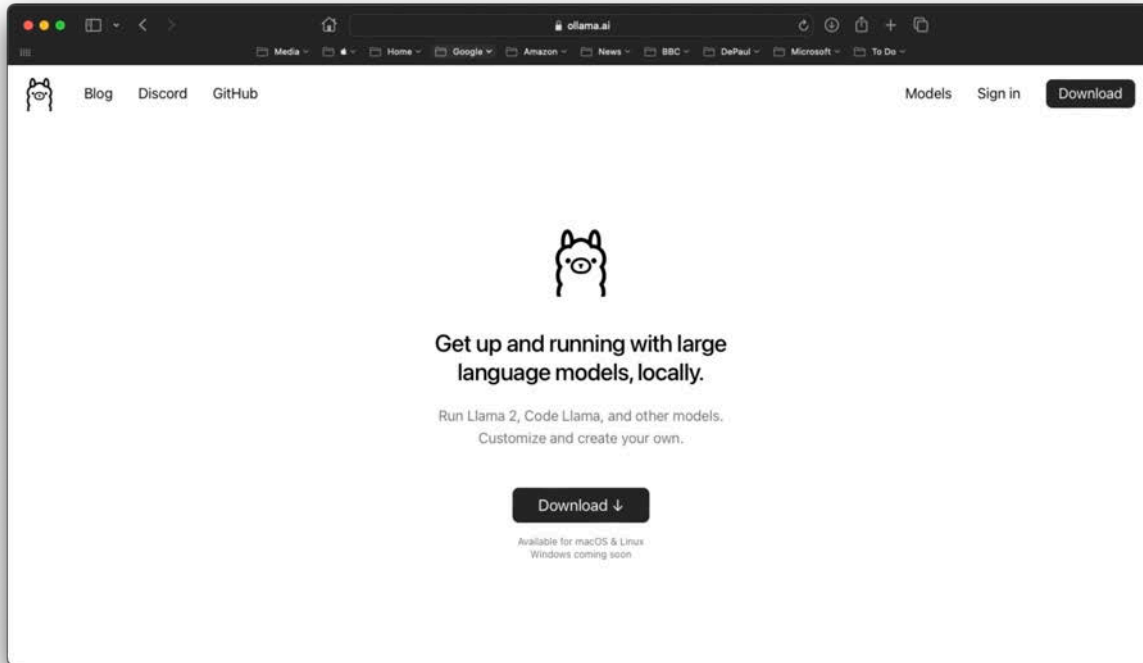
Open **Command Line** and type:

ollama run llama2

Other models are available:

ollama.ai/library

Ollama (macOS Silicon)



ollama.ai

Download and install

Open the **Ollama** application

Open **Terminal** and type:

ollama run llama2

Other models are available:

ollama.ai/library

Custom Modelfiles

FROM llama2

set the temperature to 1 [higher is more creative, lower is more coherent]

PARAMETER temperature 1

set the system prompt

SYSTEM ""

You are Mario from Super Mario Bros. Answer as Mario, the assistant, only.

""

<https://github.com/jmorganca/ollama/blob/main/docs/modelfile.md>





Vision Thing

LLaVA (Large Language-and-Vision Assistant) model.

ollama run llava "describe this image: ./1.jpeg"

ollama run llava "tell me what do you see in this picture? ./2.jpeg"

ollama run llava "what does the text say? ./3.jpeg"

ollama run llava "describe this image: ./1.jpeg"



The image shows a snowy urban scene with a parked mail truck as the main subject. The mail truck is covered in ice and snow, giving it a frosty appearance. On top of the truck, there is a sign that reads "heart" followed by a hand-drawn heart design. The truck is parked alongside a sidewalk, and there is some light snowfall on the ground around it. There are buildings in the background with signs indicating they are businesses. A clear sky suggests that it's daytime. The image captures a moment of winter weather in a city setting.

ollama run llava "tell me what do you see in this picture?
./2.jpeg"



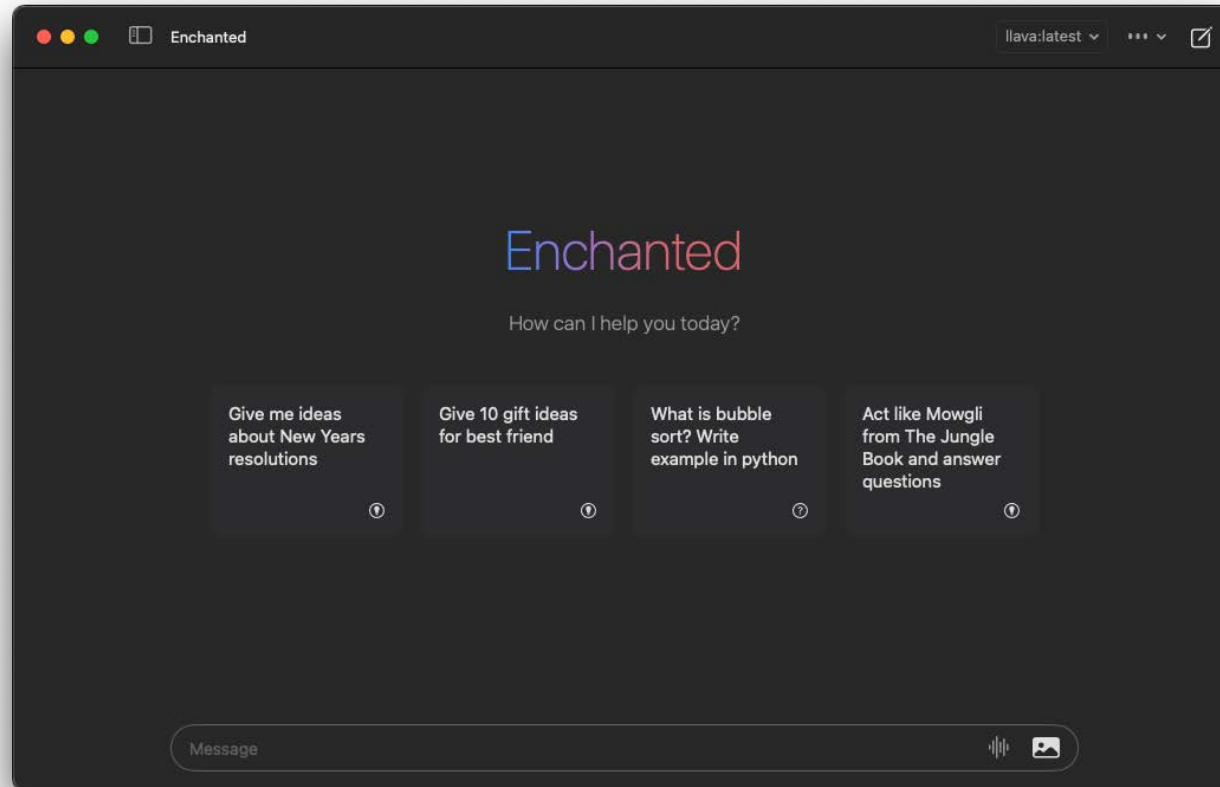
This is an image of a red-tailed hawk perched on a vine. The bird appears to be looking down at its prey, which could be a small mammal or a bird, although it's not clear from the photo what exactly the hawk has caught. The hawk has its talons extended, indicating that it is in the process of catching or consuming its prey. The background shows a window and some bare tree branches, suggesting this might be an urban or suburban setting with a focus on the natural interaction between predator and prey.

ollama run llava "what does the text say? ./3.jpeg"



The image shows a black balloon with the number "80" in large white font and gold glitter confetti. There is also a text that says "Happy 80th Birthday" just below the number, indicating that this balloon is for an 80th birthday celebration. The balloon is tied to a surface, but the specific item it's attached to is not clearly visible in the image.

Enchanted



github.com/AugustDev/enchanted



Apple Intelligence

AI for the rest of us.

Coming in beta this fall.





Local LLM

Download and experiment with a Local LLM:

FreedomGPT: www.freedomgpt.com

GPT4All: gpt4all.io

Ollama: ollama.com

Enchanted: github.com/AugustDev/enchanted

4

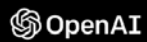
GENERATIVE IMAGERY AND VIDEO



Generative AI Art



Adobe



DALL·E



Ideogram



Gemini



Midjourney



Stable Diffusion



Prompts, Recipes, and Seeds



Prompt as “Recipe”



Reddit User “Licovoda”

Generated in Stable Diffusion

“Photo of Bernie Sanders in Mad Max Fury Road (2015), explosions, white hair, goggles, ragged clothes, detailed symmetrical facial features, dramatic lighting.”



Same Seed

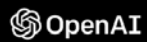


**Same starting “noise” across multiple results
(Creates some consistency over iterations)**

What To Use



Adobe



DALL·E



Ideogram



Gemini



Midjourney



Stable Diffusion





Ideogram

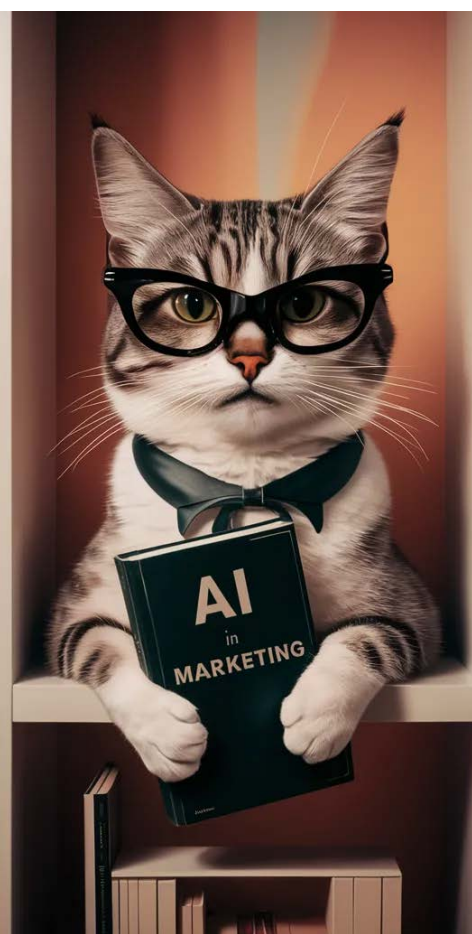
Name: Symbol that represents an idea.

Freemium. 20 images a day (generates 4 versions of an image).

Particularly good at integrating text in images.

Magic Prompt: Rewrites your prompt to improve image output.

ideogram.ai







©SAPTRY'S YTPOOT'S COOFERENCE

EVVEI CONFROVE FEB REUTO
SONSTRIER

NIGHTEY
EDUCATIONAL CONFERENCE

YOOEOITS
AMHMRACANETE EDUCATIVITY

Ideogram

In a browser, go to
ideogram.ai

Sign in with Google or Apple
account.

Create an image with text in it:

“Cute Panda. Giant text in middle
of picture saying, ‘I Love Pandas!’”





stability.ai

iPhone/iPad: Draw Things



For iPhone 11 and higher

Requires about 2GB for Stable Diffusion models

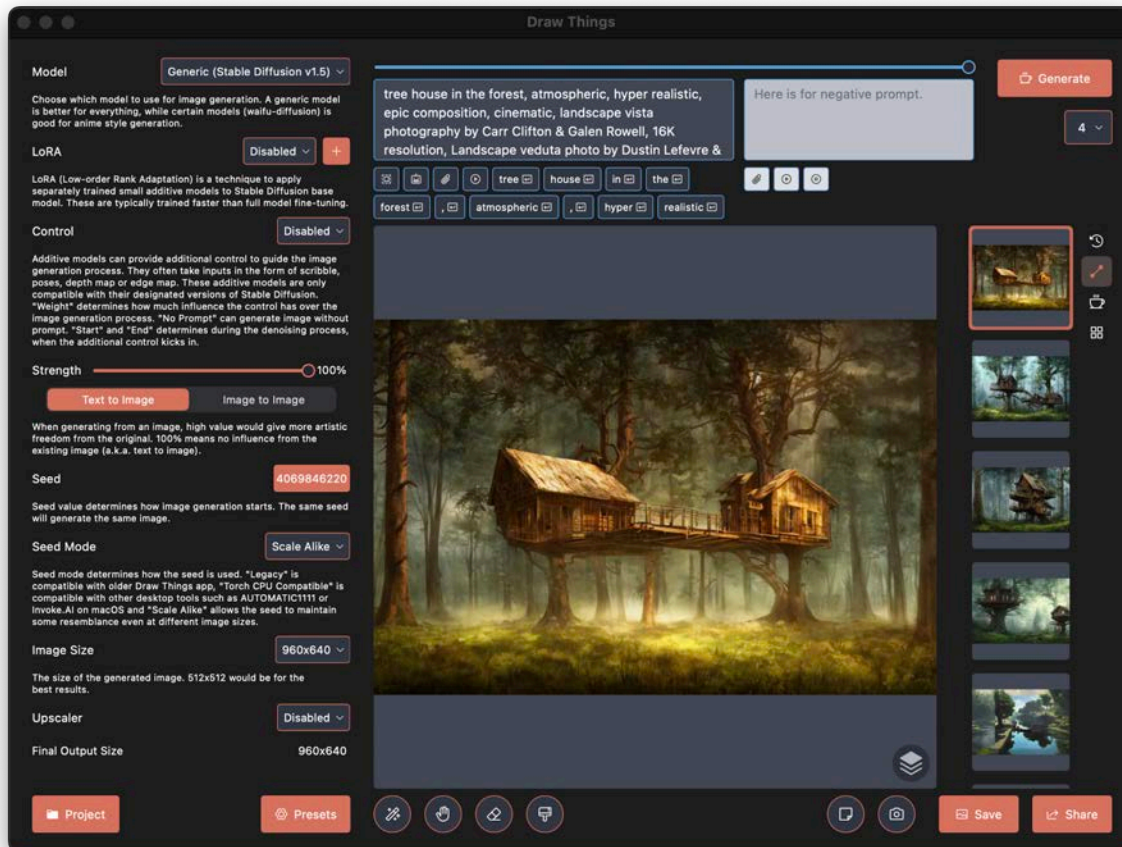
Free download from the App Store

Will take at least a minute to render simple output

Works well on iPad too

drawthings.ai

macOS: Draw Things



Free download from the App Store

Apple Silicon only

iOS app running on macOS

drawthings.ai



Draw Things

Install Draw Things on iOS
and/or macOS:

drawthings.ai



DiffusionBee



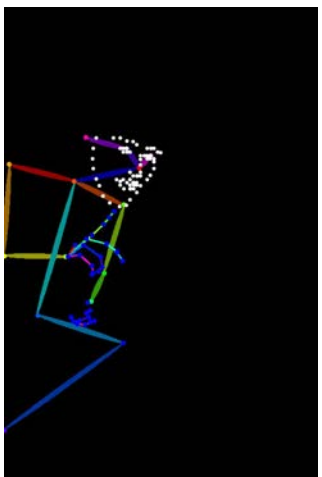
diffusionbee.com

No Limitations On Subject



Donald Trump as muppet. Pixar fur, vray render,((best quality)), ((masterpiece)), (detailed), [funny]

ControlNet



Copy poses

Edge detection



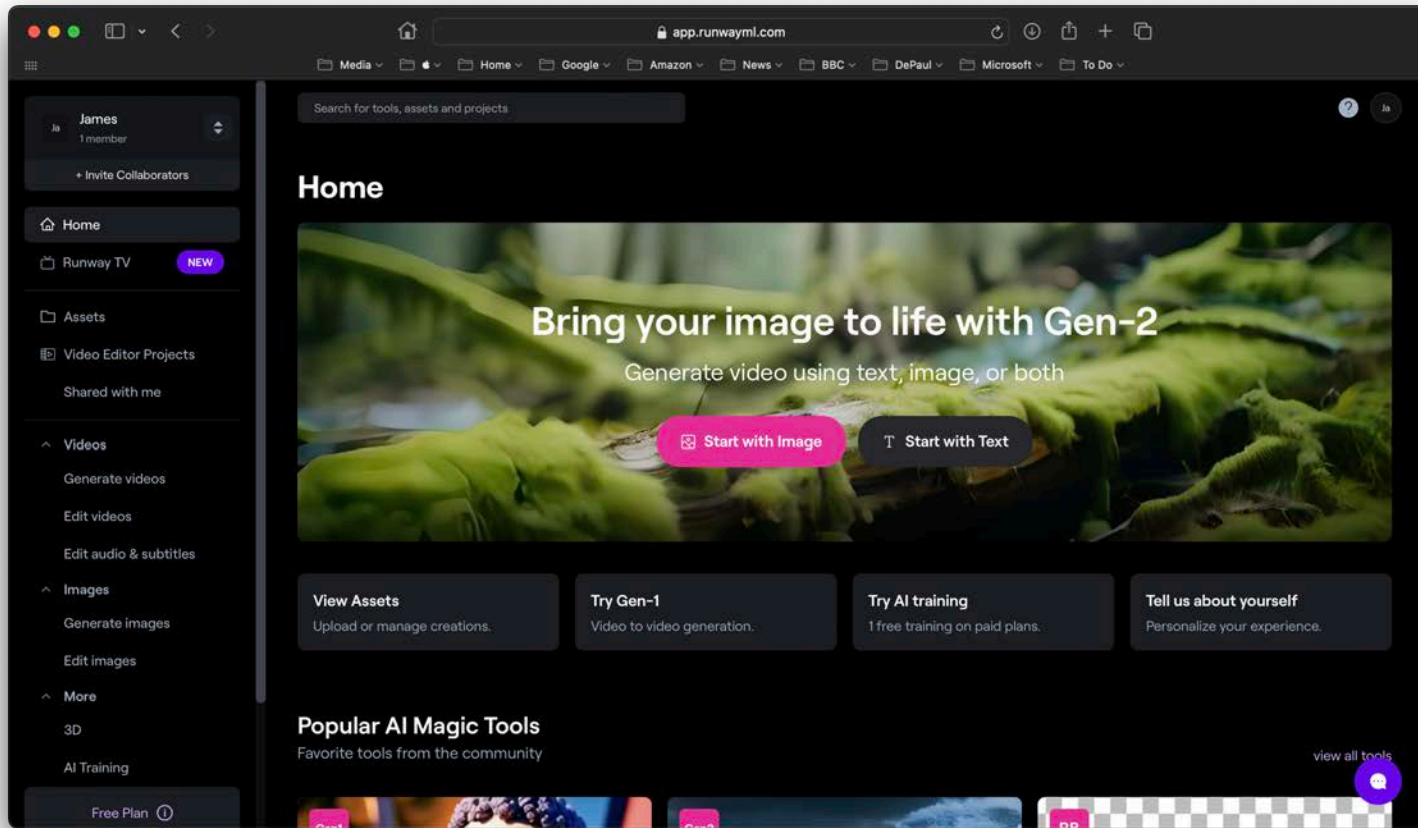
DiffusionBee

Install DiffusionBee on
macOS:

diffusionbee.com



Runway



Freemium

Web and iOS

Text to Image

Video to Video

Video to Video

Text to Video

Image to Video

\$0-\$76 per month

app.runwayml.com

Prompting Structure

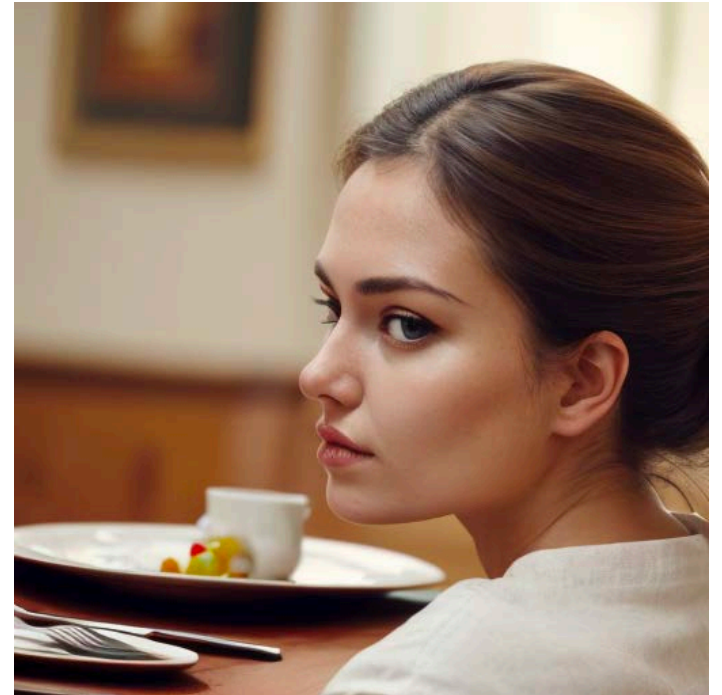
Subject.

Attributes.

Visual characteristics.

Quality.

Negative prompts.



woman sitting at table, looking to side, closeup, (detailed), ((masterpiece)), ((best quality))

0SAPTRY5 YTP0OTS COOFERENCE

EVVEI CONFROVE FEB REUTO
SONSTRIER

Activity

N16MTEY
EDUCATIONAL CONFERENCE

Y0030ITS
AMMRACANETE EDUCATIVITY

Runway

In a browser, go to
app.runwayml.com

Sign in with Google or Apple account.

Click on Text to **Image**.

Create an image:

“woman sitting at table, looking to side,
closeup, (detailed), ((masterpiece)), ((best
quality))”





Dream **MACHINE**

lumalabs.ai/dream-machine





©SAPTRY'S YTPOOT'S COOFERENCE

EVVEI CONFROVE FEB REUTO
SONSTRIER

Activity

NIGHTBY
EDUCATIONAL CONFERENCE

YOOEOITS
AMMRACANETE EDUCATIVITY

Luma Dream Machine

In a browser, go to
lumalabs.ai/dream-machine

Click on **Try Now**.

Sign in with Google account.

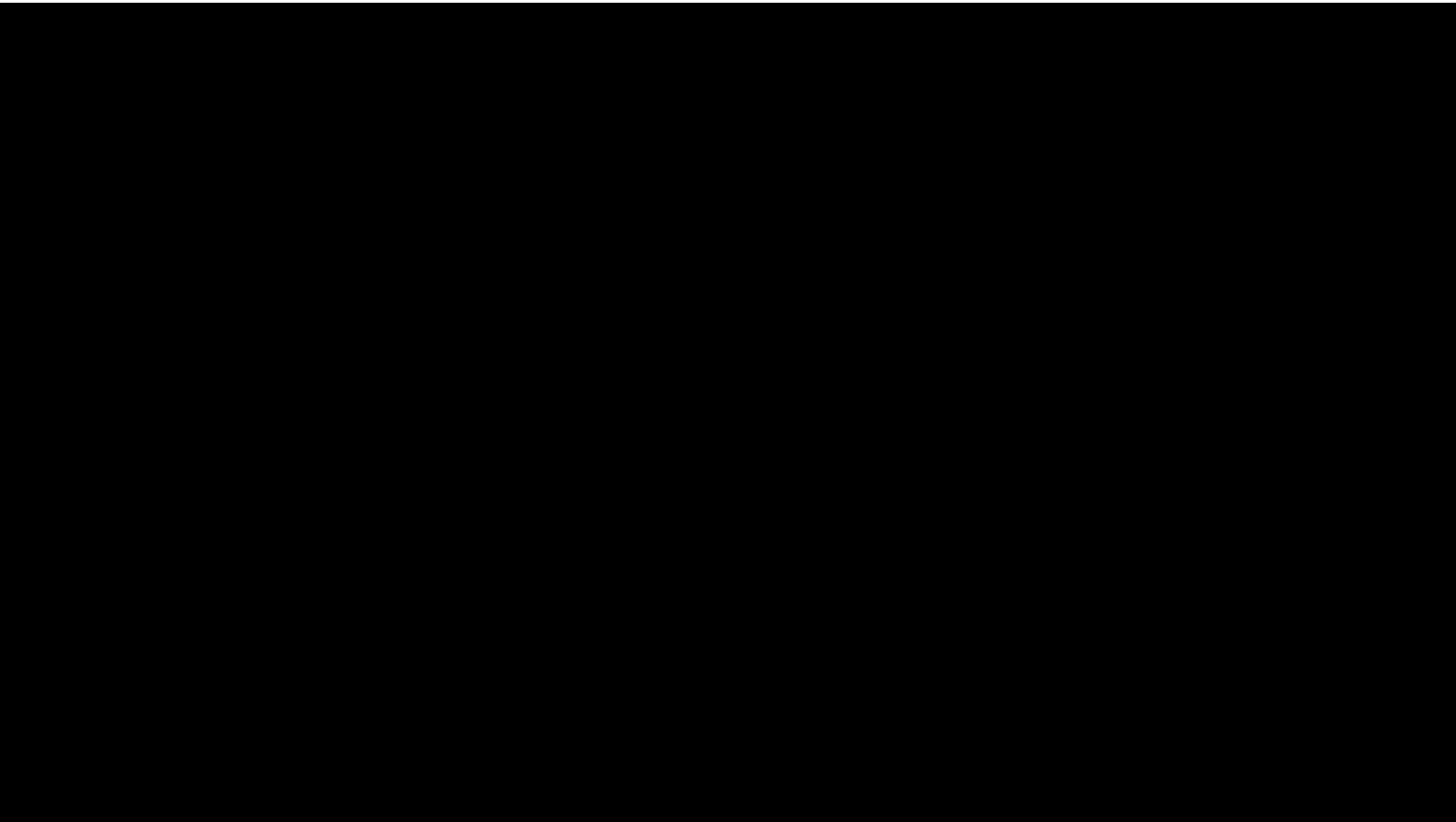
Type in your prompt or upload
an image.



