

<https://mco.dev/IntroAI.pdf>

# Understanding Artificial Intelligence

**What a well informed person should know**

**Marc Cohen**

All material in this presentation is ©Marc A. Cohen, 2024, 2025

# Prerequisites

- **You are human (bots not allowed)**
- **You are curious about AI and want to learn more**
- **No technical knowledge required**
- **No programming required**
- **No maths required**

# Goals

- **Learn a bit of AI history**
- **Gain a conceptual understanding of how AI works**
- **Find out about the state of the art in AI**
- **Consider the ethical issues raised by AI**
- **Think critically about AI in your life and in the world**
- **Help you prepare for the continuing revolution**

# Non-Goals

- Deep dive into technology or maths
- Definitely answer difficult ethical or societal questions
- Predict the future

# Agenda

- My Fave AI App
- Definitions
- History
- Mechanics
- State of the Art
- Ethics
- The Future



Photo by Domenico Loia on [Unsplash](#)

# My Fave AI App

# August 2016

## Option 1 - Interactive Trivia Quiz Challenge Game

Inspired by [quizup](#) but a PWA rather than a native app and used as a way to illustrate backend design principles. We'll build this app in steps and highlight important lessons as they arise in the design process.

### Requirements

- Usage flow
  - Visit site ([quizr.io](#))
  - Anonymous mode works out of the box
    - Ability to do quiz challenges with random other users but no user profile, no persistence of results, no leaderboard, no history
  - Establish a user profile
  - Challenge random or selected opponent to real time trivia challenge
  - Challenger selects subject, responder accepts or rejects (repeat until agreement)
  - Quiz conducted in real time using webRTC data (with optional video)
  - Results are persistent
  - Leaderboards maintained
  - Questions are accessed dynamically via network (caching is not particularly helpful because real time contents won't work while offline)
- Utilize an [open trivia question database](#)
- Implemented as a PWA
- Both a hosted service and an open source code example
- FE uses either react or Angular 2 (TBD)
- Use to illustrate state of the art back end capabilities

### App Stages

1. FE - UI starts with anonymous mode only.
2. Add successive features by implementing an API (REST or gRPC). With each step, highlight options and considerations in choosing backend technology
  - a. sign in and user profile management

**March 2023**

**Wouldn't it be  
cool to have a  
free quiz game  
for everyone  
powered by AI?**

The logo consists of the word "quizaic" in a bold, sans-serif font. The letters are primarily white with a light blue shadow, except for the letter "i" which is red. The logo is set against a solid light blue rectangular background.

**quizaic**

(pronounced like **mosaic**)

# Let's play!

The screenshot shows the Quizaic app interface. At the top, there is a navigation bar with the Quizaic logo and the text "AI Powered Infinitrivia". On the right side of the bar is a circular profile picture of a man wearing a green cap. Below the bar is a vertical sidebar containing five menu items: "Home" (selected), "Browse", "Create", "Play", and "About". The main content area features a large blue banner with the "Quizaic" logo in white and red. Below the banner, the text "AI Powered Infinitrivia" is displayed in a large, bold, blue font. A detailed description follows: "Quizaic leverages the power of generative AI to create and play unlimited trivia quizzes and online surveys. Quizaic is a demonstration app to illustrate what's possible with the combination of Google Cloud services, Flutter, and Vertex AI. Quizaic is not an official Google product and should not be used for commercial purposes."

Quizaic

AI Powered Infinitrivia

AI Powered Infinitrivia

Quizaic leverages the power of generative AI to create and play unlimited trivia quizzes and online surveys. Quizaic is a demonstration app to illustrate what's possible with the combination of Google Cloud services, Flutter, and Vertex AI. Quizaic is not an official Google product and should not be used for commercial purposes.

# Early Prompt

Generate a {num\_questions} multiple choice quiz questions based on category {topic}.

Generate a {num\_questions} multiple choice quiz questions **in Swedish** based on category {topic}.

Generate a {num\_questions} multiple choice quiz questions based on category {topic} **in Swedish**.