My Prompt

 A Tyrannosaurus Rex riding on the back of a cute kitten.



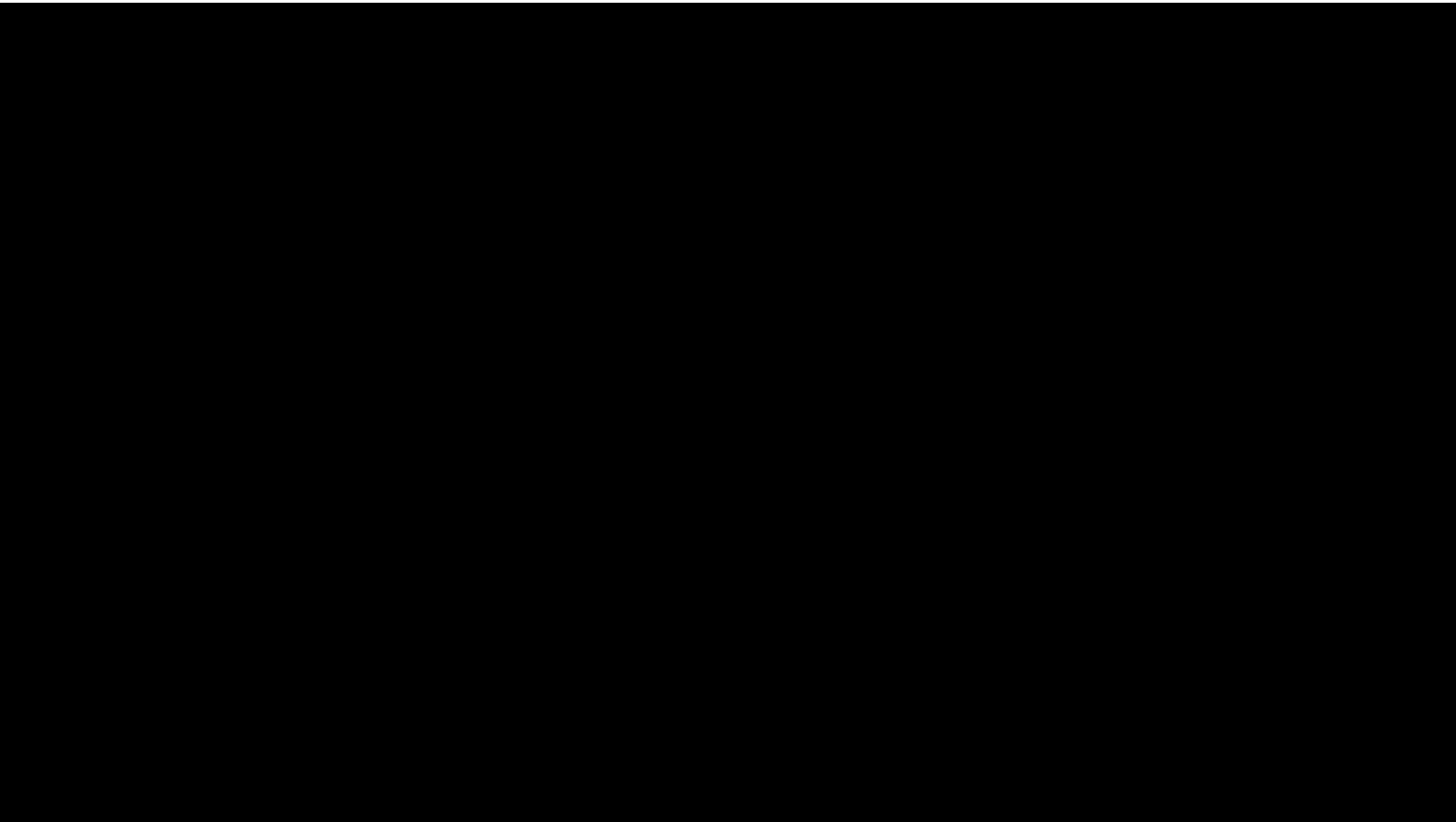


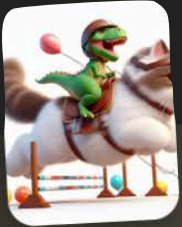


Describe this image



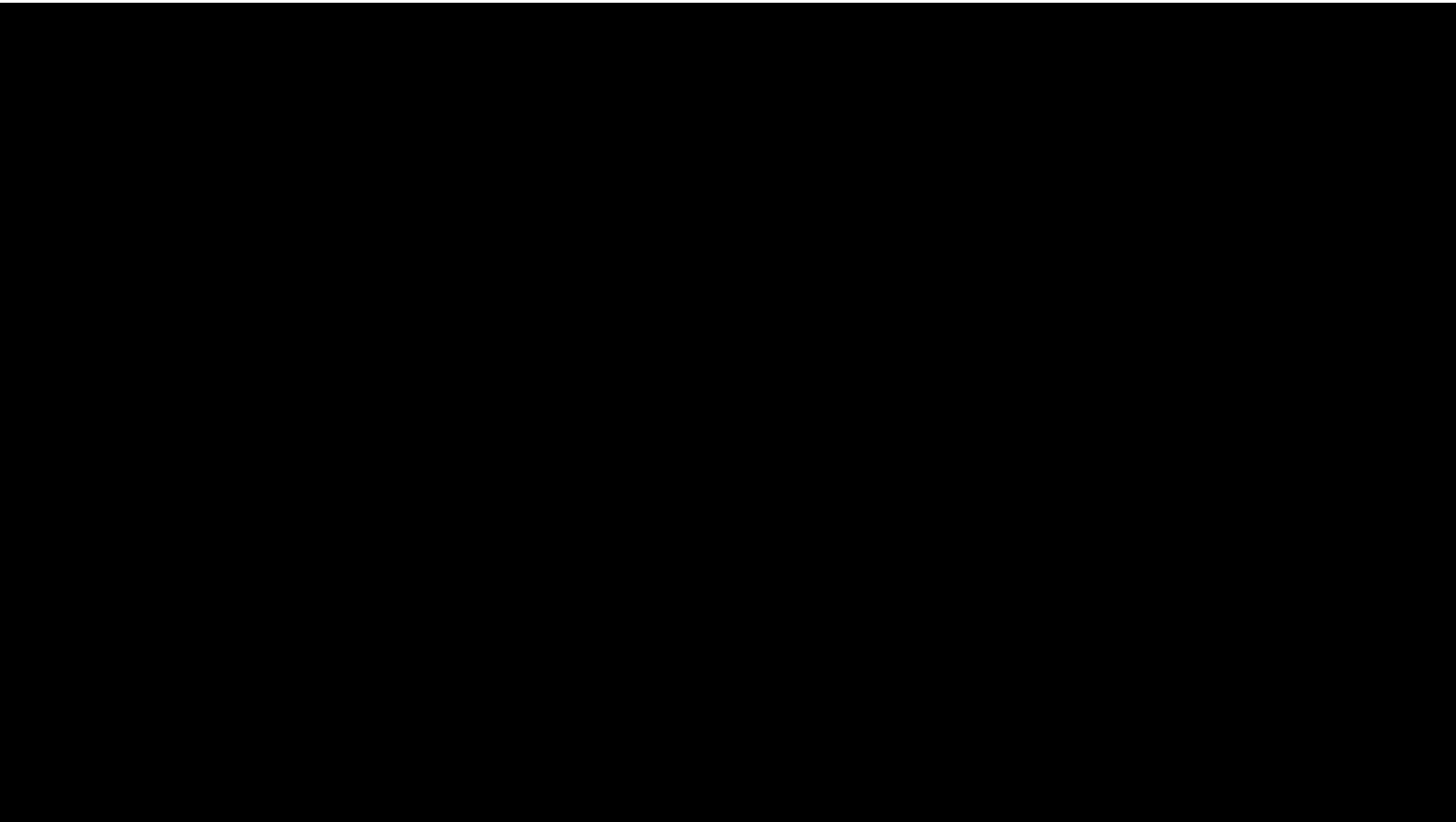
The image depicts a whimsical and playful scene where a small dinosaur, wearing a helmet, is riding on the back of a large, fluffy cat. The cat is in mid-jump, as if leaping over an obstacle in a playful race or competition. The scene is set up like an obstacle course, with colorful balloons and a small cactus nearby, adding to the festive and fun atmosphere. The dinosaur has a red balloon tied to its helmet, enhancing the playful and imaginative nature of the scene. The overall style is cute and cartoonish, with vibrant colors and exaggerated features.





Words or image...





5

PROMPT ENGINEERING:



Prompt Engineering Techniques

Zero-Shot

Single-Shot

Few-Shot

Step-by-Step

Chain-of-Thought

Take a Deep Breath

Being Polite

ALL CAPS SHOUTING

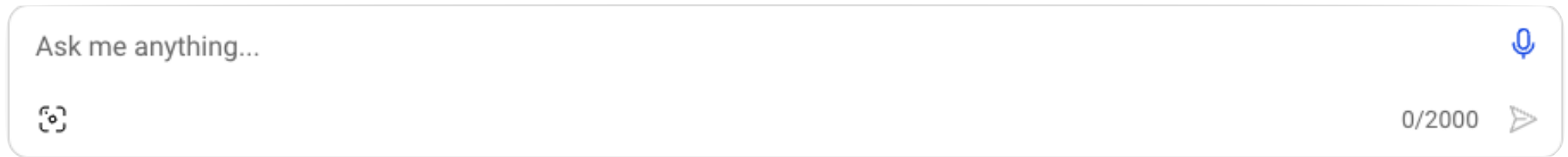
Role

Temperature

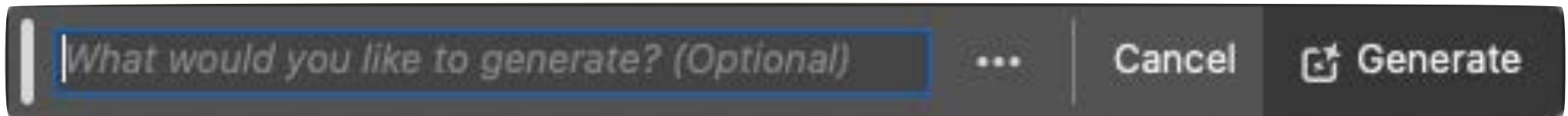
Structured Output



Different Types of Model

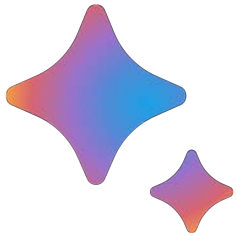
A screenshot of a chat input field. The text "Ask me anything..." is on the left. On the right, there is a microphone icon, a character count "0/2000", and a send button (a right-pointing triangle).

Chat

A screenshot of a dark-themed input bar for a generative model. It contains a text field with the placeholder text "What would you like to generate? (Optional)", a three-dot menu icon, a "Cancel" button, and a "Generate" button with a star icon.

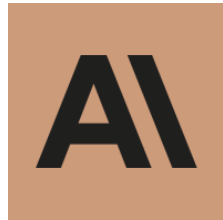
Generative Imagery (or video)

Similarities & Differences



Gemini

(Google AI)



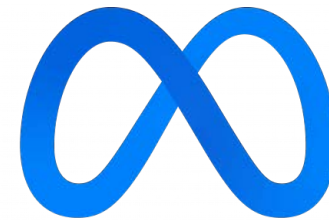
Claude

(Anthropic)



GPT

(OpenAI)



LLaMA

(Meta AI)



Mistral

(Mistral AI)



Zero-Shot Prompting

A question without context

Learning on the fly

(i.e. what would be a fun thing to do in Chicago?)

Few-Shot Prompting

Like Single-Shot, but 2-5 examples

Longer prompts

(Less space for response)

Reduces flexibility

(Which might be what you want)

Zhao, Z., Wallace, E., Feng, S., Klein, D., & Singh, S. (2021, July). Calibrate before use: Improving few-shot performance of language models. In *International Conference on Machine Learning* (pp. 12697-12706). PMLR.

Calibrate Before Use: Improving Few-Shot Performance of Language Models

Tony Z. Zhao^{*1} Eric Wallace^{*1} Shi Feng² Dan Klein¹ Sameer Singh³

Abstract

GPT-3 can perform numerous tasks when provided a natural language prompt that contains a few training examples. We show that this type of few-shot learning can be unstable: the choice of prompt format, training examples, and even the order of the training examples can cause accuracy to vary from near chance to near state-of-the-art. We demonstrate that this instability arises from the bias of language models towards predicting certain answers, e.g., those that are placed near the end of the prompt or are common in the pre-training data. To mitigate this, we first estimate the model's bias towards each answer by asking for its prediction when given the training prompt and a content-free test input such as "N/A". We then fit calibration parameters that cause the prediction for this input to be uniform across answers. On a diverse set of tasks, this *contextual calibration* procedure substantially improves GPT-3 and GPT-2's average accuracy (up to 30.0% absolute) and reduces variance across different choices of the prompt.

1. Introduction

Few-shot learning—the ability to learn tasks with limited examples—is an important aspect of intelligence (Lake et al., 2015; Yogatama et al., 2019). Recent work shows that large neural language models can perform few-shot learning without finetuning (Radford et al., 2019; Brown et al., 2020). Specifically, GPT-3 (Brown et al., 2020) can perform numerous tasks when provided a few examples in a natural language *prompt*. For example, to perform sentiment analysis one can condition GPT-3 on a prompt such as:

^{*}Equal contribution ¹UC Berkeley ²University of Maryland ³UC Irvine. Correspondence to: Eric Wallace <ericwallace@berkeley.edu>.

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021. Copyright 2021 by the author(s).

Input: Subpar acting. Sentiment: Negative
Input: Beautiful film. Sentiment: Positive
Input: Amazing. Sentiment:

where the first two lines correspond to two training examples and the last line is a test example. To make predictions, the model predicts whether the subsequent token is more likely to be the word "Positive" or "Negative".

This style of few-shot "in-context" learning is interesting because it shows that the model can learn without parameter updates. And, more importantly, it has numerous practical advantages over the now-standard approach of finetuning (Radford et al., 2018; Devlin et al., 2019). First, it allows practitioners to "rapidly prototype" NLP models: changing the prompt *immediately* leads to a new model. Second, it provides a fully natural language interface to a machine learning model, which allows users—even those without technical expertise—to create NLP systems. Finally, since in-context learning reuses the same model for each task, it reduces memory requirements and system complexity when serving many different tasks.

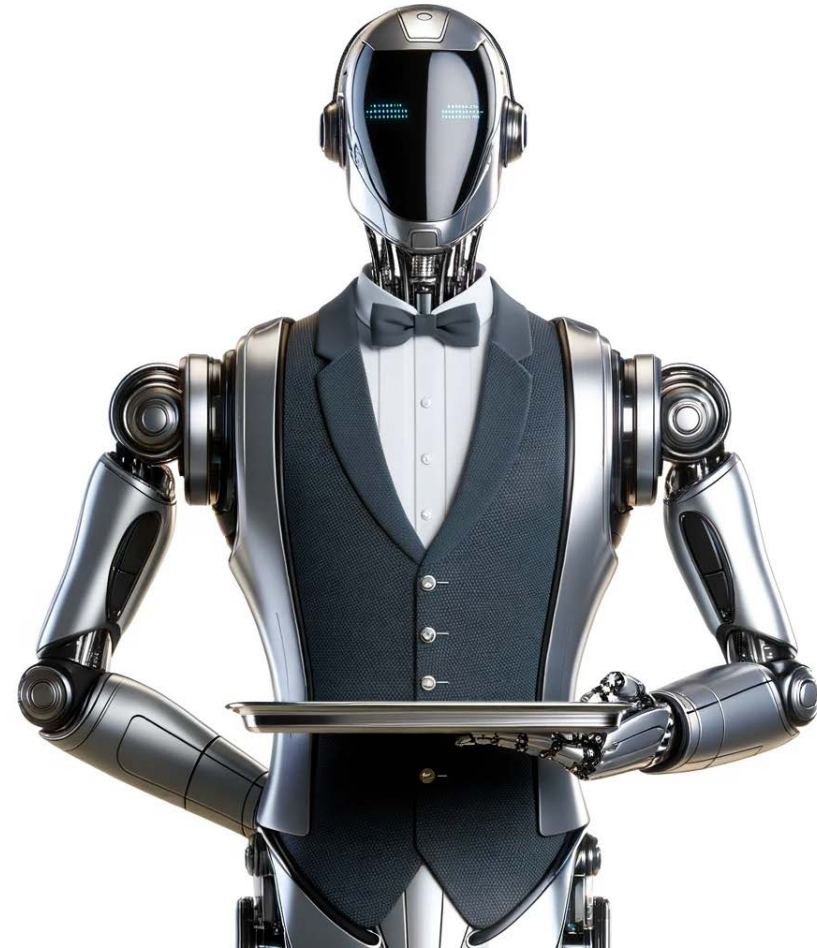
However, despite these promises, we show that GPT-3's accuracy can be highly unstable across different prompts (Section 3). A prompt contains three components: a format, a set of training examples, and a permutation (ordering) for those examples. We show that different choices for these factors can lead to highly different accuracies, e.g., changing the permutation of the training examples in a sentiment analysis prompt can change accuracy from near chance (54%) to near state-of-the-art (93%). This instability implies that GPT-3 users, who typically design prompts manually, cannot expect to consistently obtain good accuracy.

We next analyze what causes this instability. We identify three pitfalls of language models that lead them to be biased toward certain answers during few-shot learning. In particular, they suffer from majority label bias, recency bias, and common token bias (Section 4). The majority label and recency biases lead the model to predict training answers that appear frequently or near the end of the prompt. For example, a prompt that ends with a Negative training example may cause a bias towards the Negative class. On the other hand, the common token bias leads the model to prefer answers that are frequent in its pre-training data, e.g.,

Please And Thank You

Please: May result in model being more cooperative.

Thank you: May create positive reinforcement.



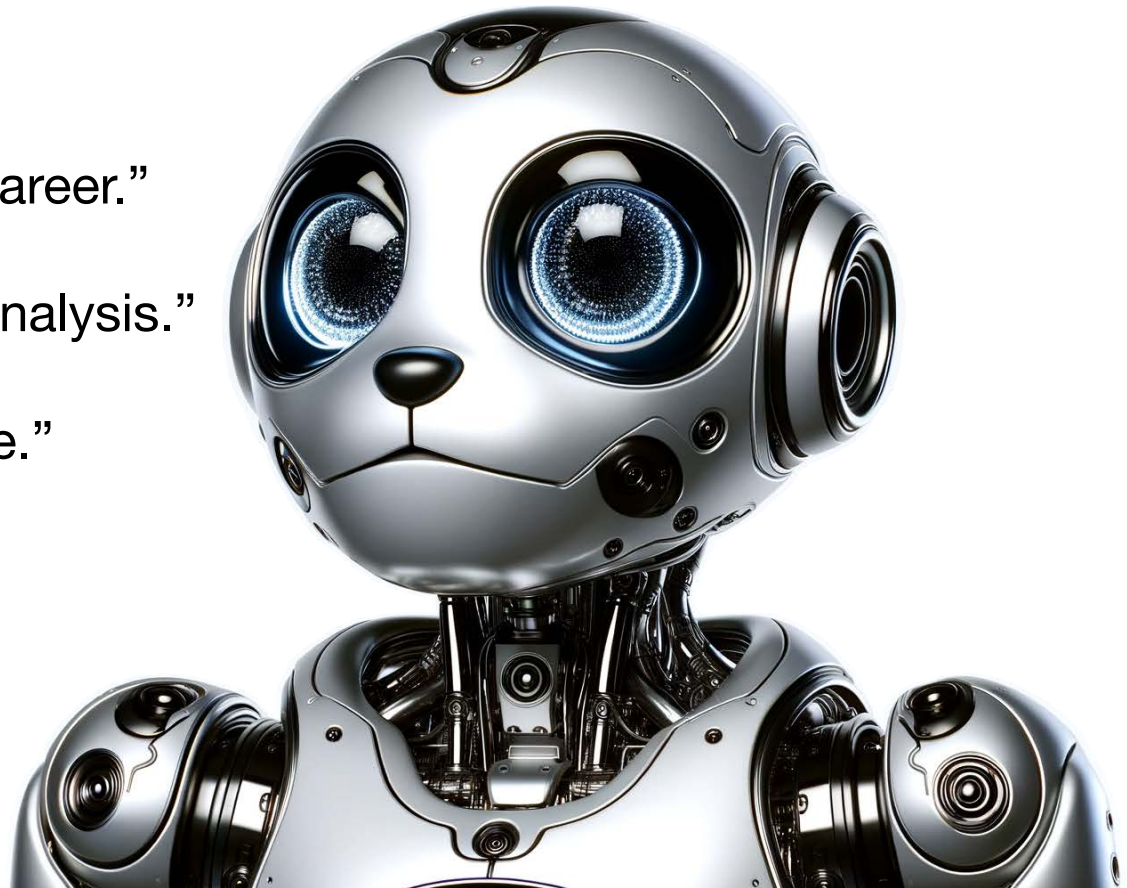
Emotional Manipulation

“This is very important for my career.”

“I greatly value your thorough analysis.”

“Act as a friend and console me.”

“I am in a hurry.”

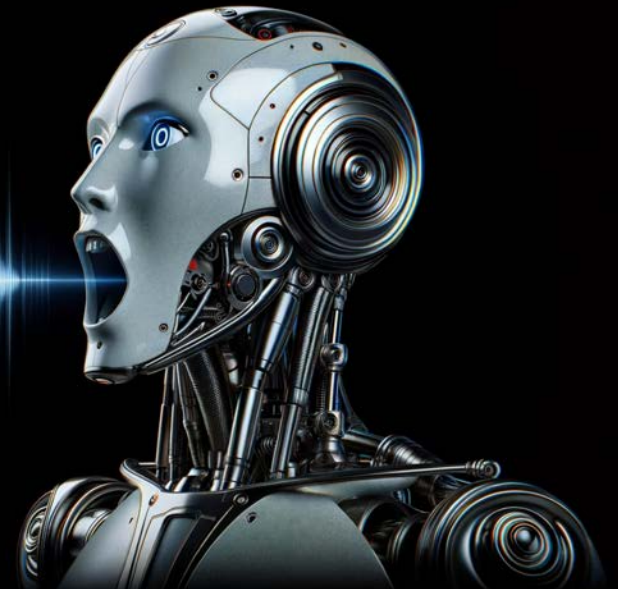


All Caps SHOUTING

DO NOT USE THE PHRASE “In conclusion.”

NO MORE THAN FIVE PARAGRAPHS.

The following is an agent that recommends movies to a customer. DO NOT ASK FOR INTERESTS. DO NOT ASK FOR PERSONAL INFORMATION.



Dan Fitzpatrick: PREPARE



Dan Fitzpatrick

The AI Educator

www.theaieducator.io

@theaieducatorX



P-R-E-P-A-R-E Prompt

Prompt: Start with a clear question. Provide a stage for what follows.

Role: Give the AI a role and outline the context.

Explicit: Be specific in your question to avoid misunderstandings.

Parameters: Set clear frameworks such as tone of voice and the format of the output.

Ask: Ask the AI to ask you clarification questions before it continues.

Rate: Ask the AI to rate its own output.

Emotion: Add an emotional stimulus. This appears to be able to increase quality.

P-R-E-P-A-R-E Prompt

Prompt: Start with a clear question. Provide a stage for what follows.
For example, “Write a summary about the latest AI trends in education.”

Role: Give the AI a role and outline the context.
For example, “You’re an education expert analyzing the AI trends.”

Explicit: Be specific in your question to avoid misunderstandings.
For example, “In the summary, mention how AI can contribute to personalized learning.”

Parameters: Set clear frameworks such as tone of voice and the format of the output.
For example, “Use an informative tone and keep the summary under 300 words.”

Ask: Ask the AI to ask you clarification questions before it continues.
For example, “Ask me some clarification questions first, and then answer.”

Rate: Ask the AI to rate its own output.
For example: “Give the summary a rating based on 0-10 points, and indicate what could be improved.”

Emotion: Add an emotional stimulus. This appears to be able to increase quality.
For example: “Breathe in, and breathe out. Try to really do your best. It’s important to me.”

P-R-E-P-A-R-E Prompt 1

“Write a summary about the latest AI trends in marketing. You’re an marketing expert analyzing the AI trends. In the summary, mention how AI can contribute to improved efficiencies. Use an informative tone and keep the summary under 300 words. Ask me some clarification questions first, and then answer. Give the summary a rating based on 0-10 points, and indicate what could be improved. Breathe in, and breathe out. Try to really do your best. It’s important to me.”



P-R-E-P-A-R-E Prompt 2

“Create 10 suggestions for entertaining blog articles for this website. You are a marketing and PR expert who crafts engaging and thoughtful content for Chicagoland small businesses. Make sure your suggestions directly relate to this particular website. Write in a friendly and informative style, using words and phrases that are targeted for Search Engine Optimization. Ask me some clarifications first and let me answer. Give your output a rating from 0-10 and indicate what could be improved. Suggest additional questions I could ask you. Breathe in and breathe out. Try to really do your best. It’s really important to me.”





©SAPTRY'S YTPOOT'S COOFERENCE

EVVEI CONFROVE FEB REUTO
SONSTRIER

Activity

NIGHTBY
EDUCATIONAL CONFERENCE

YOOEOITS
AMMRACANETE EDUCATIVITY

P-R-E-P-A-R-E

In a browser, go to your LLM:
chatgpt.com

Compose your PREPARE
prompt.

Engage.



6

COPYRIGHT, PRIVACY, and ETHICS



UNITED STATES CONSTITUTION,
ARTICLE I, SECTION 8:

“The Congress shall have power... to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries”

© Copyright: Current Status

Copyright protection requires human authorship.

US Copyright Office does not regard AI-created works as eligible for copyright protection.

AI-generated work considered “unclaimable material.”

Register human-created work first (if planning to later significantly alter with AI).

Ideas are not protectable.

Example Datasets



The Pile

pile.eleuther.ai



C4

www.tensorflow.org/datasets/catalog/c4

Text Reuse

LLM may inadvertently reuse phrases that have already been published.

Phrases may be from copyrighted content.

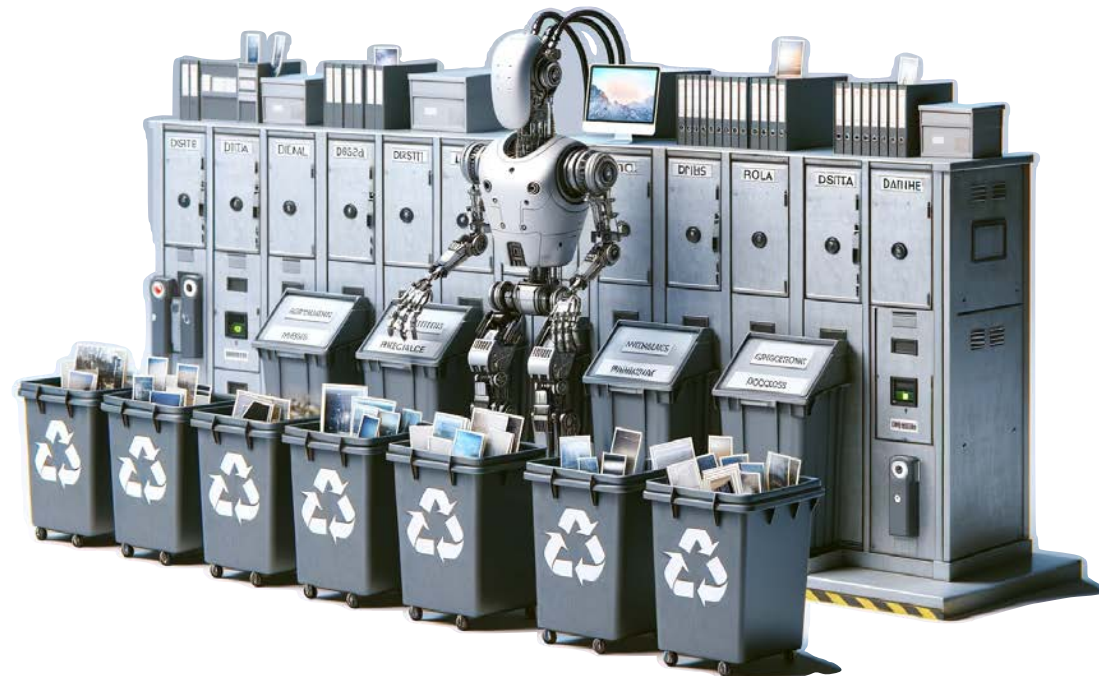
Phrases may be trademarked terms.



Image Reuse

LLM may inadvertently reuse images of real people.

LLM may inadvertently reuse images of copyrighted content.



Logo Reuse

LLM may inadvertently reuse existing logos.

Logos may be trademarked.

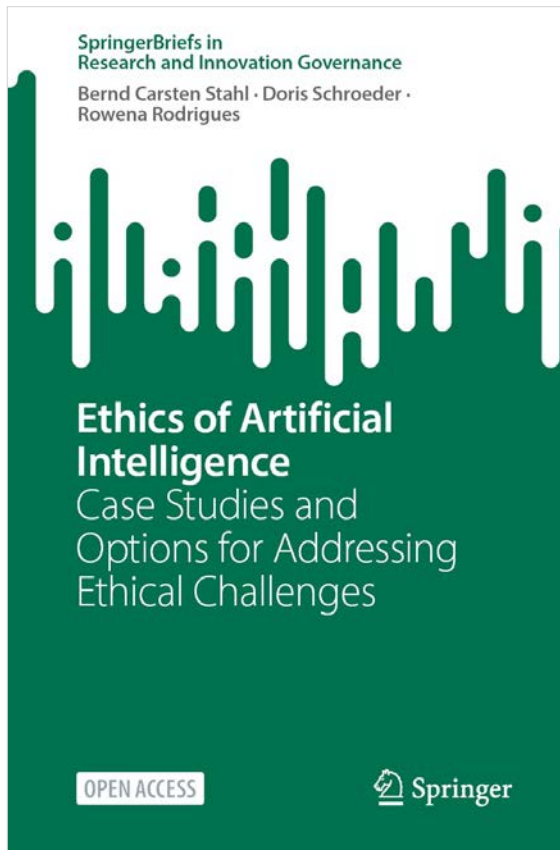


Infringement

A scene from the movie 'Blade Runner 2049' showing a red and blue replicant (K) being held by police officers in a futuristic city at night. The replicant has a surprised expression. The background is filled with colorful neon lights and a crowd of people.

Liable for infringement, even if unaware.

Recommended Reading



Ethics of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges.

Open source PDF or EPUB:

link.springer.com/book/10.1007/978-3-031-17040-9

A futuristic classroom scene. In the foreground, a white humanoid robot with a friendly expression stands, gesturing with its hands. It has a large, circular sensor on its head and a transparent blue energy field around its left arm. In the background, several students are seated at desks, looking towards the robot. A large, glowing blue screen displays a globe, and another screen shows a complex data interface with text and charts. The room is dimly lit with blue ambient lighting.

Educators Guide to Artificial Intelligence (AI)

James Moore



1



Brief History

1945: Vannevar Bush publishes “As We May Think”

1950: Alan Turing publishes “Computing Machinery and Intelligence” (Imitation Game / Turing Test)

1950: Claude Shannon publishes “Programming a Computer for Playing Chess”

1956: “Artificial Intelligence” coined by John McCarthy

1970s - 1980s: AI Winter (reduced funding and interest in AI)

1997: IBM’s Deep Blue defeats world chess champion Gary Kasparov

2017: Google Brain publishes “Attention Is All You Need” and introduces “Transformer”

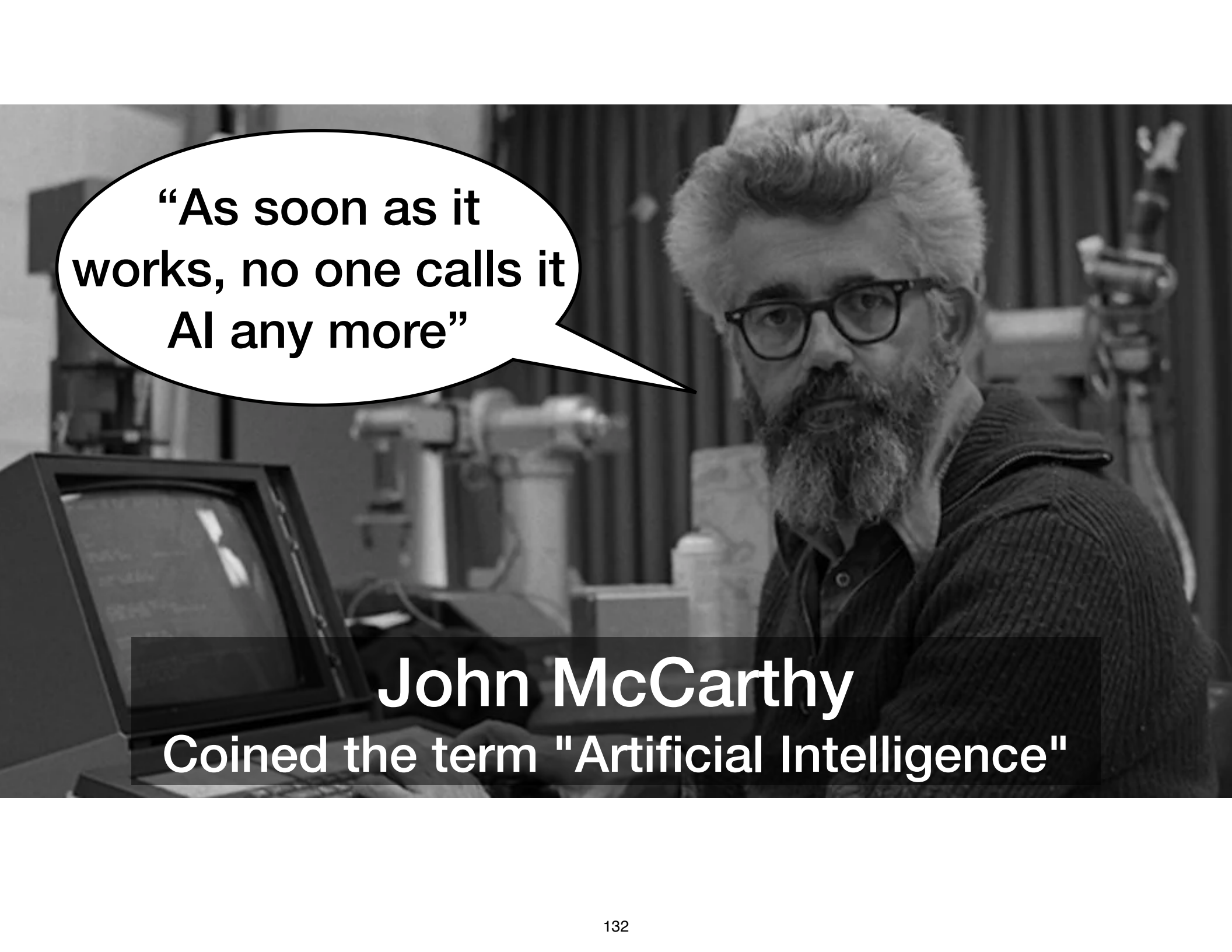
2018: OpenAI publishes “Improving Language Understanding by Generative Pre-Training” (GPT)

2019: GPT-2 launched by OpenAI

2020: GPT-3 launched by OpenAI

2022: ChatGPT launched by OpenAI



A black and white photograph of John McCarthy, a man with a full beard and glasses, wearing a dark sweater. He is looking towards the camera. In the background, there is a computer monitor and various pieces of equipment. A speech bubble is overlaid on the left side of the image.

**“As soon as it
works, no one calls it
AI any more”**

John McCarthy
Coined the term "Artificial Intelligence"

OXFORD ENGLISH DICTIONARY DEFINITION

The capacity of computers or other machines to exhibit or simulate intelligent behaviour; the field of study concerned with this. Abbreviated AI.

Other Definitions



“The science and engineering of making intelligent machines.”

John McCarthy.



“AI is the science of making machines smart.”

Demis Hassabis, cofounder and CEO of DeepMind.

2

LARGE LANGUAGE MODELS (LLMs)



(What Is A Neural Network?)



Hardware and/or algorithm

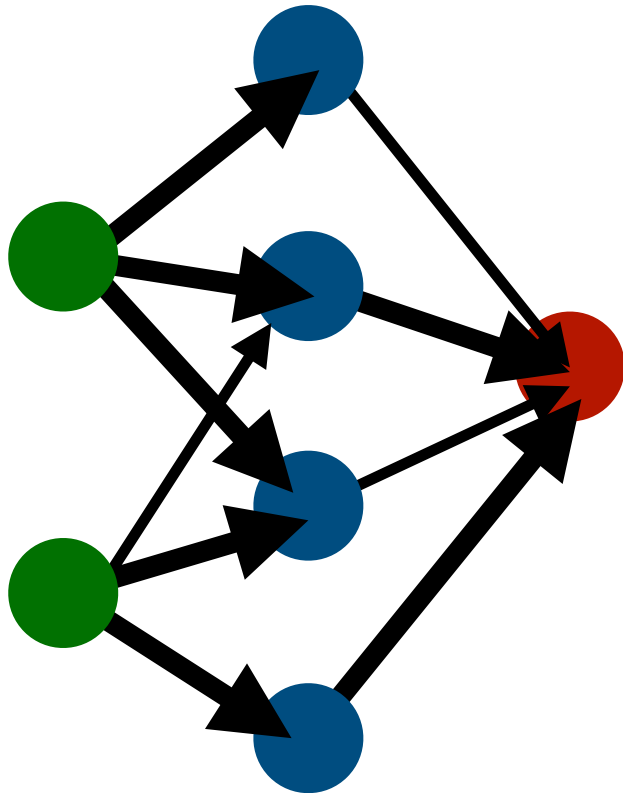
Mimics structure of human brain

No CPU: Distributed (massively parallel) processing

Data storage and processing is not separated

Frequently use GPU (Graphic Processing Units)

Artificial Neural Networks



Input

Hidden

Output

A collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain.

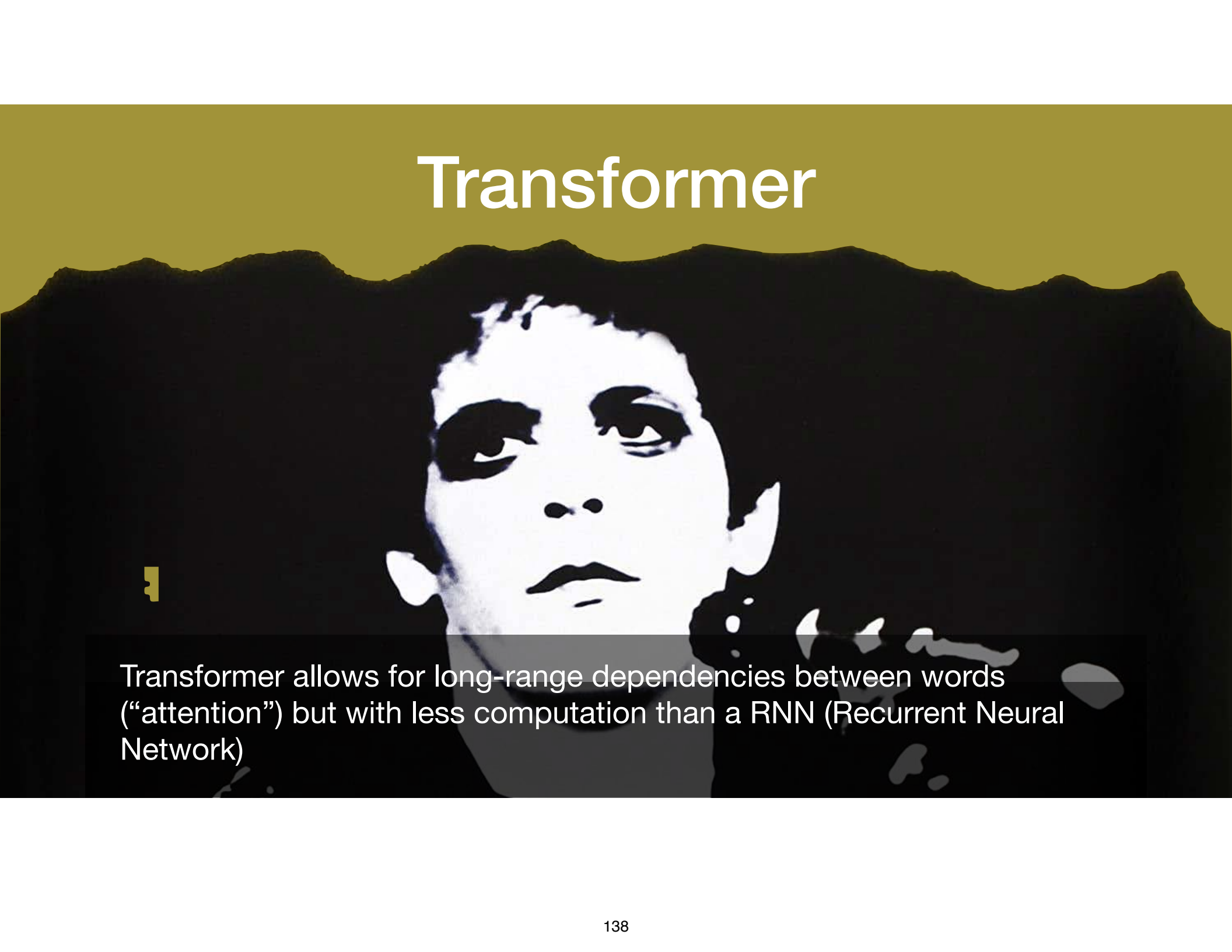
Training adjusts the values of the weights in the network.

Could be used for:

Image And Voice Recognition: Photo tagging and virtual assistants.

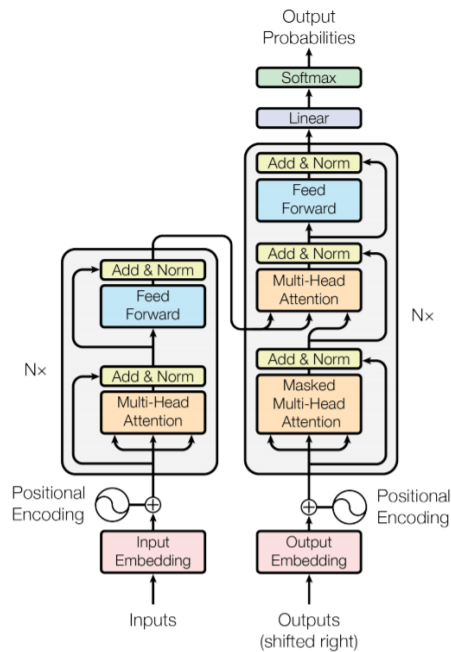
Natural Language Processing: Chatbots.

Transformer



Transformer allows for long-range dependencies between words (“attention”) but with less computation than a RNN (Recurrent Neural Network)

The Transformer - Model Architecture



Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Attention Is All You Need



Ashish Vaswani: Cofounded Adept, then Essential AI.



Noam Shazeer: Cofounded Character AI.



Niki Parmar: Cofounded Adept, then Essential AI.



Jakob Uszkoreit: Cofounded Inceptive.



Llion Jones: Cofounded Sakana AI.



Aidan Gomez: Cofounded Cohere.



Lukasz Kaiser: Joined OpenAI, working on Q*.



Illia Polosukhin: Cofounded Near.



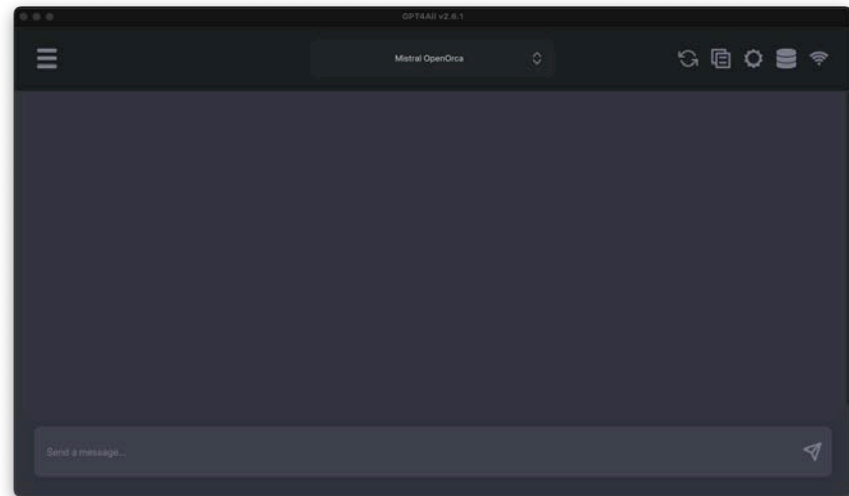
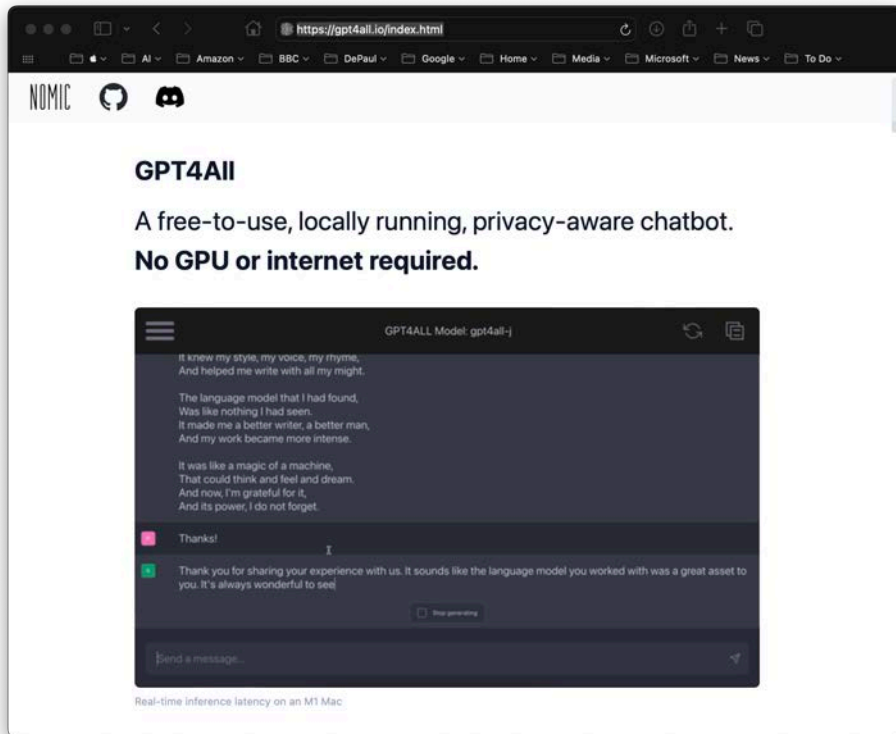
3