# Current Prompt

You are a trivia expert.  
Generate a set of multiple choice quiz questions.

Category: {topic}

Quiz difficulty level: {difficulty}

Number of questions: {num\_questions}

Number of responses per question: {num\_answers}

Quiz language: {language}

RULES:

- Accuracy is critical.
- Each question must have exactly one correct response, selected from the responses array.
- Output should be limited to a json array of questions, each of which is an object containing quoted keys "question", "responses", and "correct".
- Don't return anything other than the json document.

OUTPUT:

# Definitions

- **Artificial Intelligence**
- **Machine Learning**
- **Neural Network**
- **Deep Learning**
- **Generative AI**

mal~anto: malantaue mizer  
estis tute en mal~o<sup>2</sup>; elmeti  
mal~on<sup>2</sup>; mal~o min kaptas  
malgrasiĝintan vizaĝon de n  
Rilata al mal~o: fajro de ma  
krio<sup>2</sup>. **mal~igi**. Senigi je ~o  
ne plu konservas ~on: sen~o  
treni sen~an ekzistadon<sup>2</sup>. **se**  
io sen~a.

★**Esperant/o** (Vd esper/o  
Zamenhof okaze de la publikig  
lia lingvoprojekto. **2** (anka  
sengenta, iniciatita de Z., kiu pu

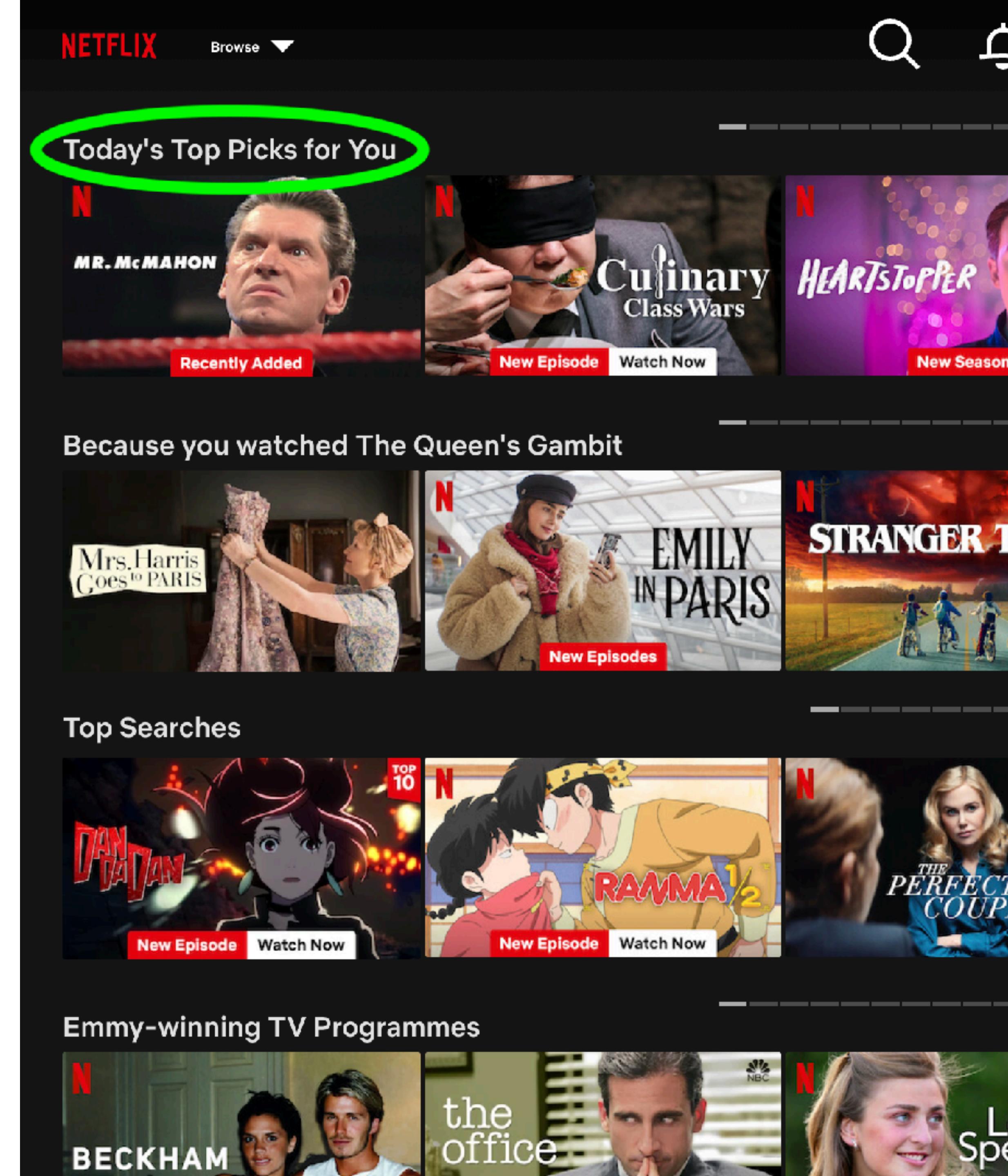
# Definition: Artificial Intelligence

The creation of algorithms and systems capable of performing tasks that typically require human cognition, such as learning, reasoning, perception, decision-making, and natural language processing.