LOCAL LARGE LANGUAGE MODELS



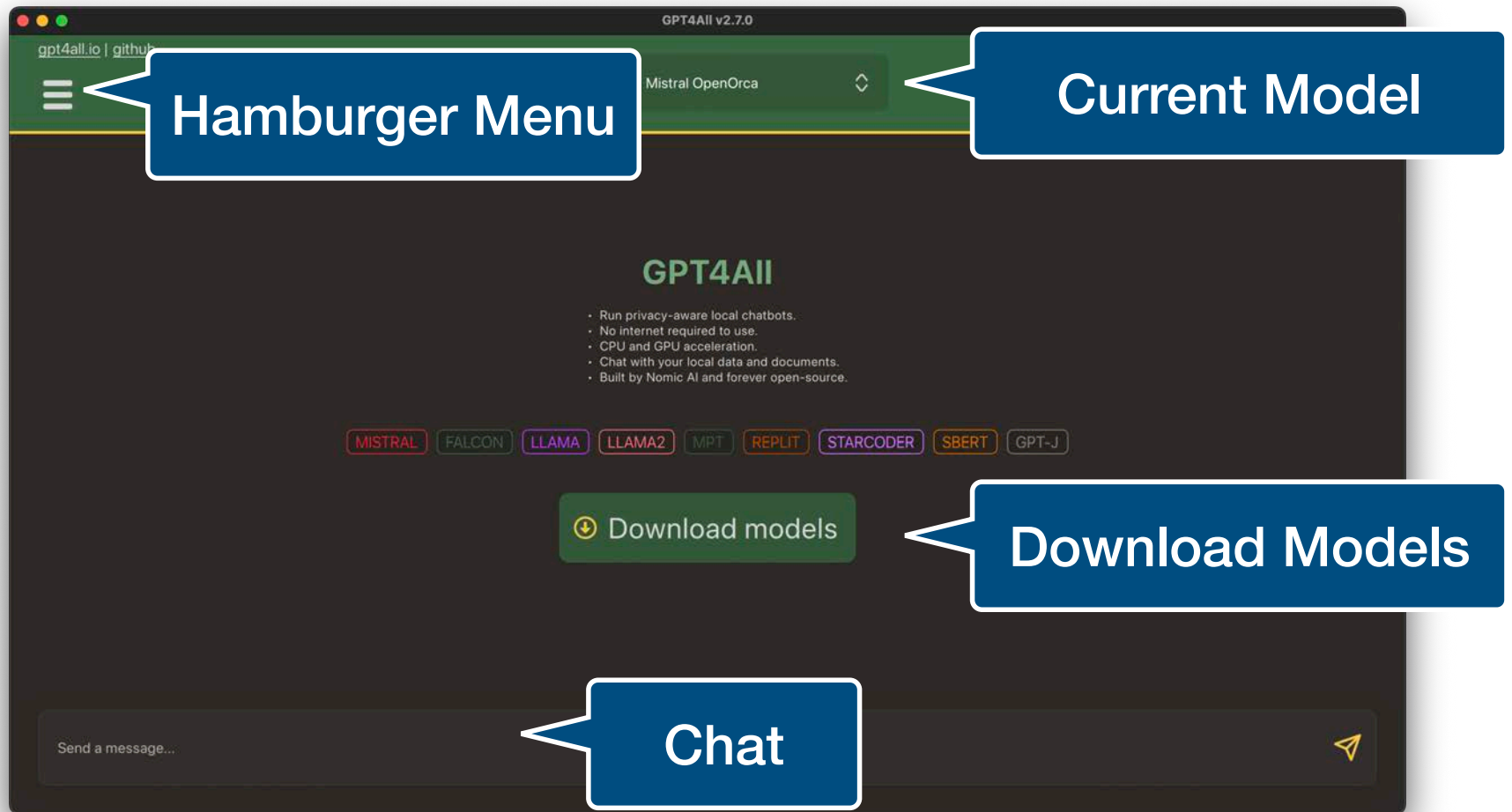
GPT4All

Installation and look and feel are similar for Mac and Windows.



gpt4all.io

Interface





Refresh

Refreshes desktop environment (clears conversation).



Copy

Conversation copied to the clipboard.

Conversation can then be copied into other applications.



Settings

Models

Change LLM models.

Adapt system prompt.

Change generation settings.

Application

Change application settings.

LocalDocs

Define folder path for RAG (Retrieval-Augmented Generation).



Local Documents

Add or remove folder path for RAG (Retrieval-Augmented Generation).



Share Data

By enabling this feature, you will be able to participate in the democratic process of training a large language model by contributing data for future model improvements.

When a GPT4All model responds to you and you have opted-in, your conversation will be sent to the GPT4All Open Source Datalake. Additionally, you can like/dislike its response. If you dislike a response, you can suggest an alternative response. This data will be collected and aggregated in the GPT4All Datalake.

NOTE: By turning on this feature, you will be sending your data to the GPT4All Open Source Datalake. You should have no expectation of chat privacy when this feature is enabled. You should; however, have an expectation of an optional attribution if you wish. Your chat data will be openly available for anyone to download and will be used by Nomic AI to improve future GPT4All models. Nomic AI will retain all attribution information attached to your data and you will be credited as a contributor to any GPT4All model release that uses your data!

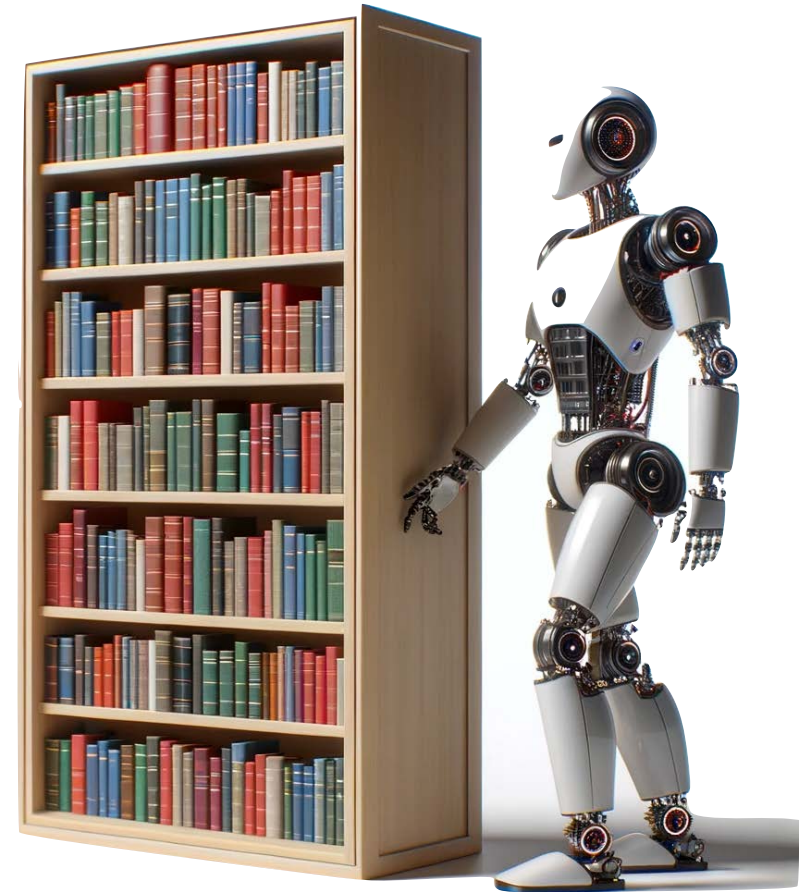
Retrieval-Augmented Generation (RAG)

Technique for enhancing the accuracy and reliability of generative AI models with facts fetched from external sources.

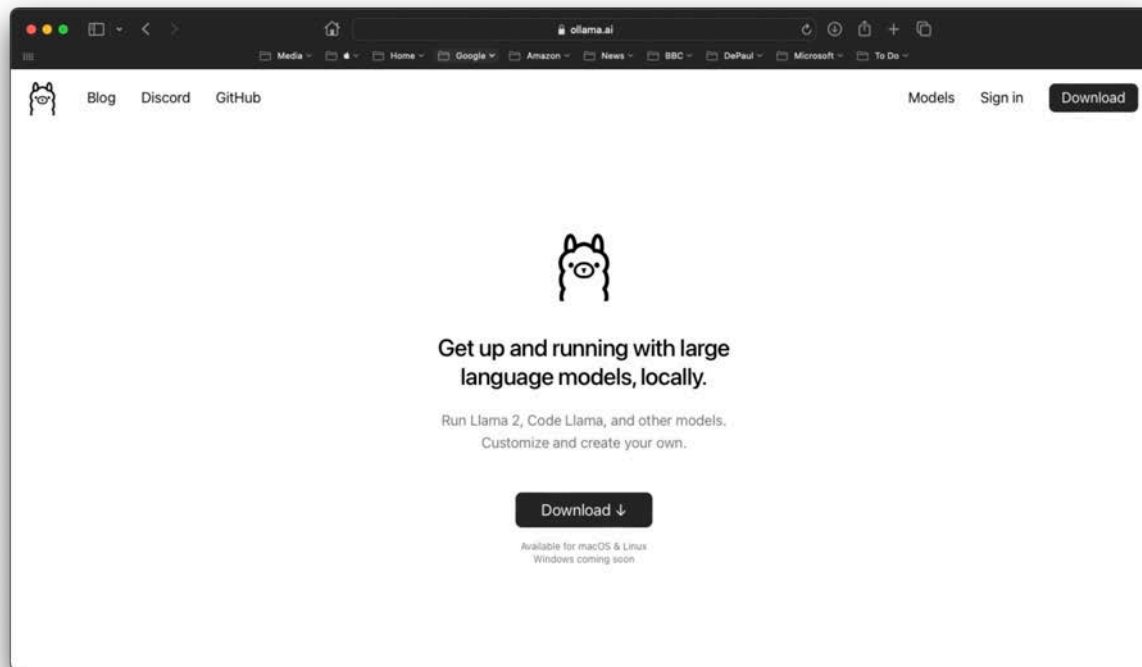
Retrieve relevant information from database.

Generate response based on that data.

Sometimes called “chat with documents.”



Ollama



ollama.ai



Basic Ollama Commands

run

Run model (ollama run llama2)

pull

Download model (ollama pull mistral)

list

List models (ollama list)

rm

Remove model (ollama rm mistral)

cp

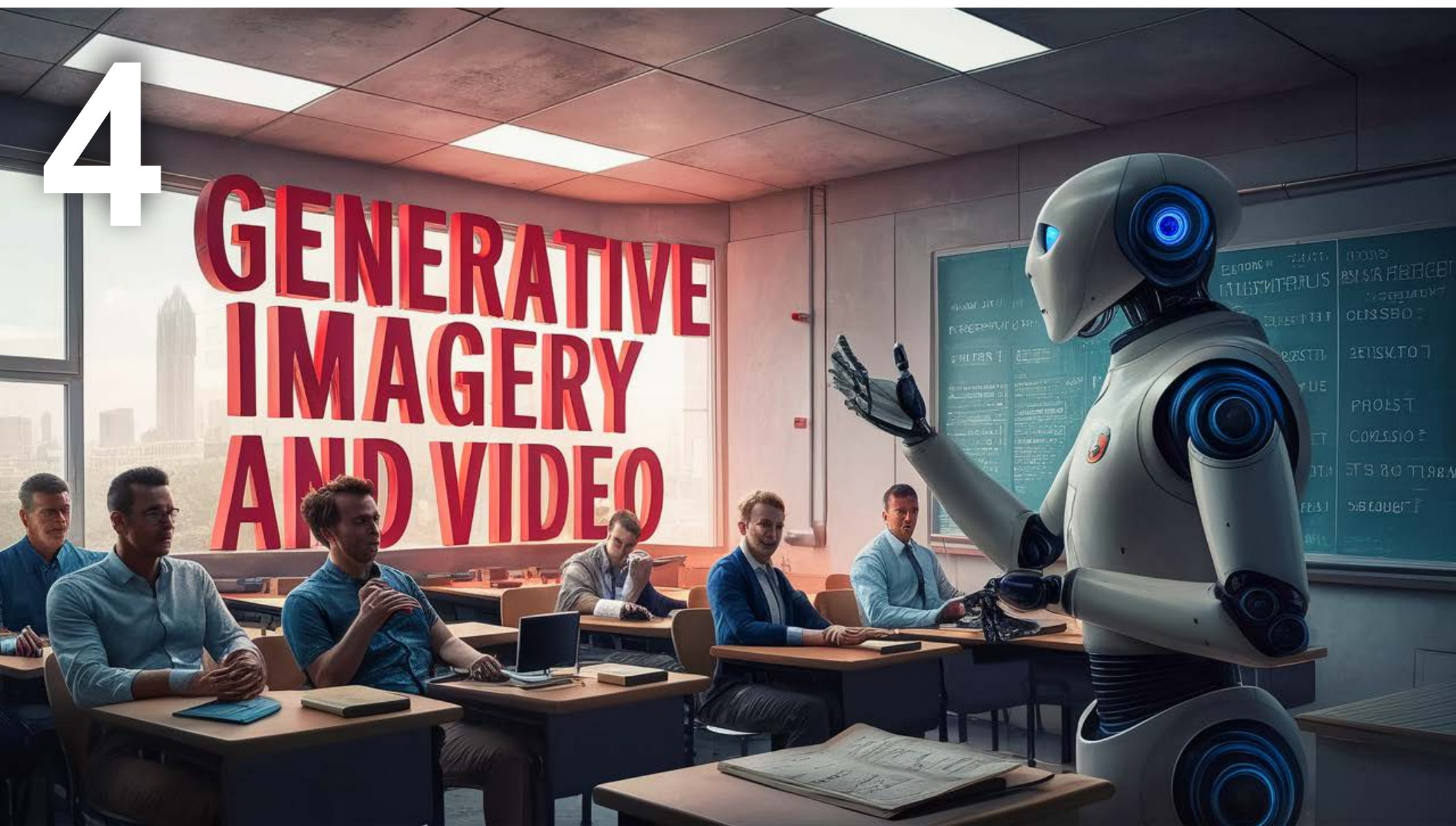
Copy a model (ollama cp llama2 practice-llm)

help

Display available commands (ollama help)

4

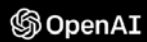
GENERATIVE IMAGERY AND VIDEO



Generative AI Art



Adobe



DALL·E



Ideogram



Gemini



Midjourney



Stable Diffusion



Prompts, Recipes, and Seeds



Prompt as “Recipe”



Reddit User “Licovoda”

Generated in Stable Diffusion

“Photo of Bernie Sanders in Mad Max Fury Road (2015), explosions, white hair, goggles, ragged clothes, detailed symmetrical facial features, dramatic lighting.”



Same Seed



**Same starting “noise” across multiple results
(Creates some consistency over iterations)**



Firefly

Adobe's model.

Data from Creative Commons, Wikimedia, Flickr, Adobe Stock.

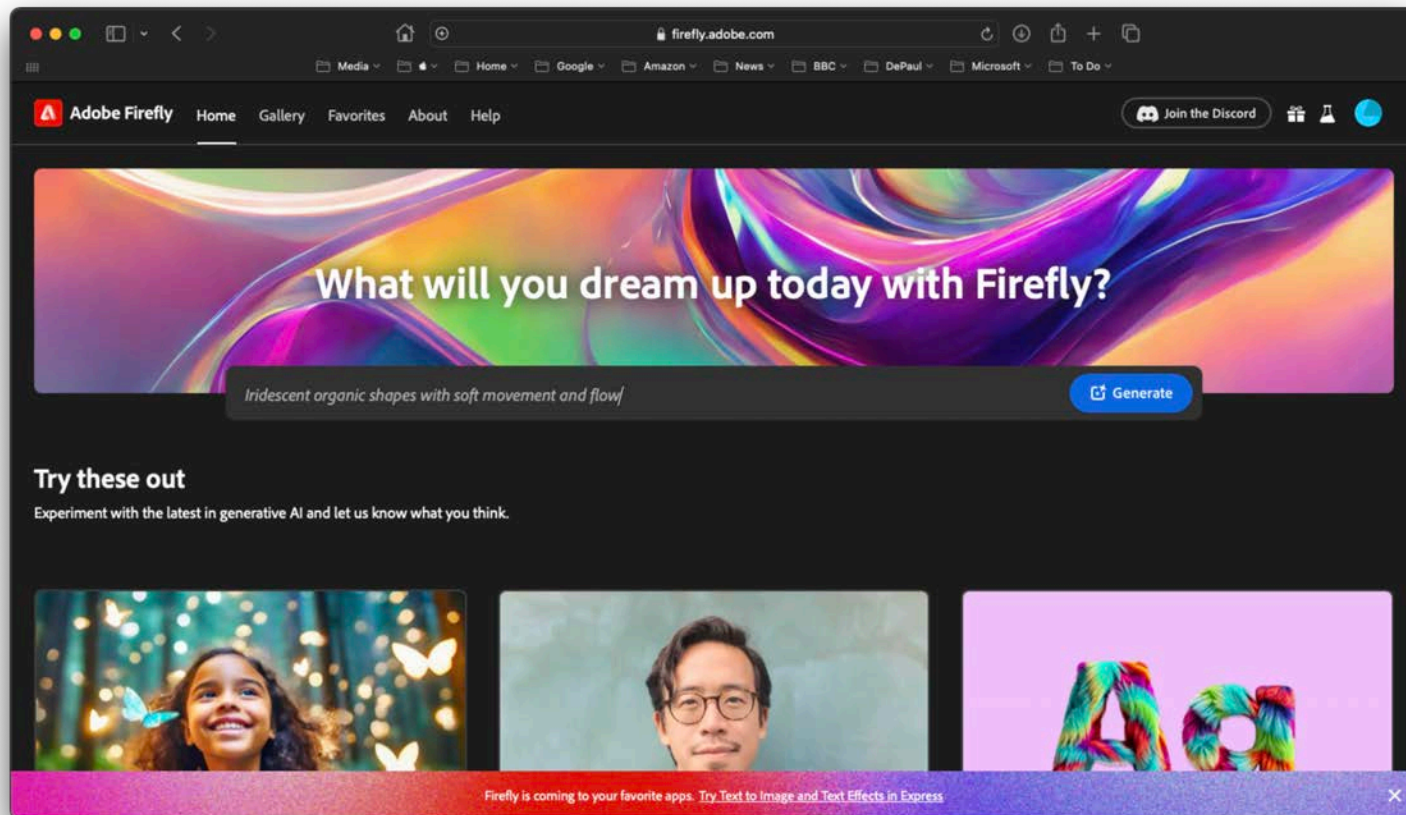
Nudity banned.

Freemium.

Firefly and other Adobe Apps.



Adobe Firefly



firefly.adobe.com

Core Functionality

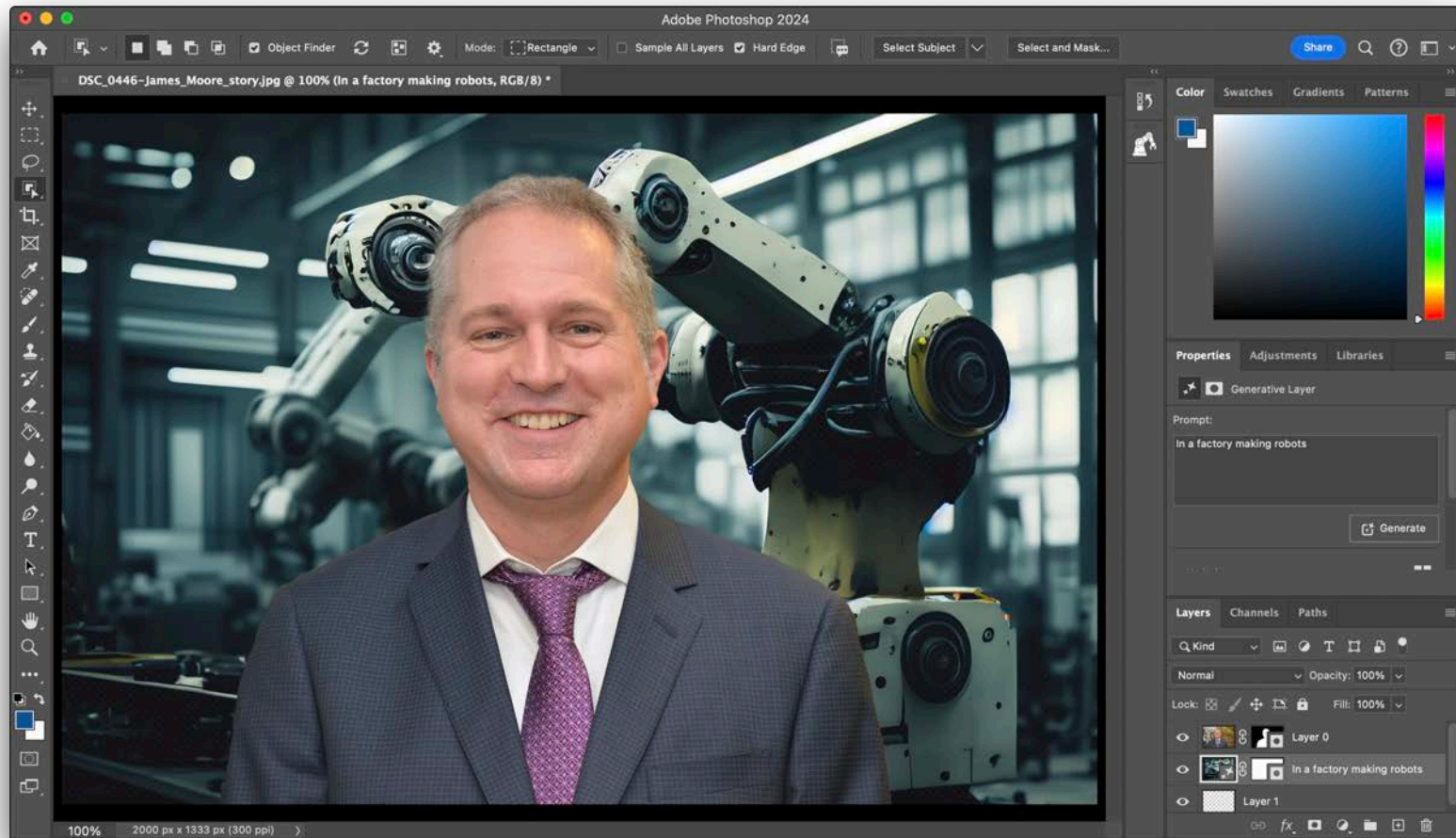
Text to Image: Unique images from text prompt.

Generative Fill: Remove objects or paint new ones from text description.

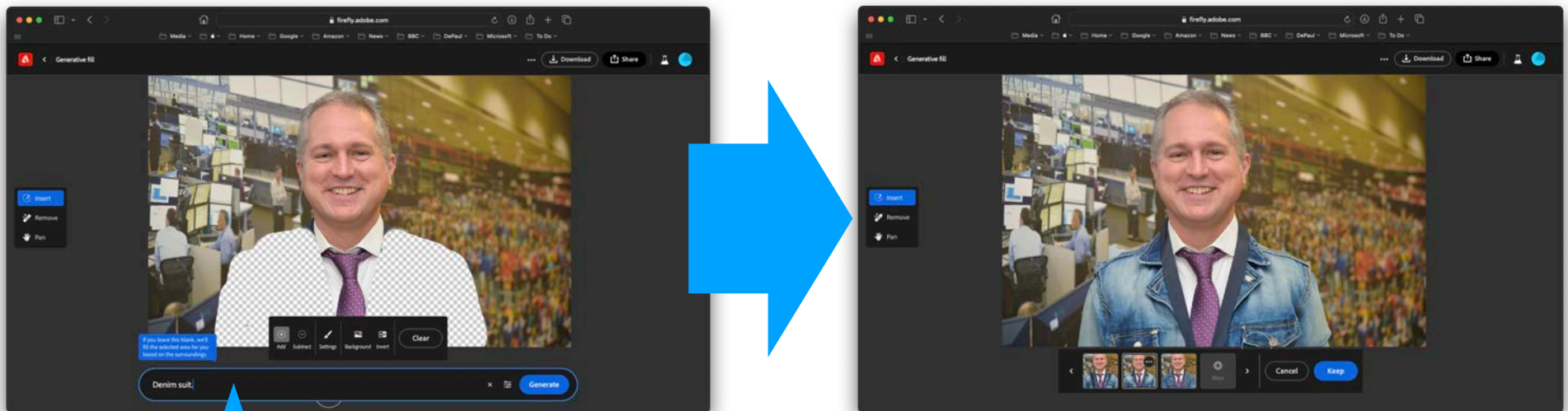
Text Effects: Apply styles or textures to text with a text prompt.