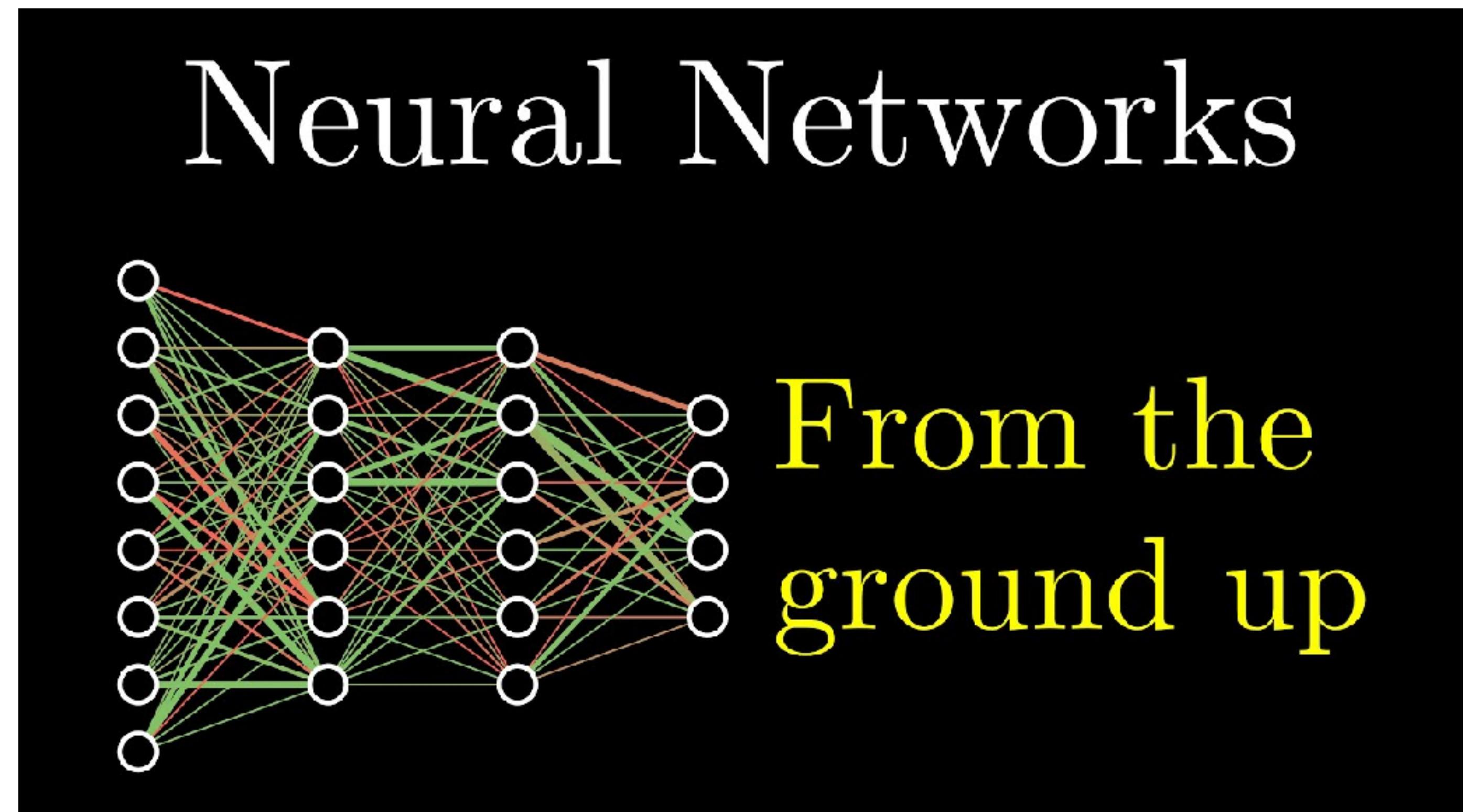
# Definition: Machine Learning

The use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data.



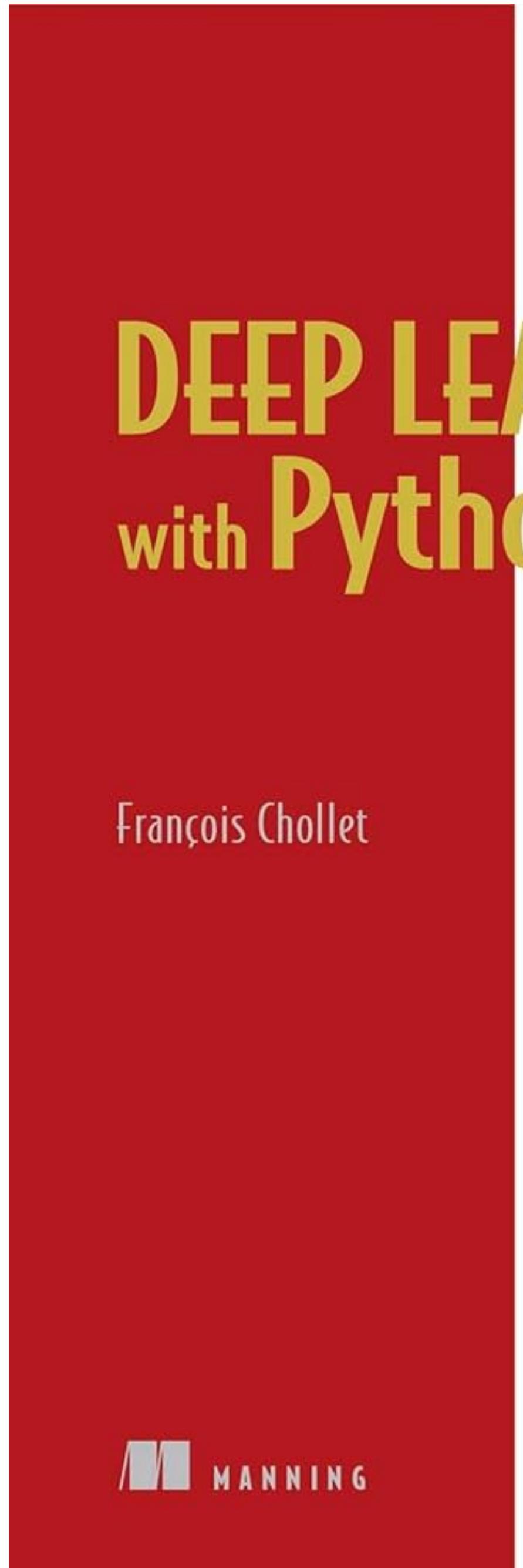
# Definition: Neural Networks

A machine learning (ML) technique that uses a network of interconnected layers of nodes to process data in a way that bears some resemblance to the human brain.



# Definition: Deep Learning

A type of machine learning that uses multi-layer neural networks to train computers to process data and make human-like decisions. Deep learning systems learn from large amounts of data, including images, text, audio, and video to recognize patterns, make predictions, and automate tasks.