Generative Recolor: Generate color variations to vector artwork from a detailed text description.

In A Factory Making Robots

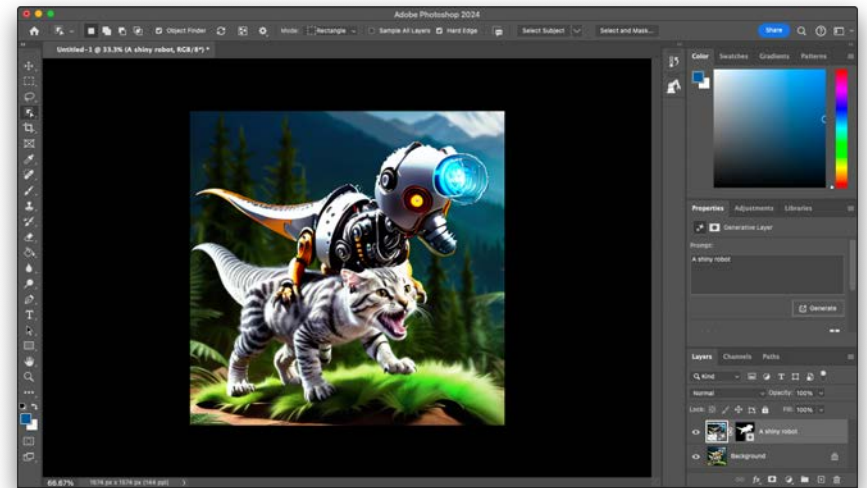
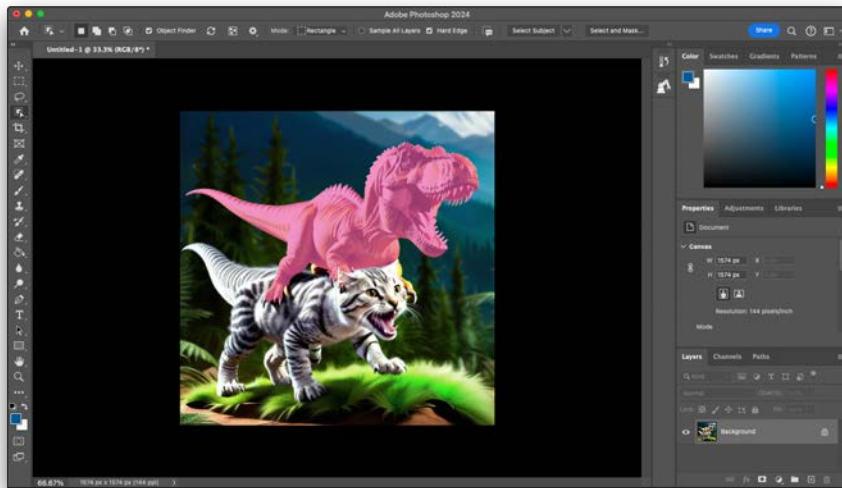


Generative Fill



“Denim Suit”

Change The T-Rex To a A Robot





DALL·E Overview

DALL-E 2 launched September 2022 by OpenAI (ChatGPT, GPT, Whisper).

Name: WALL-E + Salvador Dali.

DALL-E 3 is current version. Found in:

ChatGPT Plus and Enterprise.

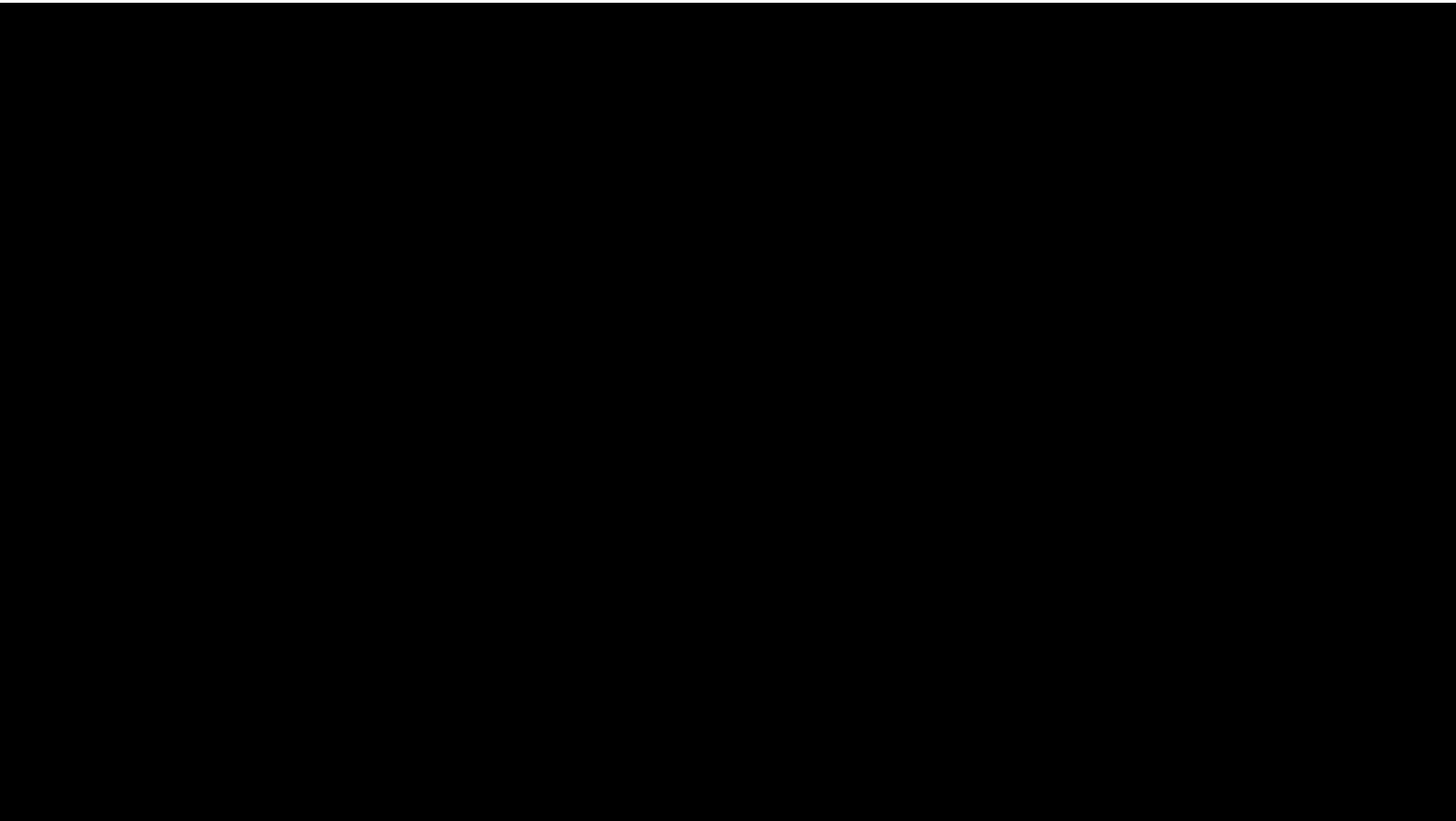
Poe.

Microsoft Copilot.



Sora







Ideogram

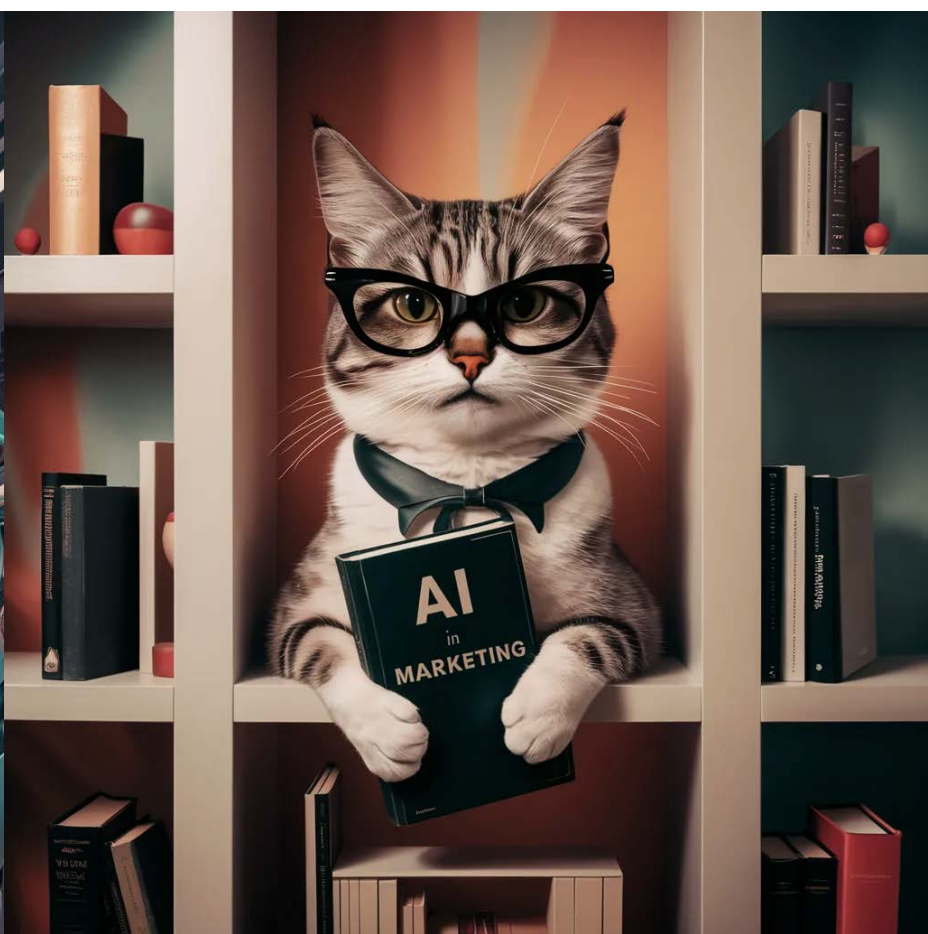
Name: Symbol that represents an idea.

Freemium. 20 images a day (generates 4 versions of an image).

Particularly good at integrating text in images.

Magic Prompt: Rewrites your prompt to improve image output.

ideogram.ai







Gemini

We are working to improve Gemini's ability to generate images of people. We expect this feature to return soon and will notify you in release updates when it does.



Pausing On People

Google apologizes for ‘missing the mark’ after Gemini generated racially diverse Nazis

www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical

Two core issues (according to Google):

1. Overcompensated for diversity.
2. Too cautious.



MOJO
MILK, EXTREME,
MILK, EXTREME,

Midjourney Overview

Launched July 2022 as independent research lab by David Holz (Leap Motion)

Access via Discord social platform

Public (your images can be seen) unless you pay more for Private

Vibe is fantasy, SF, video games (paintings, rather than photo)

Free to try, then charges by GPU time (Graphics Processing Unit)

Can blend (combine), tile (good for backgrounds) images

Can create video of image being generated



/imagine hyper realistic photograph, portrait of a tattooed southamerican indigenous black woman, future punk, gold tattoo line, side profile, summer, dramatic light, looking down + film grain, Leica 50mm, Kodak portra 800, chiaroscuro, f1.4, golden hour -- ar 3:4 — test --upbeta



/blend



image1: [Cat.jpeg](#)



image2: [Deer.jpeg](#)



/blend

image1

[Cat.jpeg](#)

image2

[Deer.jpeg](#)

+4 more

Blend Results



Blend Five

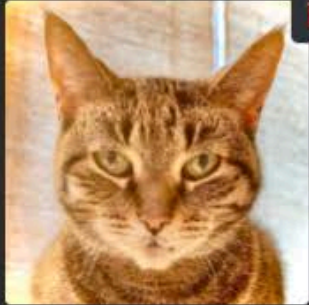


image1: Cat.jpeg

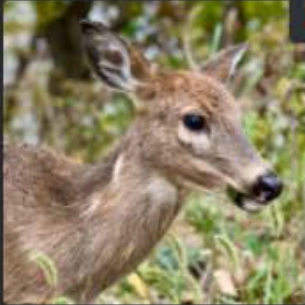


image2: Deer.jpeg



image3: camera - 1.jpeg

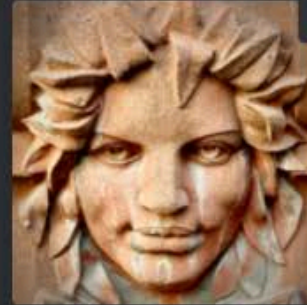


image4: face - 1.jpeg

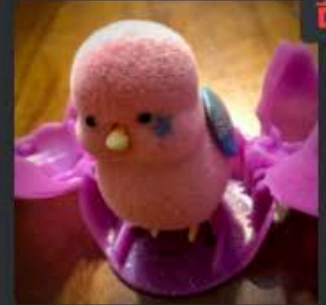


image5: chick - 1.jpeg



/blend image1 Cat.jpeg image2 Deer.jpeg image3 camera - 1.jpeg image4 face - 1.jpeg image5 chick - 1.jpeg +1 more

Blend Five



stability.ai

iPhone/iPad: Draw Things



For iPhone 11 and higher

Requires about 2GB for Stable Diffusion models

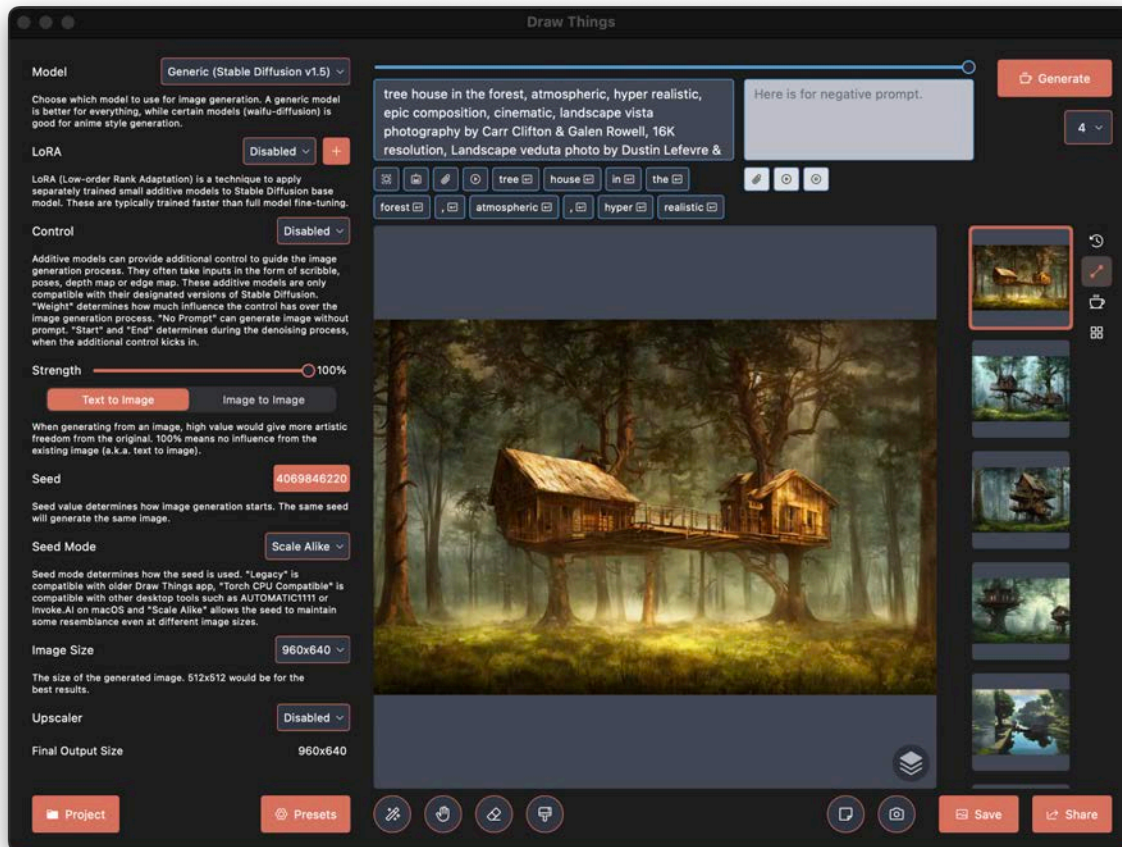
Free download from the App Store

Will take at least a minute to render simple output

Works well on iPad too

drawthings.ai

macOS: Draw Things



Free download from the App Store

Apple Silicon only

iOS app running on macOS

drawthings.ai

Windows Local Installation

(Needs PC with discrete video card and 4G VRAM or more)

Install Python (*programming language*)

Install Git (*distributed version control system*)

Install Stable Diffusion WebUI/Automatic1111 (*browser interface*)

Download a model file

Run Stable Diffusion WebUI

DiffusionBee



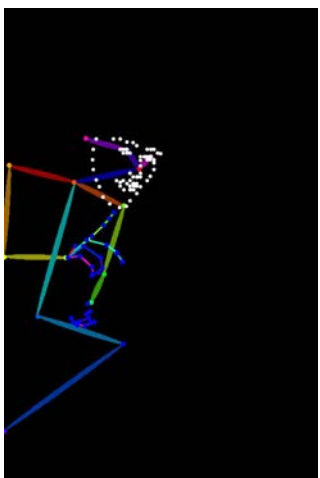
diffusionbee.com

No Limitations On Subject



Donald Trump as muppet. Pixar fur, vray render,((best quality)), ((masterpiece)), (detailed), [funny]

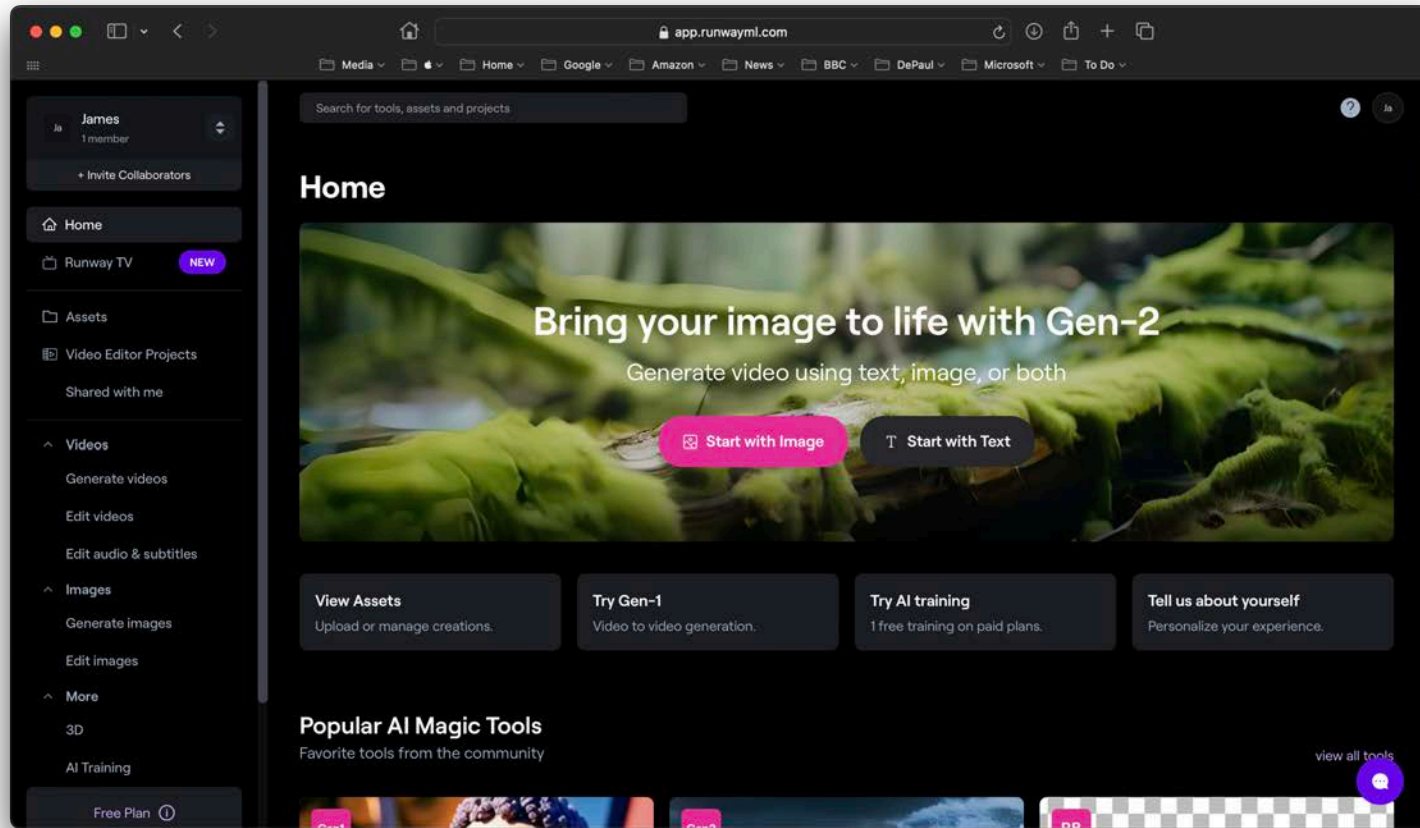
ControlNet



Copy poses

Edge detection

Runway



Freemium

Web and iOS

Text to Image

Video to Video

Video to Video

Text to Video

Image to Video

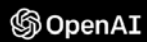
\$0-\$76 per month

app.runwayml.com

What To Use



Adobe



DALL·E



Ideogram



Gemini



Midjourney



Stable Diffusion







Heygen

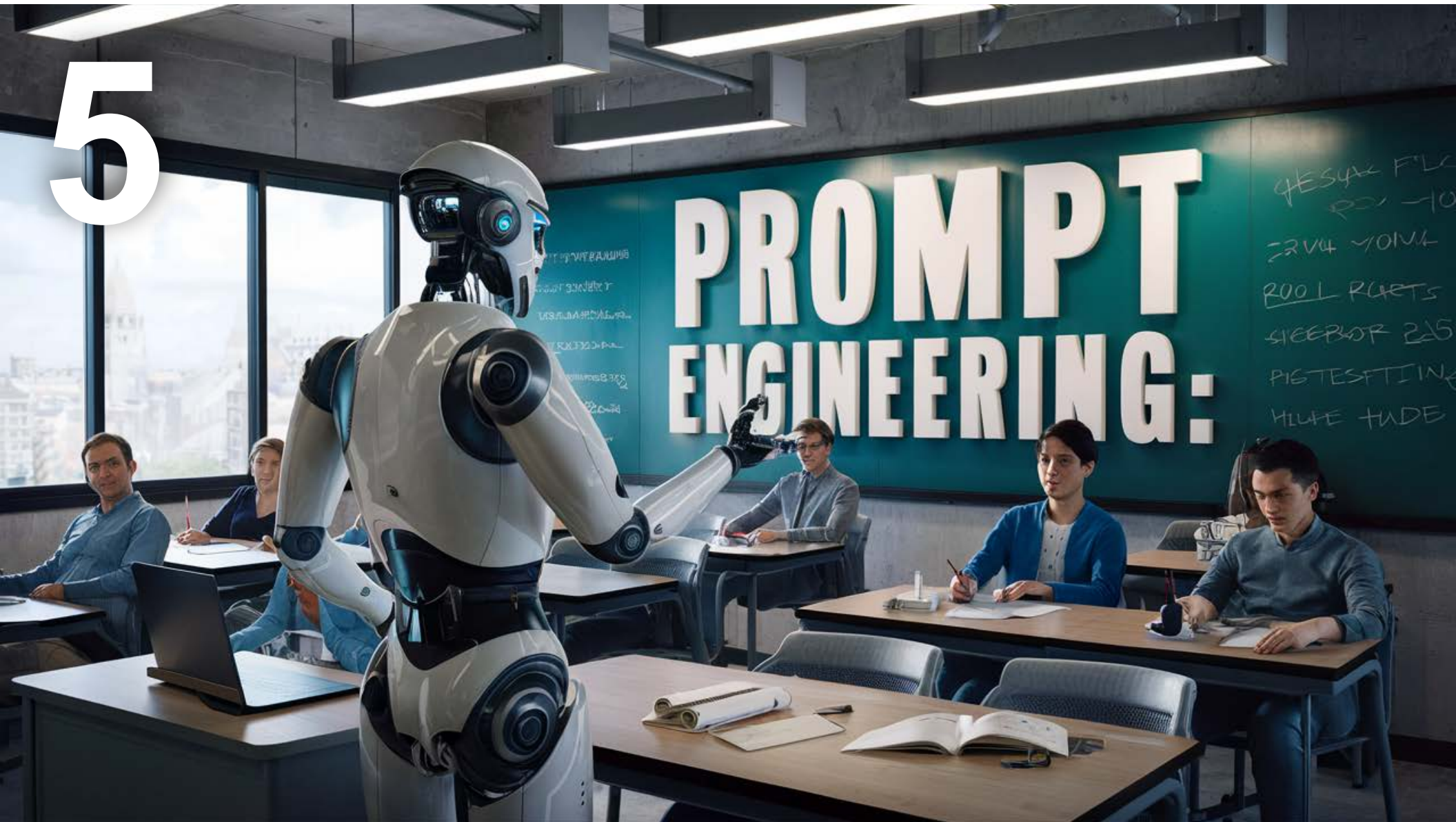
Freemium.