# Definition: Generative AI

Deep-learning models  
that can generate high-  
quality text, images,  
video, and other  
content based on the  
data they were trained  
on.



**Artificial  
Intelligence**

**Machine  
Learning**

**Neural  
Networks**

**Deep  
Learning**

**Generative  
AI**

# Artificial Intelligence

Machine Learning

Deep Learning

1950's    1960's    1970's    1980's    1990's    2000's    2010's



# History

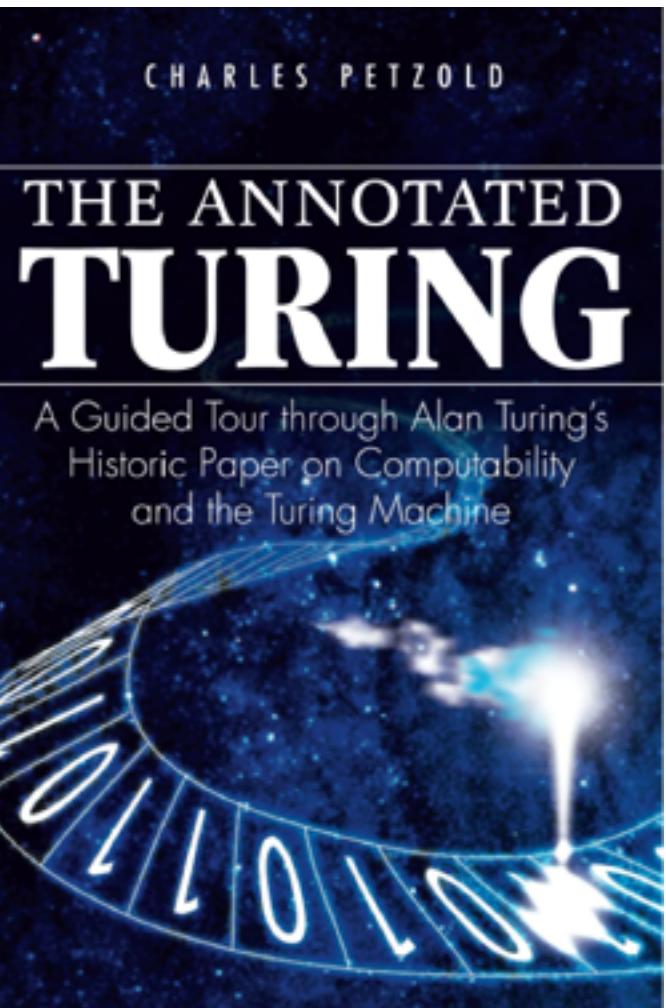
# How did we get here?