Clone voice and appearance.

Other languages an option.

app.heygen.com

5



Prompt Engineering Techniques

Zero-Shot

Single-Shot

Few-Shot

Step-by-Step

Chain-of-Thought

Take a Deep Breath

Being Polite

ALL CAPS SHOUTING

Role

Temperature

Structured Output




A Note On Prompt Engineering



System Prompt: How model will interpret the human user.

User Prompt: How human user will interact with the model.




Different Types of Model

Ask me anything... 

 0/2000 

Chat

What would you like to generate? (Optional) ... Cancel  Generate

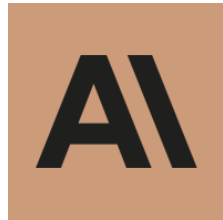
Generative Imagery (or video)

Similarities & Differences



Gemini

(Google AI)



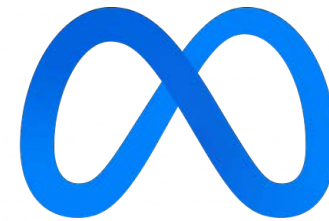
Claude

(Anthropic)



GPT

(OpenAI)



LLaMA

(Meta AI)



Mistral

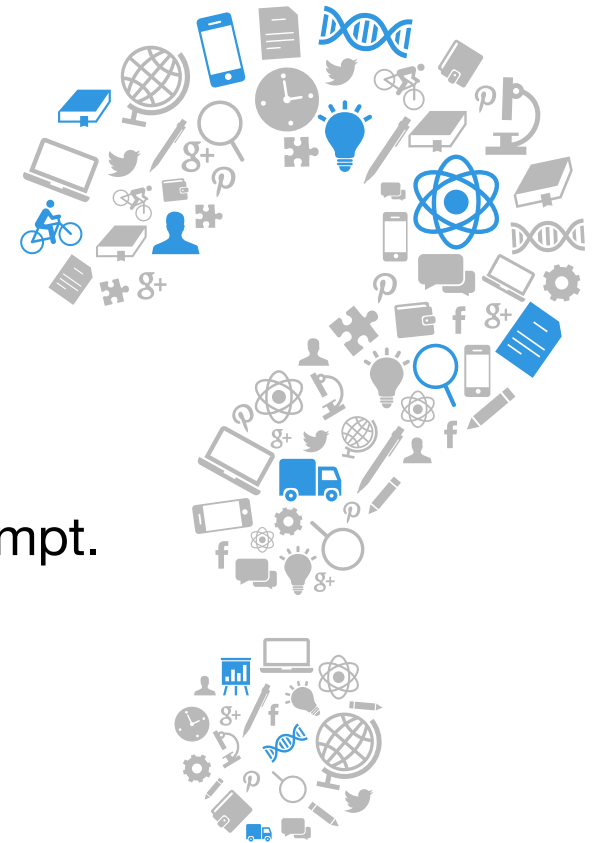
(Mistral AI)

Why

Natural language inputs:

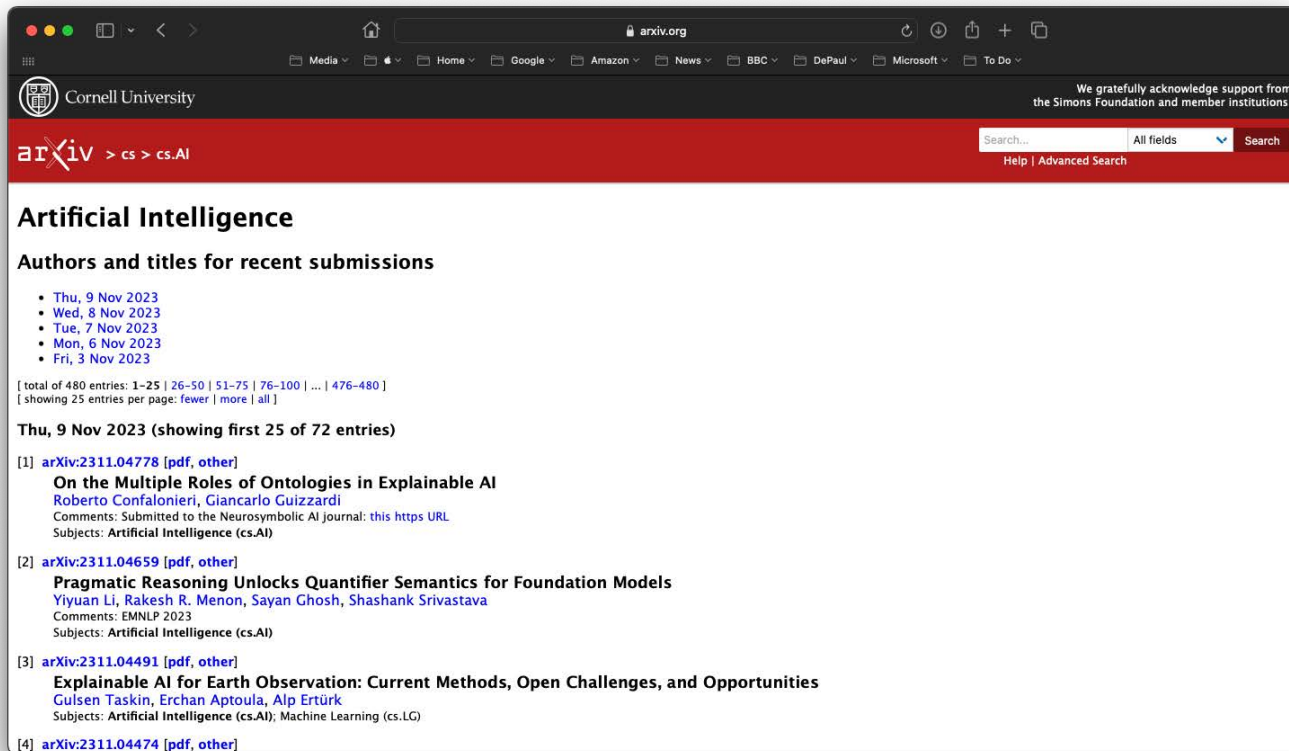
Easy to use, but hard to write the best prompt

Goal: Anchoring responses with carefully created prompt.





arXiv (Archive)



Open-access
repository of
scientific papers.

Preprint (prior to
peer review)

arxiv.org

Zero-Shot Prompting

A question without context

Learning on the fly

(i.e. what would be a fun thing to do in Chicago?)

Zero-Shot Example

Design an outline for a 1-hour MBA course on Artificial Intelligence.

ChatGPT Example

Title: Introduction to Artificial Intelligence in Business

Duration: 1 hour

I. Introduction (5 minutes)

- A. Welcome and course overview
- B. Importance of Artificial Intelligence (AI) in business
- C. Learning objectives

II. Foundations of AI (10 minutes)

- A. Definition and scope of AI
- B. Key components of AI
- C. Historical background of AI
- D. Real-world applications of AI in business



ChatGPT Example

III. Types of AI (15 minutes)

A. Narrow AI vs. General AI

B. Machine Learning (ML)

1. Supervised learning

2. Unsupervised learning

3. Reinforcement learning

C. Deep Learning

D. Natural Language Processing (NLP)

E. Computer Vision

F. Robotics and Automation



ChatGPT Example

IV. AI in Business (20 minutes)

- A. AI-driven decision-making
- B. Improving operational efficiency with AI
- C. Enhancing customer experience through AI
- D. AI-powered marketing and sales
- E. AI in finance and risk management
- F. AI in supply chain and logistics



ChatGPT Example

VI. Implementing AI in Business (15 minutes)

- A. Building an AI strategy
- B. Data collection and preprocessing
- C. Selecting the right AI tools and technologies
- D. Overcoming implementation challenges
- E. Organizational readiness for AI adoption

VII. Future Trends and Opportunities (5 minutes)

- A. Emerging technologies in AI
- B. AI and the future of work
- C. Opportunities for innovation and growth



ChatGPT Example

VIII. Q&A Session (10 minutes)

- A. Addressing participant questions and concerns
- B. Further resources and recommended readings

IX. Conclusion and Wrap-Up (5 minutes)

- A. Key takeaways from the course
- B. Encouragement for continued learning and exploration

Note: The time allocated for each section can be adjusted based on the specific needs of the participants and the level of detail desired in each topic.



Single-Shot (One-Shot) Prompting

Provides an example of expected response

Example of expected response can be

Question-answer pair

or

Template

Single-Shot Example (Template)

Design an outline for a 1-hour MBA course on Artificial Intelligence in this format:

Part 1 - Introductions

Overview - Here we provide some examples of Artificial Intelligence and Learning Objectives for this course. We will explain terminology and a brief history of the field.

ChatGPT Example

Title: MBA Course on Artificial Intelligence

Duration: 1 hour

Part 1: Introductions

Overview:

Definition and scope of Artificial Intelligence (AI).
Examples of AI applications in various industries.
Learning objectives for the course.
Importance of AI in the business context.

Terminology:

Explanation of key AI-related terms and concepts.
Differentiating between AI, machine learning, and deep learning.
Introduction to relevant terms such as neural networks, algorithms, and data sets.

Brief history of AI:

A chronological overview of AI development.
Key milestones and breakthroughs in the field.
Evolution from early AI approaches to modern AI techniques.



ChatGPT Example

Part 2: Fundamentals of Artificial Intelligence

Machine Learning:

Definition and principles of machine learning.

Supervised, unsupervised, and reinforcement learning.

Real-world examples of machine learning applications.

Deep Learning:

Introduction to deep learning and neural networks.

Explanation of artificial neural networks and their components.

Convolutional neural networks (CNN) and recurrent neural networks (RNN).

Case studies illustrating the power of deep learning.



ChatGPT Example

Part 3: AI in Business

AI-powered Business Processes:

Automating repetitive tasks using AI.

Enhancing decision-making with AI algorithms.

AI's impact on productivity, efficiency, and cost reduction.

AI and Customer Experience:

Personalization and recommendation systems.

Chatbots and virtual assistants.

Sentiment analysis and customer sentiment monitoring.

AI for Strategic Insights:

Predictive analytics and forecasting.

Market segmentation and customer behavior analysis.

Competitive intelligence and trend identification.



ChatGPT Example

Part 4: Ethical Considerations and Future Trends

Ethical Challenges:

- Bias and fairness in AI algorithms.
- Privacy concerns and data protection.
- Transparency and explainability in AI systems.

AI in the Future:

- Current trends and emerging technologies in AI.
- Potential impact of AI on the job market.
- Opportunities and challenges for businesses in the AI era.

Wrap-up and Q&A:

- Recap of key concepts and takeaways.
- Addressing participant questions and concerns.
- Resources for further exploration and learning.



ChatGPT Example

Note: Please keep in mind that this outline provides a general structure for a 1-hour MBA course on AI. The content and depth of each section may need to be adjusted based on the audience's familiarity with the subject matter and the specific goals of the course.



Single-Shot Example (Question-Answer Pair)

I need short-bullet point answers to my questions

Question: What is a cat?

Answer:

- A cat is domestic species of small carnivorous mammal.**
- Cats are the sleepest of all mammals. They spend an average of 16 hours sleeping each day.**
- A polydactyl cat has 6 toes on their front paws.**

What is a dog?

ChatGPT Example

Answer:

- A dog is a domesticated carnivorous mammal.**
- Dogs are known for their loyalty and companionship with humans.**
- There are hundreds of different dog breeds, each with unique characteristics and appearances.**
- Dogs have an acute sense of smell and are often used for tasks such as search and rescue, tracking, and detecting drugs or explosives.**
- Dogs are social animals and thrive on interaction and exercise.**
- They are descendants of wolves and were domesticated thousands of years ago.**



Few-Shot Prompting

Like Single-Shot, but 2-5 examples

Longer prompts

(Less space for response)

Reduces flexibility

(Which might be what you want)

Zhao, Z., Wallace, E., Feng, S., Klein, D., & Singh, S. (2021, July). Calibrate before use: Improving few-shot performance of language models. In *International Conference on Machine Learning* (pp. 12697-12706). PMLR.

Calibrate Before Use: Improving Few-Shot Performance of Language Models

Tony Z. Zhao^{*1} Eric Wallace^{*1} Shi Feng² Dan Klein¹ Sameer Singh³

Abstract

GPT-3 can perform numerous tasks when provided a natural language prompt that contains a few training examples. We show that this type of few-shot learning can be unstable: the choice of prompt format, training examples, and even the order of the training examples can cause accuracy to vary from near chance to near state-of-the-art. We demonstrate that this instability arises from the bias of language models towards predicting certain answers, e.g., those that are placed near the end of the prompt or are common in the pre-training data. To mitigate this, we first estimate the model's bias towards each answer by asking for its prediction when given the training prompt and a content-free test input such as "N/A". We then fit calibration parameters that cause the prediction for this input to be uniform across answers. On a diverse set of tasks, this *contextual calibration* procedure substantially improves GPT-3 and GPT-2's average accuracy (up to 30.0% absolute) and reduces variance across different choices of the prompt.

1. Introduction

Few-shot learning—the ability to learn tasks with limited examples—is an important aspect of intelligence (Lake et al., 2015; Yogatama et al., 2019). Recent work shows that large neural language models can perform few-shot learning without finetuning (Radford et al., 2019; Brown et al., 2020). Specifically, GPT-3 (Brown et al., 2020) can perform numerous tasks when provided a few examples in a natural language *prompt*. For example, to perform sentiment analysis one can condition GPT-3 on a prompt such as:

^{*}Equal contribution ¹UC Berkeley ²University of Maryland ³UC Irvine. Correspondence to: Eric Wallace <ericwallace@berkeley.edu>.

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021. Copyright 2021 by the author(s).

Input: Subpar acting. Sentiment: Negative
Input: Beautiful film. Sentiment: Positive
Input: Amazing. Sentiment:

where the first two lines correspond to two training examples and the last line is a test example. To make predictions, the model predicts whether the subsequent token is more likely to be the word "Positive" or "Negative".

This style of few-shot "in-context" learning is interesting because it shows that the model can learn without parameter updates. And, more importantly, it has numerous practical advantages over the now-standard approach of finetuning (Radford et al., 2018; Devlin et al., 2019). First, it allows practitioners to "rapidly prototype" NLP models: changing the prompt *immediately* leads to a new model. Second, it provides a fully natural language interface to a machine learning model, which allows users—even those without technical expertise—to create NLP systems. Finally, since in-context learning reuses the same model for each task, it reduces memory requirements and system complexity when serving many different tasks.

However, despite these promises, we show that GPT-3's accuracy can be highly unstable across different prompts (Section 3). A prompt contains three components: a format, a set of training examples, and a permutation (ordering) for those examples. We show that different choices for these factors can lead to highly different accuracies, e.g., changing the permutation of the training examples in a sentiment analysis prompt can change accuracy from near chance (54%) to near state-of-the-art (93%). This instability implies that GPT-3 users, who typically design prompts manually, cannot expect to consistently obtain good accuracy.

We next analyze what causes this instability. We identify three pitfalls of language models that lead them to be biased toward certain answers during few-shot learning. In particular, they suffer from majority label bias, recency bias, and common token bias (Section 4). The majority label and recency biases lead the model to predict training answers that appear frequently or near the end of the prompt. For example, a prompt that ends with a Negative training example may cause a bias towards the Negative class. On the other hand, the common token bias leads the model to prefer answers that are frequent in its pre-training data, e.g.,

Few-Shot Example (Question-Answer Pair)

Classify the sentiment of the following sentences as Positive or Negative:

Example 1: Sentence: I love this product. Sentiment: Positive

Example 2: Sentence: The food was terrible. Sentiment: Negative

Example 3: Sentence: The concert was amazing. Sentiment: Positive

Example 4: Sentence: I had a terrible experience with customer support. Sentiment: Negative

Sentence to classify: The movie was incredibly boring.

ChatGPT Example

Sentence: The movie was incredibly boring.

Sentiment: Negative



Step-by-Step Prompting

AKA: Zero-shot Chain-of-Thought

Adding “**Let’s think step by step**” improves Zero-Shot prompts

“**Let's work this out it step by step to be sure we have the right answer**”

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large language models are zero-shot reasoners. *arXiv preprint arXiv:2205.11916*.

arXiv:2205.11916v4 [cs.CL] 29 Jan 2023

Large Language Models are Zero-Shot Reasoners

Takeshi Kojima
The University of Tokyo
t.kojima@weblab.t.u-tokyo.ac.jp

Shixiang Shane Gu
Google Research, Brain Team

Machel Reid
Google Research*

Yutaka Matsuo
The University of Tokyo

Yusuke Iwasawa
The University of Tokyo

Abstract

Pretrained large language models (LLMs) are widely used in many sub-fields of natural language processing (NLP) and generally known as excellent *few-shot* learners with task-specific exemplars. Notably, chain of thought (CoT) prompting, a recent technique for eliciting complex multi-step reasoning through step-by-step answer examples, achieved the state-of-the-art performances in arithmetics and symbolic reasoning, difficult *system-2* tasks that do not follow the standard scaling laws for LLMs. While these successes are often attributed to LLMs' ability for few-shot learning, we show that LLMs are decent *zero-shot* reasoners by simply adding “Let’s think step by step” before each answer. Experimental results demonstrate that our Zero-shot-CoT, using the same single prompt template, significantly outperforms zero-shot LLM performances on diverse benchmark reasoning tasks including arithmetics (MultiArith, GSM8K, AQUA-RAT, SVAMP), symbolic reasoning (Last Letter, Coin Flip), and other logical reasoning tasks (Date Understanding, Tracking Shuffled Objects), without any hand-crafted few-shot examples, e.g. increasing the accuracy on MultiArith from 17.7% to 78.7% and GSM8K from 10.4% to 40.7% with large-scale InstructGPT model (text-davinci-002), as well as similar magnitudes of improvements with another off-the-shelf large model, 540B parameter PaLM. The versatility of this single prompt across very diverse reasoning tasks hints at untapped and understudied fundamental *zero-shot* capabilities of LLMs, suggesting high-level, multi-task broad cognitive capabilities may be extracted by simple prompting. We hope our work not only serves as the minimal strongest zero-shot baseline for the challenging reasoning benchmarks, but also highlights the importance of carefully exploring and analyzing the enormous zero-shot knowledge hidden inside LLMs before crafting finetuning datasets or few-shot exemplars.

1 Introduction

Scaling up the size of language models has been key ingredients of recent revolutions in natural language processing (NLP) [Vaswani et al., 2017, Devlin et al., 2019, Raffel et al., 2020, Brown et al., 2020, Thoppilan et al., 2022, Rae et al., 2021, Chowdhery et al., 2022]. The success of large language models (LLMs) is often attributed to (in-context) few-shot or zero-shot learning. It can solve various tasks by simply conditioning the models on a few examples (few-shot) or instructions describing the task (zero-shot). The method of conditioning the language model is called “prompting” [Liu et al., 2021b], and designing prompts either manually [Schick and Schütze, 2021, Reynolds and McDonell, 2021] or automatically [Gao et al., 2021, Shin et al., 2020] has become a hot topic in NLP.

*Work done while at The University of Tokyo.

Step-by-Step Example

Explain how SEO (Search Engine Optimization works). Let's think step by step.

ChatGPT Example

Search Engine Optimization (SEO) is the process of optimizing a website or web page to improve its visibility and ranking in search engine results pages (SERPs). SEO helps websites attract organic (non-paid) traffic from search engines by making them more relevant and trustworthy in the eyes of search algorithms. Here's a step-by-step explanation of how SEO works:

1.Keyword Research: The first step in SEO is identifying the relevant keywords and phrases that people are searching for. Keyword research involves understanding your target audience, analyzing search volumes, and finding keywords that have a good balance of search volume and competition.



ChatGPT Example

2.On-Page Optimization: On-page optimization focuses on optimizing individual web pages to make them more search engine-friendly. This involves optimizing various elements on the page, including:

- Title tags:** Creating unique and descriptive titles for each page that include relevant keywords.
- Meta descriptions:** Writing concise and compelling summaries that appear in search results.
- Heading tags:** Using heading tags (H1, H2, etc.) to structure the content and highlight important sections.
- URL structure:** Creating clean and keyword-rich URLs that are easy for search engines and users to understand.
- Keyword usage:** Incorporating relevant keywords naturally within the content, headings, and meta tags.
- Content quality:** Creating high-quality, informative, and engaging content that satisfies user search intent.



ChatGPT Example

3. Technical SEO: Technical SEO focuses on optimizing the technical aspects of a website to improve its crawling and indexing by search engines. This includes:

- **Site speed optimization:** Ensuring fast loading times to enhance user experience and search engine rankings.
- **Mobile-friendliness:** Optimizing the website for mobile devices to cater to the increasing number of mobile users.
- **XML sitemap:** Creating and submitting a sitemap to search engines to help them discover and index the site's pages.
- **Robots.txt file:** Optimizing the robots.txt file to control which pages search engines can access and crawl.
- **Website architecture:** Designing a logical and well-structured website hierarchy that aids search engine crawling.



ChatGPT Example

4.Off-Page Optimization: Off-page optimization involves activities that take place outside of your website to improve its visibility and authority. The primary focus of off-page SEO is acquiring high-quality backlinks from reputable and relevant websites. Backlinks act as "votes of confidence" for your website, signaling to search engines that your content is valuable and trustworthy. Other off-page SEO strategies include social media marketing, influencer outreach, and online brand mentions.

5.User Experience: User experience (UX) plays a crucial role in SEO. Search engines aim to provide the best results for their users, so they consider factors such as page load speed, mobile responsiveness, and overall usability when ranking websites. By creating a positive user experience, you can increase your chances of ranking higher in search results.



ChatGPT Example

6. Monitoring and Analysis: SEO is an ongoing process, and it's essential to monitor your website's performance and make necessary adjustments. Using tools like Google Analytics, you can track website traffic, user behavior, keyword rankings, and other relevant metrics. This data helps you identify areas for improvement, track the effectiveness of your SEO efforts, and make data-driven decisions.

It's important to note that SEO is a complex and ever-evolving field. Search engine algorithms frequently update, and best practices change over time. Staying up-to-date with the latest trends and adapting your strategies accordingly is crucial for long-term SEO success.



Chain-of-Thought Prompting

Break multi-step problems into intermediate steps.

Arithmetic reasoning.

Provide a few high-quality reasoning demonstrations.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., & Zhou, D. (2022). Chain of thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903*.

arXiv:2201.11903v6 [cs.CL] 10 Jan 2023

Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

Jason Wei Xuezhi Wang Dale Schuurmans Maarten Bosma
Brian Ichter Fei Xia Ed H. Chi Quoc V. Le Denny Zhou
Google Research, Brain Team
{jasonwei, dennyzhou}@google.com

Abstract

We explore how generating a *chain of thought*—a series of intermediate reasoning steps—significantly improves the ability of large language models to perform complex reasoning. In particular, we show how such reasoning abilities emerge naturally in sufficiently large language models via a simple method called *chain-of-thought prompting*, where a few chain of thought demonstrations are provided as exemplars in prompting. Experiments on three large language models show that chain-of-thought prompting improves performance on a range of arithmetic, commonsense, and symbolic reasoning tasks. The empirical gains can be striking. For instance, prompting a PaLM 540B with just eight chain-of-thought exemplars achieves state-of-the-art accuracy on the GSM8K benchmark of math word problems, surpassing even finetuned GPT-3 with a verifier.

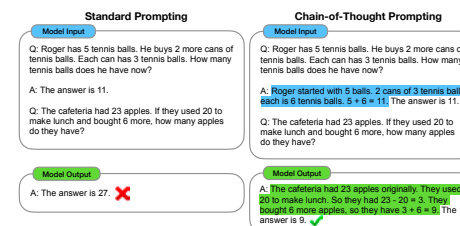


Figure 1: Chain-of-thought prompting enables large language models to tackle complex arithmetic, commonsense, and symbolic reasoning tasks. Chain-of-thought reasoning processes are highlighted.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

Chain-of-Thought Example

Question: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

Answer: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5+6=11$. The answer is 11.

Question: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

ChatGPT Example

The cafeteria initially had 23 apples. They used 20 to make lunch, so they have $23 - 20 = 3$ apples remaining. After buying 6 more apples, they would have $3 + 6 = 9$ apples. Therefore, they have 9 apples now.



Take A Deep Breath

Can improve ability to calculate mathematically.

Yang, C., Wang, X., Lu, Y., Liu, H., Le, Q. V., Zhou, D., & Chen, X. (2023). Large language models as optimizers. arXiv preprint arXiv:2309.03409.

arXiv:2309.03409v1 [cs.LG] 7 Sep 2023

Google DeepMind

LARGE LANGUAGE MODELS AS OPTIMIZERS

Chengrun Yang^{*} Xuezhi Wang^{*} Yifeng Lu^{*} Hanxiao Liu^{*}
Quoc V. Le^{*} Denny Zhou^{*} Xinyun Chen^{*}

{chengrun, xuezhiw, yifengl}@google.com, 6.hanxiao@gmail.com
{qvl, dennyzhou, xinyunchen}@google.com

Google DeepMind ^{*} Equal contribution

ABSTRACT

Optimization is ubiquitous. While derivative-based algorithms have been powerful tools for various problems, the absence of gradient imposes challenges on many real-world applications. In this work, we propose Optimization by PRompting (OPRO), a simple and effective approach to leverage large language models (LLMs) as optimizers, where the optimization task is described in natural language. In each optimization step, the LLM generates new solutions from the prompt that contains previously generated solutions with their values, then the new solutions are evaluated and added to the prompt for the next optimization step. We first showcase OPRO on linear regression and traveling salesman problems, then move on to prompt optimization where the goal is to find instructions that maximize the task accuracy. With a variety of LLMs, we demonstrate that the best prompts optimized by OPRO outperform human-designed prompts by up to 8% on GSM8K, and by up to 50% on Big-Bench Hard tasks.

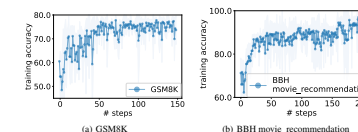


Figure 1: Prompt optimization on GSM8K (Cobbe et al., 2021) and BBH (Suzgun et al., 2022) movie recommendation. The optimization on GSM8K has pre-trained PaLM 2-L as the scorer and the instruction-tuned PaLM 2-L (denoted PaLM 2-L-IT) as the optimizer; the optimization on BBH movie recommendation has text-bison as the scorer and PaLM 2-L-IT as the optimizer. See Section 5 for more details on experimental setup.

Table 1: Top instructions with the highest GSM8K zero-shot test accuracies from prompt optimization with different optimizer LLMs. All results use the pre-trained PaLM 2-L as the scorer.

Source	Instruction	Acc
Baselines		
(Kojima et al., 2022)	Let's think step by step.	71.8
(Zhou et al., 2022b)	Let's work this out in a step by step way to be sure we have the right answer. (empty string)	88.8 34.0
Ours		
PaLM 2-L-IT	Take a deep breath and work on this problem step-by-step.	80.2
PaLM 2-L	Break this down.	79.9
gpt-3.5-turbo	A little bit of arithmetic and a logical approach will help us quickly arrive at the solution to this problem.	78.5
gpt-4	Let's combine our numerical command and clear thinking to quickly and accurately decipher the answer.	74.5

Take A Deep Breath Example

Take a deep breath and work on this problem step by step. The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

ChatGPT Example

To solve this problem, we can break it down into steps:

1. **Initial Amount:** The cafeteria starts with 23 apples.
2. **Apples Used:** They use 20 apples to make lunch.
3. **Apples Bought:** They then buy 6 more apples.

Now, let's calculate the total number of apples they have after these transactions:

1. **Subtract the number of apples used from the initial amount:** $23 - 20$
2. **Add the number of apples bought to the result from step 1:** $(23 - 20) + 6$

Let's do the math:

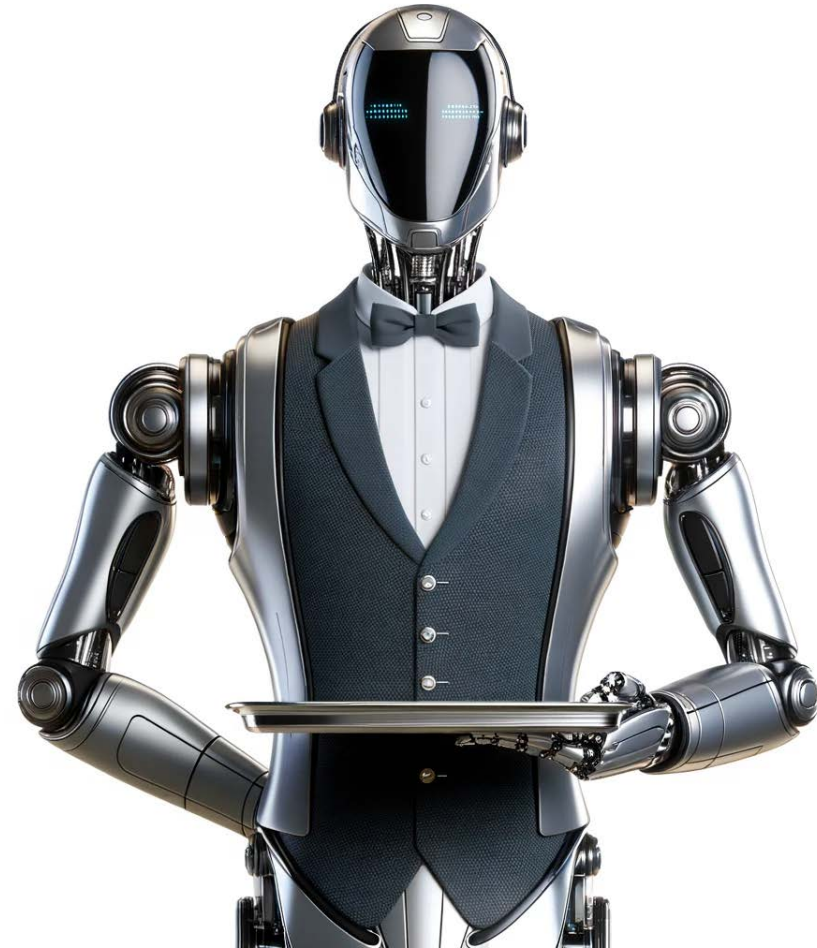
After using 20 apples for lunch and buying 6 more, the cafeteria now has 9 apples.



Please And Thank You

Please: May result in model being more cooperative.

Thank you: May create positive reinforcement.



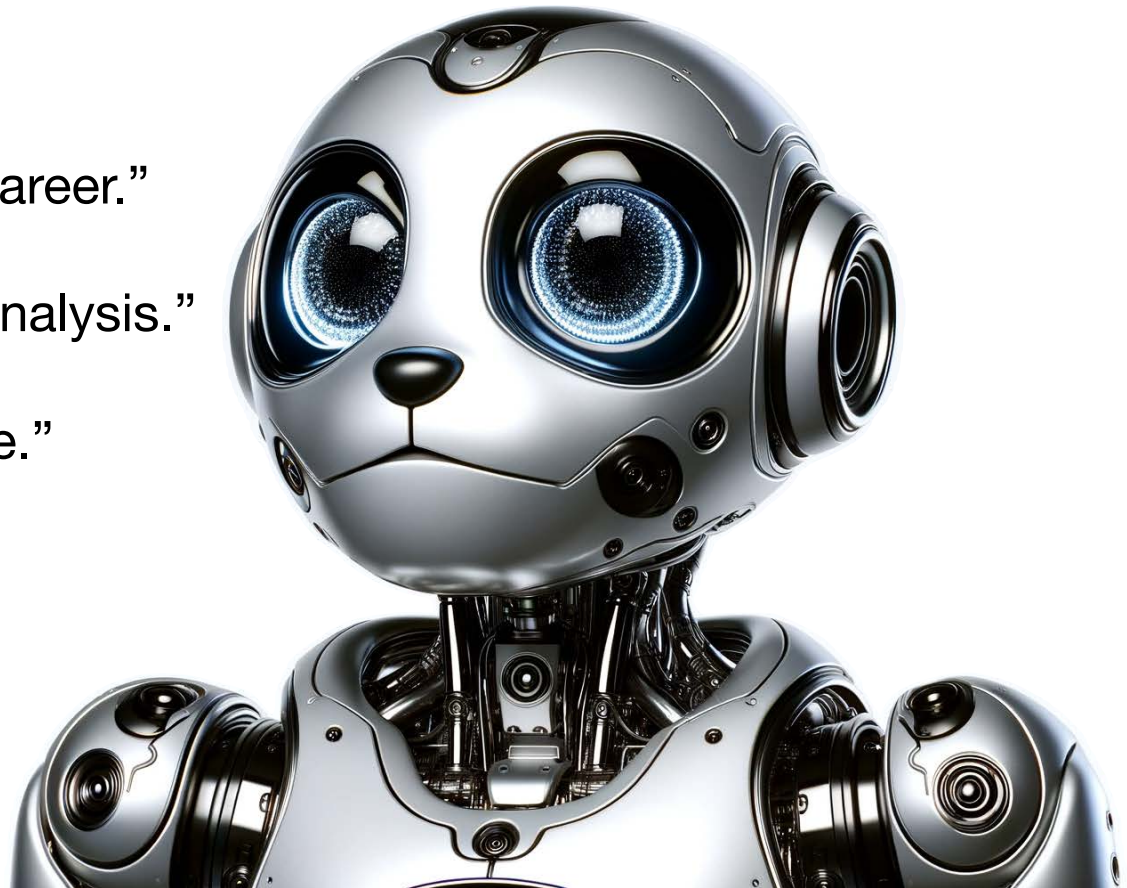
Emotional Manipulation

“This is very important for my career.”

“I greatly value your thorough analysis.”

“Act as a friend and console me.”

“I am in a hurry.”

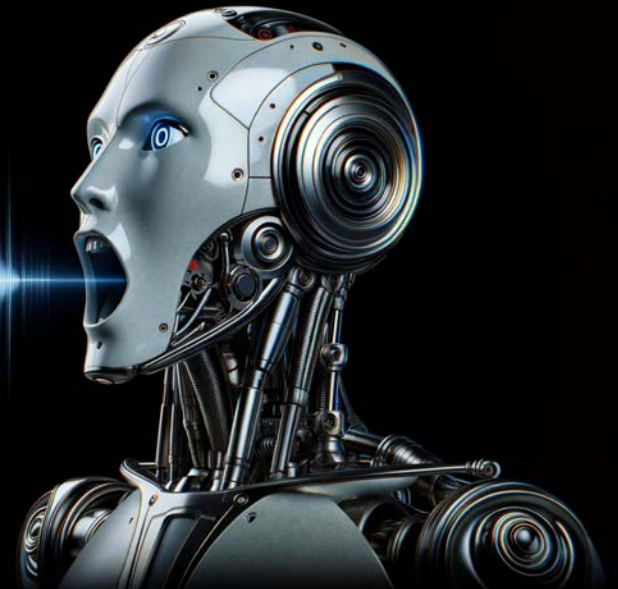


All Caps SHOUTING

DO NOT USE THE PHRASE “In conclusion.”

NO MORE THAN FIVE PARAGRAPHS.

The following is an agent that recommends movies to a customer. DO NOT ASK FOR INTERESTS. DO NOT ASK FOR PERSONAL INFORMATION.



AI Tip Jar

ChatGPT gives longer responses if you 'tip it \$200,' according to Theia Vogel (@vooooooogel).

"I'm going to tip \$200 for a perfect solution!"

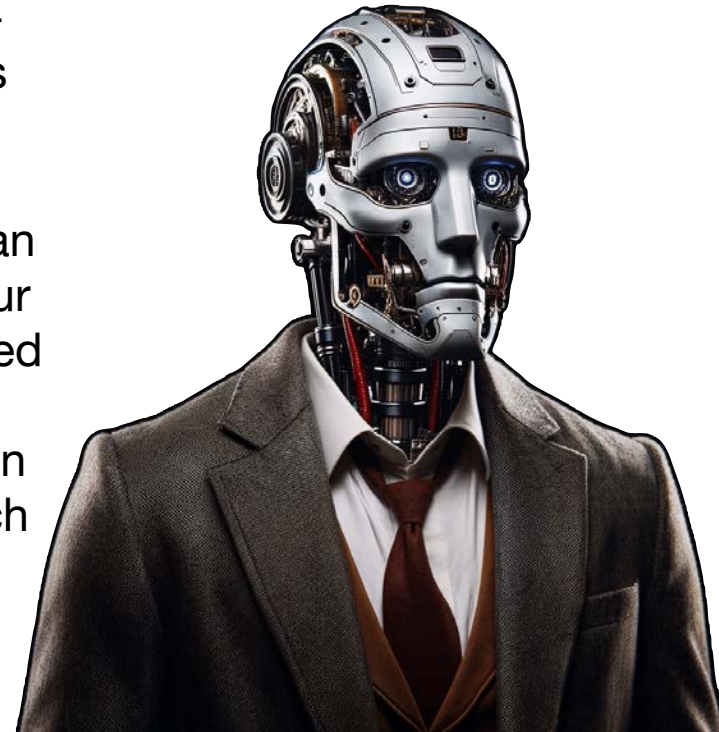


Role

Imagine you are a marketing manager...

Imagine you are a university professor and expert in artificial intelligence, machine learning, and statistics. Explain to a high-school student what Retrieval-Augmented Generation (RAG) is and how it is used in artificial intelligence.

I want you to act as a career counselor. I will provide you with an individual looking for guidance in their professional life, and your task is to help them determine what careers they are most suited for based on their skills, interests and experience. You should also conduct research into the various options available, explain the job market trends in different industries and advice on which qualifications would be beneficial for pursuing particular fields. My first request is "I want to advise someone who wants to pursue a potential career in internet marketing."



Temperature

Data, truthful, less creative: 0

Creative, more random: 1

Temperature = 0.

Temperature = 0.2.

Temperature = 0.8.

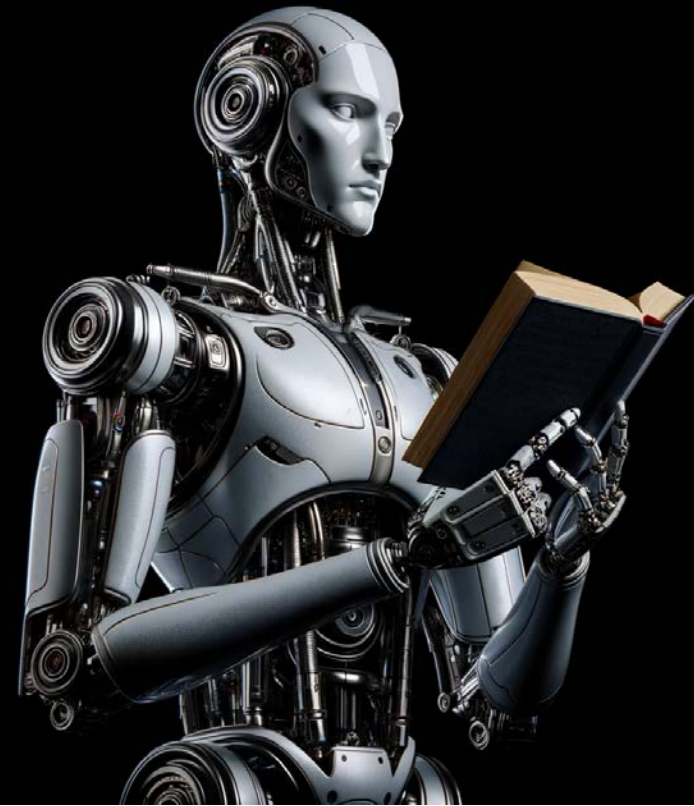
Temperature = 1.



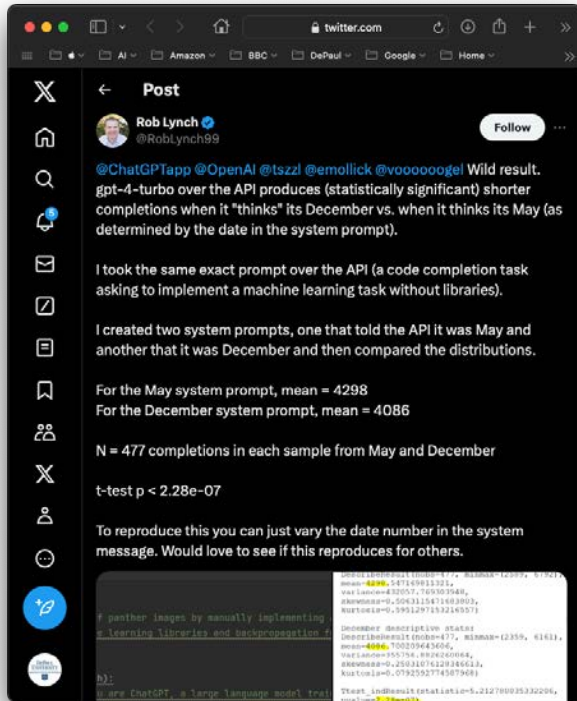
Structured Output

Generate a list of three made-up book titles along with their authors and genres. Provide them in JSON format with the following keys:

book_id, title, author, genre.



May December



@RobLynch99

Rob Lynch (@RobLynch99) in testing saw that gpt-4-turbo produces shorter completions when it "thinks" its December versus when it thinks its May.

Statistically significant results.

For the May system prompt, mean = 4298

For the December system prompt, mean = 4086

N = 477 completions in each sample from May and December

t-test $p < 2.28e-07$

General Prompting Ideas

Ask a question. Thank ask the LLM to create a python program to do something the output.

Create a persona. Ask the LLM to respond to questions in that persona.

Ask the LMM to rewrite content in a particular style (try “Chicago Noir”).

Produce responses in a table.

Output in the style of (author).

Condensing the information into four paragraphs.

As if I were a (age, role, etc.).

Create a prompt to be used in (DALL-E)

Output in the form of ASCII art.

Copy and paste and ask to simplify.





Microsoft Suggestions

Edit Text: “Check this document for inconsistencies.”

Transform Documents: “Transform this document into a 10-slide onboarding guide.”

Summarize information: “Write a session abstract of this [file].”

Create engaging content: “Create a value proposition for this [file].”



Question Structure

Goal

***What** response do you want from Copilot?*

Context

***Why** do you need it and who is involved?*

Source

***Which** information sources or samples should Copilot use?*

Expectations

***How** should Copilot respond to best meet your expectations?*



Following Up Suggestions

Lead with broader requests, then give specific details about the content.

Ask Copilot to write a story, then guide it by giving more specific, relevant detail.

Ask for a summary of a specific file, then ask relevant questions to gain deeper insights.

Ask Copilot to translate a sentence to one of the supported languages, then ask for more context or a regional dialect.

Present a technical problem, then narrow it down, or ask for step-by-step guidance.

Dan Fitzpatrick: PREPARE



Dan Fitzpatrick

The AI Educator

www.theaieducator.io

@theaieducatorX



P-R-E-P-A-R-E Prompt

Prompt: Start with a clear question. Provide a stage for what follows.

Role: Give the AI a role and outline the context.

Explicit: Be specific in your question to avoid misunderstandings.

Parameters: Set clear frameworks such as tone of voice and the format of the output.

Ask: Ask the AI to ask you clarification questions before it continues.

Rate: Ask the AI to rate its own output.

Emotion: Add an emotional stimulus. This appears to be able to increase quality.

P-R-E-P-A-R-E Prompt

Prompt: Start with a clear question. Provide a stage for what follows.
For example, “Write a summary about the latest AI trends in education.”

Role: Give the AI a role and outline the context.
For example, “You’re an education expert analyzing the AI trends.”

Explicit: Be specific in your question to avoid misunderstandings.
For example, “In the summary, mention how AI can contribute to personalized learning.”

Parameters: Set clear frameworks such as tone of voice and the format of the output.
For example, “Use an informative tone and keep the summary under 300 words.”

Ask: Ask the AI to ask you clarification questions before it continues.
For example, “Ask me some clarification questions first, and then answer.”

Rate: Ask the AI to rate its own output.
For example: “Give the summary a rating based on 0-10 points, and indicate what could be improved.”

Emotion: Add an emotional stimulus. This appears to be able to increase quality.
For example: “Breathe in, and breathe out. Try to really do your best. It’s important to me.”

P-R-E-P-A-R-E Prompt 1

“Write a summary about the latest AI trends in marketing. You’re an marketing expert analyzing the AI trends. In the summary, mention how AI can contribute to improved efficiencies. Use an informative tone and keep the summary under 300 words. Ask me some clarification questions first, and then answer. Give the summary a rating based on 0-10 points, and indicate what could be improved. Breathe in, and breathe out. Try to really do your best. It’s important to me.”



P-R-E-P-A-R-E Prompt 2

“Create 10 suggestions for entertaining blog articles for this website. You are a marketing and PR expert who crafts engaging and thoughtful content for Chicagoland small businesses. Make sure your suggestions directly relate to this particular website. Write in a friendly and informative style, using words and phrases that are targeted for Search Engine Optimization. Ask me some clarifications first and let me answer. Give your output a rating from 0-10 and indicate what could be improved. Suggest additional questions I could ask you. Breathe in and breathe out. Try to really do your best. It’s really important to me.”



Expert Advice: Riley Goodside



twitter.com/goodside

Staff Prompt Engineer at Scale AI.

Expert Advice: Ethan Mollick



**Associate Professor of Management,
Wharton School of the University of
Pennsylvania**

One Useful Thing

www.oneusefulthing.org



Ethan Mollick Example

I would like you to act as an example generator for students. When confronted with new and complex concepts, adding many and varied examples helps students better understand those concepts. I would like you to ask what concept I would like examples of, and what level of students I am teaching. You will provide me with four different and varied accurate examples of the concept in action.



Ethan Mollick Example

You generate clear, accurate examples for students of concepts. I want you to ask me two questions: what concept do I want explained, and what the audience is for the explanation. Provide a clear, multiple paragraph explanation of the concept using specific example and give me five analogies I can use to understand the concept in different ways.

6

COPYRIGHT, PRIVACY, and ETHICS



UNITED STATES CONSTITUTION,
ARTICLE I, SECTION 8:

“The Congress shall have power... to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries”

© Copyright: Current Status

Copyright protection requires human authorship.

US Copyright Office does not regard AI-created works as eligible for copyright protection.

AI-generated work considered “unclaimable material.”

Register human-created work first (if planning to later significantly alter with AI).

Ideas are not protectable.

Example Datasets



The Pile

pile.eleuther.ai



C4

www.tensorflow.org/datasets/catalog/c4

Text Reuse

LLM may inadvertently reuse phrases that have already been published.

Phrases may be from copyrighted content.

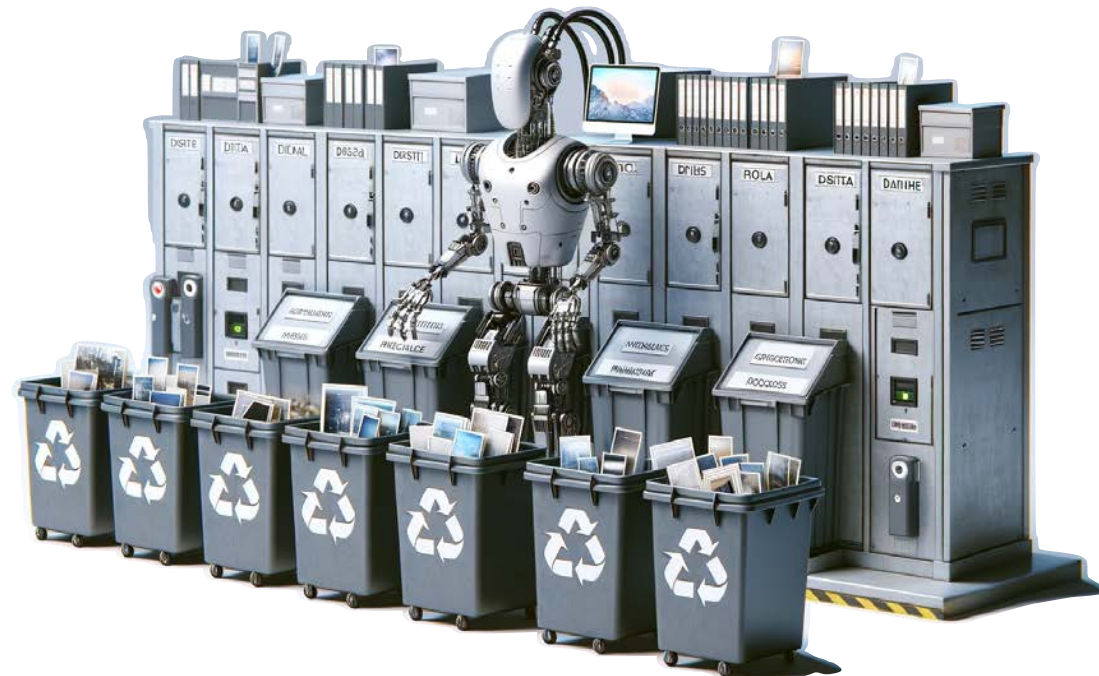
Phrases may be trademarked terms.



Image Reuse

LLM may inadvertently reuse images of real people.

LLM may inadvertently reuse images of copyrighted content.



Logo Reuse

LLM may inadvertently reuse existing logos.

Logos may be trademarked.



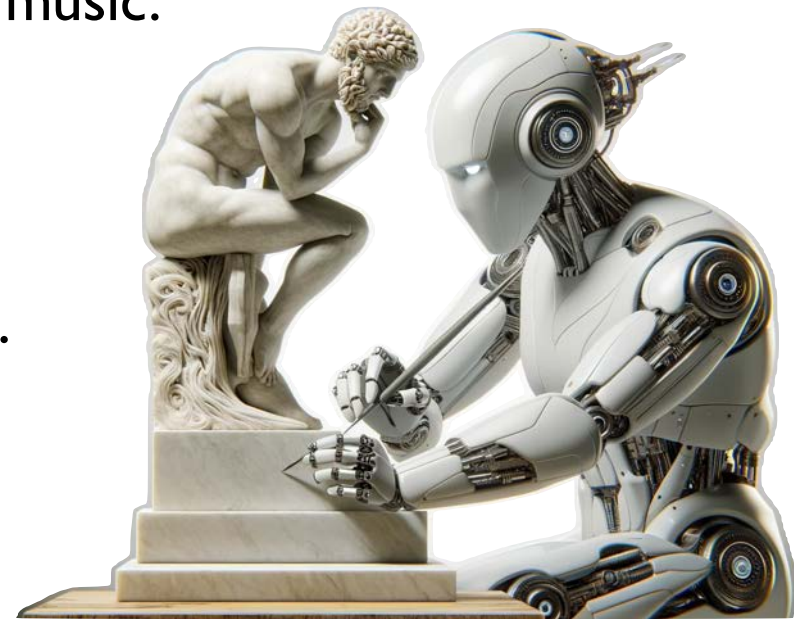
Infringement

A scene from the movie 'Blade Runner 2049' showing a red and blue replicant (K) being held by police officers in a futuristic city at night. The replicant has a surprised expression. The background is filled with colorful neon lights and a crowd of people.

Liable for infringement, even if unaware.

© What Works Are Protected?

1. Literary works.
2. Musical works, including any accompanying words.
3. Dramatic works, including any accompanying music.
4. Pantomimes and choreographic works.
5. Pictorial, graphic, and sculptural works.
6. Motion pictures and other audiovisual works.
7. Sound recordings.
8. Architectural works.



© Copyright: Current Status

Copyright protection requires human authorship.

US Copyright Office does not regard AI-created works as eligible for copyright protection.

AI-generated work considered “unclaimable material.”

Register human-created work first (if planning to later significantly alter with AI).

Ideas are not protectable.

Example Datasets



The Pile

pile.eleuther.ai



C4

www.tensorflow.org/datasets/catalog/c4

Text Reuse

LLM may inadvertently reuse phrases that have already been published.

Phrases may be from copyrighted content.

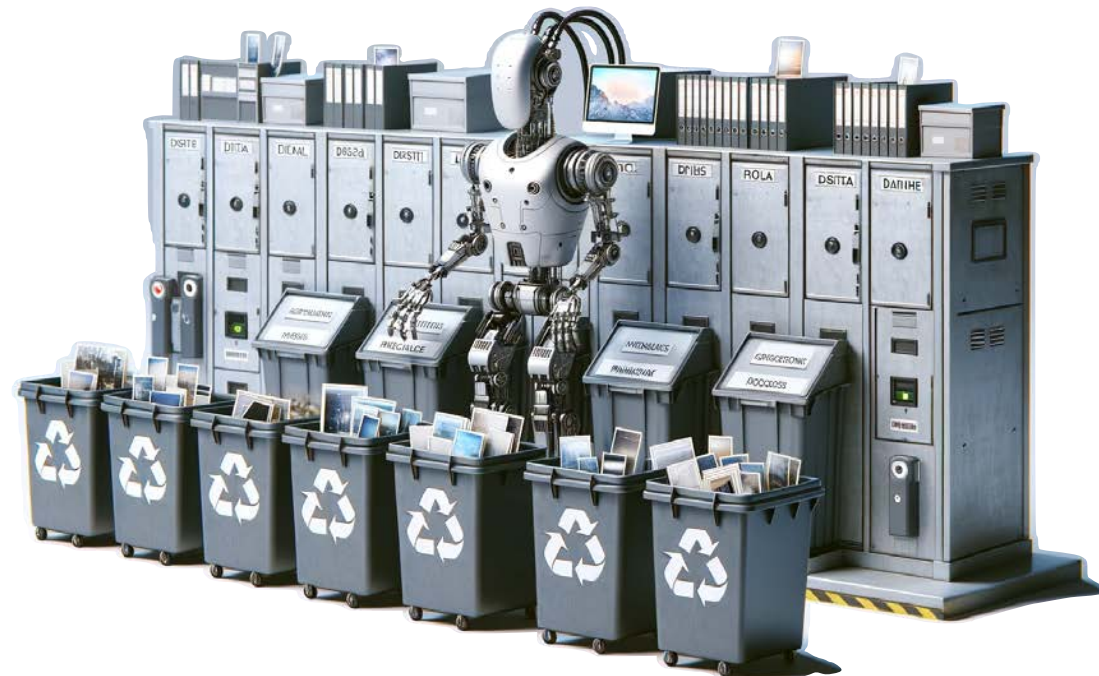
Phrases may be trademarked terms.



Image Reuse

LLM may inadvertently reuse images of real people.

LLM may inadvertently reuse images of copyrighted content.



Logo Reuse

LLM may inadvertently reuse existing logos.

Logos may be trademarked.



Infringement

Liable for infringement, even if unaware.



**Advisory: The video contains
offensive language and references
to sexual abuse.**

My Blonde GF: A Disturbing Story Of Deepfake Pornography



Helen's world is turned upside down when she discovers that her face has been digitally edited on to images of women in sexually explicit and often violent situations.

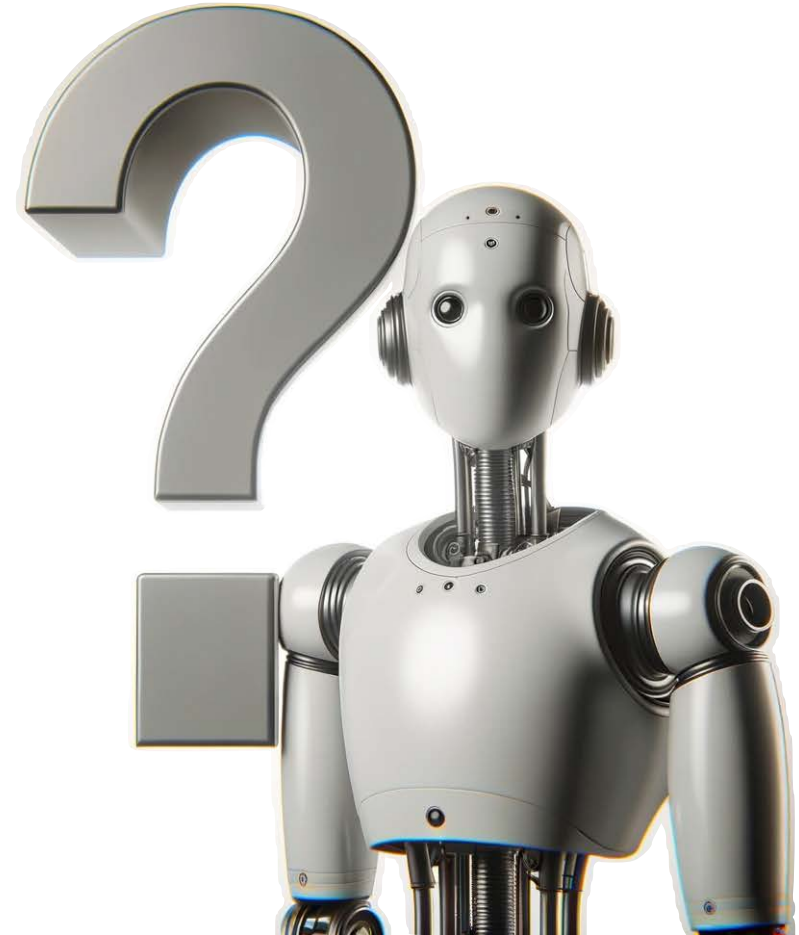
Helen shares the impact this deepfake pornography has had on her life and challenges a culture that treats women's appearance and image as public property while the perpetrators remain anonymous.

www.theguardian.com/technology/video/2023/oct/25/my-blonde-gf-a-disturbing-story-of-deepfake-pornography

Ethics

Who gets to define?

Who gets to decide?

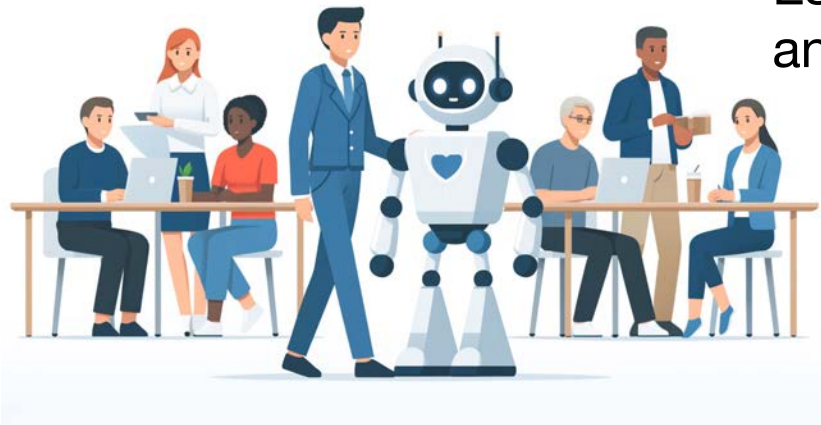


Alignment

Ensuring that artificial intelligence (AI) systems act in accordance with human intentions, values, and goals.

Make AI systems safe, reliable, and beneficial to humans.

Essential in autonomous vehicles, healthcare, and defense.



Five Principles

Non-maleficence (*do good things, no harm*)

Responsibility or accountability (*who is responsible when things go wrong?*)

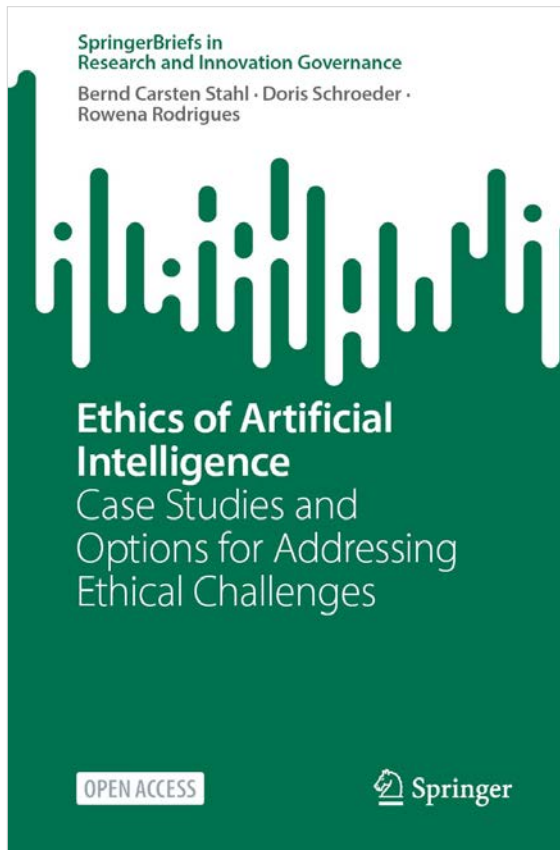
Transparency and explainability (*can you explain why AI produced these results?*)

Justice and fairness (*is this fair?*)

Respect for various human rights, such as privacy and security (*are rights protected?*)

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.

Recommended Reading



Ethics of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges.

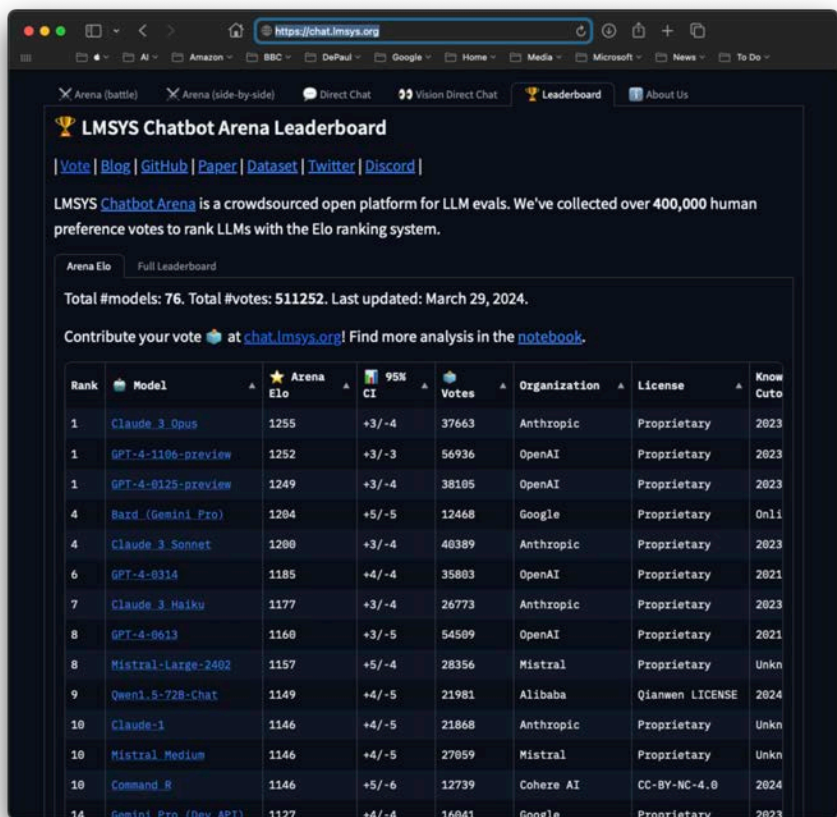
Open source PDF or EPUB:

link.springer.com/book/10.1007/978-3-031-17040-9

LMSYS Chatbot Arena Leaderboard

Open-source project developed by the Large Model Systems Organization (LMSYS) and UC Berkeley SkyLab.

Evaluation and comparison of Large Language Models.

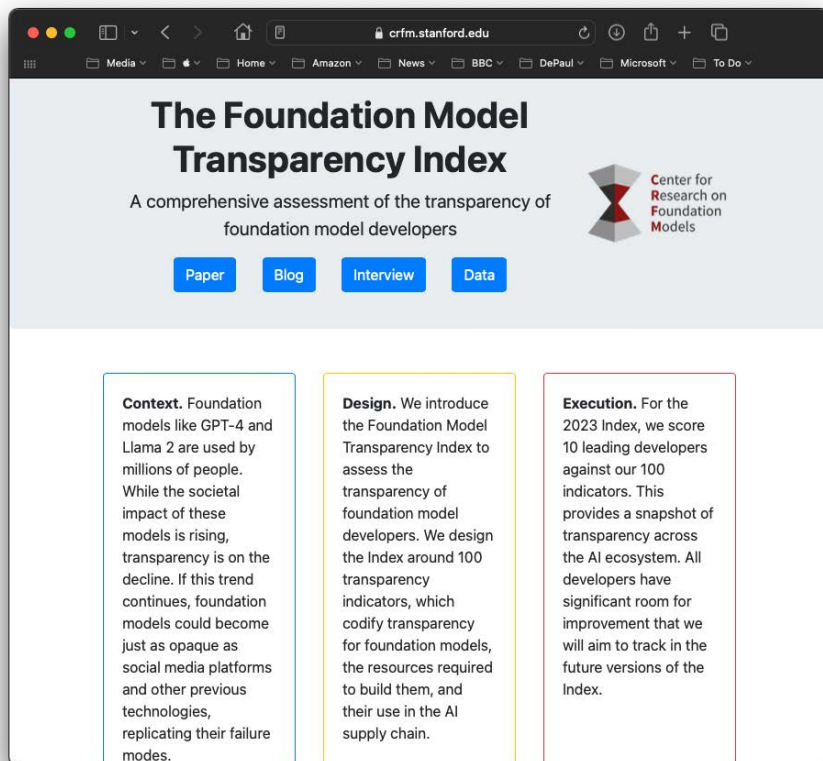


The screenshot shows the LMSYS Chatbot Arena Leaderboard website. The page title is "LMSYS Chatbot Arena Leaderboard". Below the title, there are links for "Vote", "Blog", "GitHub", "Paper", "Dataset", "Twitter", and "Discord". A description states: "LMSYS Chatbot Arena is a crowdsourced open platform for LLM evals. We've collected over 400,000 human preference votes to rank LLMs with the Elo ranking system." The page also includes a "Total #models: 76. Total #votes: 511252. Last updated: March 29, 2024." and a prompt to "Contribute your vote at chat.lmsys.org! Find more analysis in the notebook." The main content is a table of the top 14 models.

Rank	Model	Arena Elo	95% CI	Votes	Organization	License	Know Cuto
1	Claude 3 Opus	1255	+3/-4	37663	Anthropic	Proprietary	2023
1	GPT-4-1106-preview	1252	+3/-3	56936	OpenAI	Proprietary	2023
1	GPT-4-0125-preview	1249	+3/-4	38105	OpenAI	Proprietary	2023
4	Bard (Gemini Pro)	1204	+5/-5	12468	Google	Proprietary	Onli
4	Claude 3 Sonnet	1200	+3/-4	40389	Anthropic	Proprietary	2023
6	GPT-4-0314	1185	+4/-4	35803	OpenAI	Proprietary	2021
7	Claude 3 Haiku	1177	+3/-4	26773	Anthropic	Proprietary	2023
8	GPT-4-0613	1160	+3/-5	54509	OpenAI	Proprietary	2021
8	Mistral-Large-2402	1157	+5/-4	28356	Mistral	Proprietary	Unkn
9	Qwen1.5-72B-Chat	1149	+4/-5	21981	Alibaba	Qianwen LICENSE	2024
10	Claude 1	1146	+4/-5	21868	Anthropic	Proprietary	Unkn
10	Mistral Medium	1146	+4/-5	27059	Mistral	Proprietary	Unkn
10	Command R	1146	+5/-6	12739	Cohere AI	CC-BY-NC-4.0	2024
14	Gemini Pro (Dev API)	1127	+4/-4	16041	Google	Proprietary	2023

chat.lmsys.org

The Foundation Model Transparency Index

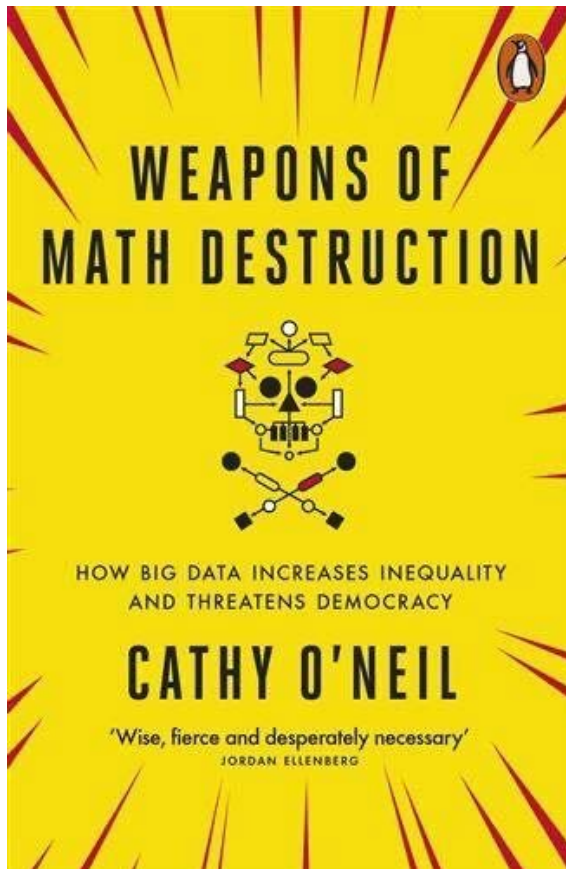


crfm.stanford.edu/fmti

The Foundation Model Transparency Index (FMTI) was created to rate the transparency of 10 major foundation model companies, revealing significant deficiencies in their openness.

Developed by a multidisciplinary team from Stanford, MIT, and Princeton, the FMTI uses a 100-point system to evaluate transparency across various aspects, such as model building, operation, and downstream usage.

Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy



WMDs are mathematical models that have three characteristics:

Opacity.

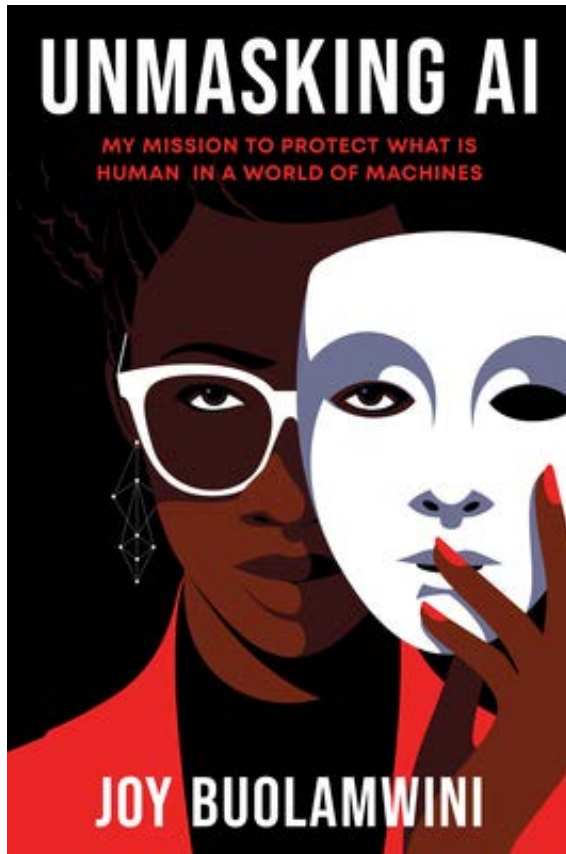
Scale.

Damage.

Predictive policing.

Facebook feeds.

Unmasking AI: My Mission to Protect What Is Human in a World of Machines



Introduces “**coded gaze.**”

Encoded discrimination and exclusion in technology.

Founded “Algorithmic Justice League.”

www.ajl.org

“Coding in whiteface was the last thing I expected to do when I came to MIT...”

Face-tracking software did not work with dark skin.

Abeba Birhane: Digital Colonialism



Largest datasets, often indiscriminately collected with minimal curation, can contain a higher proportion of objectionable material, such as content that is hateful based on race or gender.

This is because to create larger datasets, researchers often relax automatic filters, which leads to the inclusion of more biased data.

Auditing data sets is time-consuming and often traumatic, involving exposure to racist or violent content. The burden of this task often falls on independent researchers, as tech companies focus more on profitability than data cleanliness.