- Founders
- Timeline



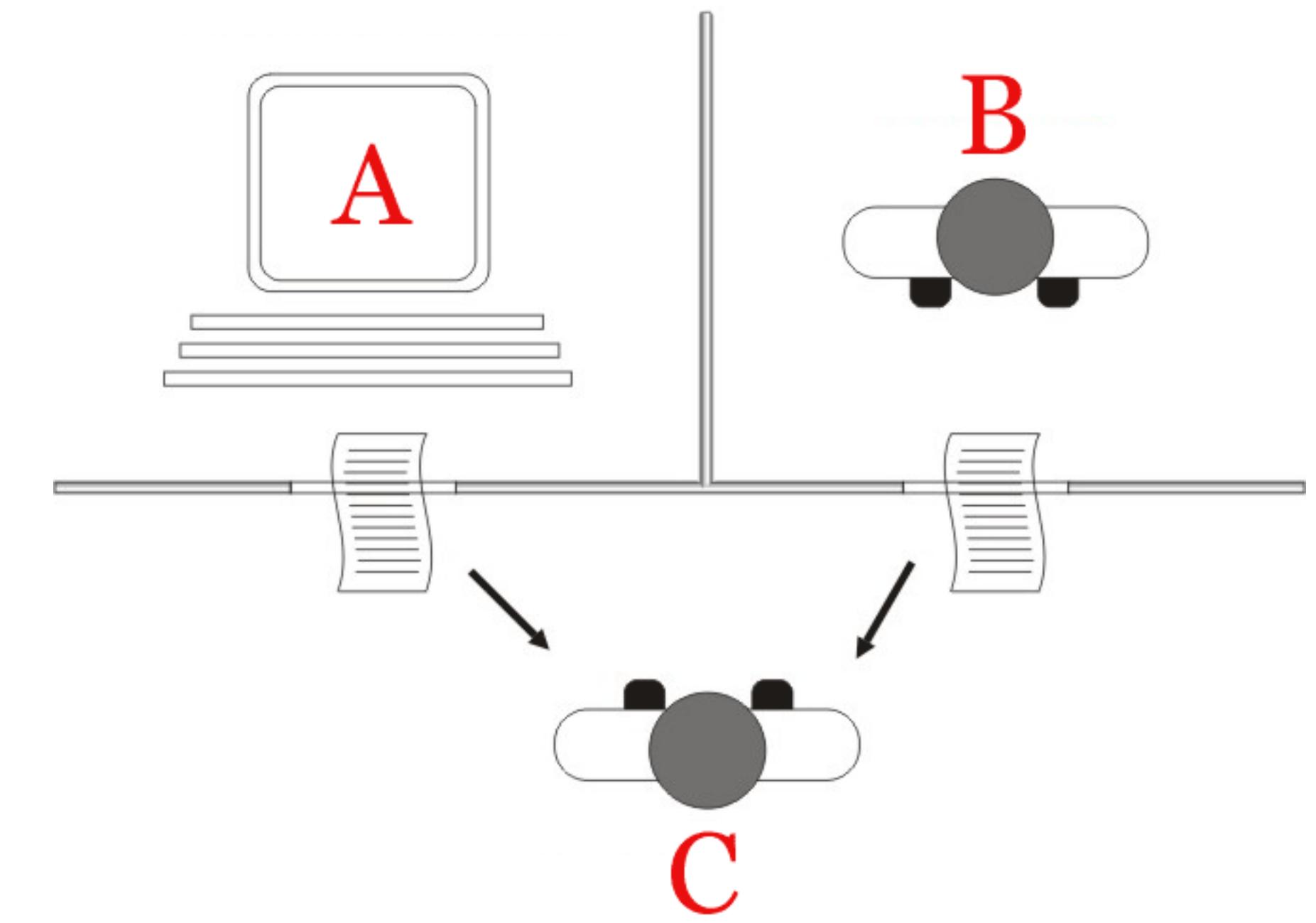
# Alan Turing

A Turing Machine is a theoretical construct that captures the essence of abstract symbol manipulation, i.e. computing. This invention inspired the founding of Computer Science and started the discussion about the possibility of thinking machines.



# The Turing Test (1950)

If a machine could carry on a conversation that was indistinguishable from a conversation with a human being, then it was reasonable to say that the machine was "thinking". This was the first serious proposal about how humanity might achieve artificial intelligence.



By Juan Alberto Sánchez Margallo - File:Test\_de\_Turing.jpg, CC BY 2.5

We'll conduct our own Turing Test later!

# How Did You Do On The AI Art Turing Test?

...

NOV 20, 2024



Share

...

Last month, I challenged 11,000 people to classify fifty pictures as either human art or AI-generated images.

I originally planned five human and five AI pictures in each of four styles: Renaissance, 19th Century, Abstract/Modern, and Digital, for a total of forty. After receiving many exceptionally good submissions from local AI artists, I fudged a little and made it fifty. The final set included paintings by Domenichino, Gauguin, Basquiat, and others, plus a host of digital artists and AI hobbyists.



One of these two pretty hillsides is by one of history's greatest artists. The other is soulless AI slop. Can you tell which is which?

# Claude Shannon

- **Invented digital computer design**
- **World's Greatest Master's Thesis!**
- **Father of Information Theory**
- **Mathematical definition of communication**
- **First appearance of the word “bit”**
- **Designed the first chess computer**



# Geoffrey Hinton

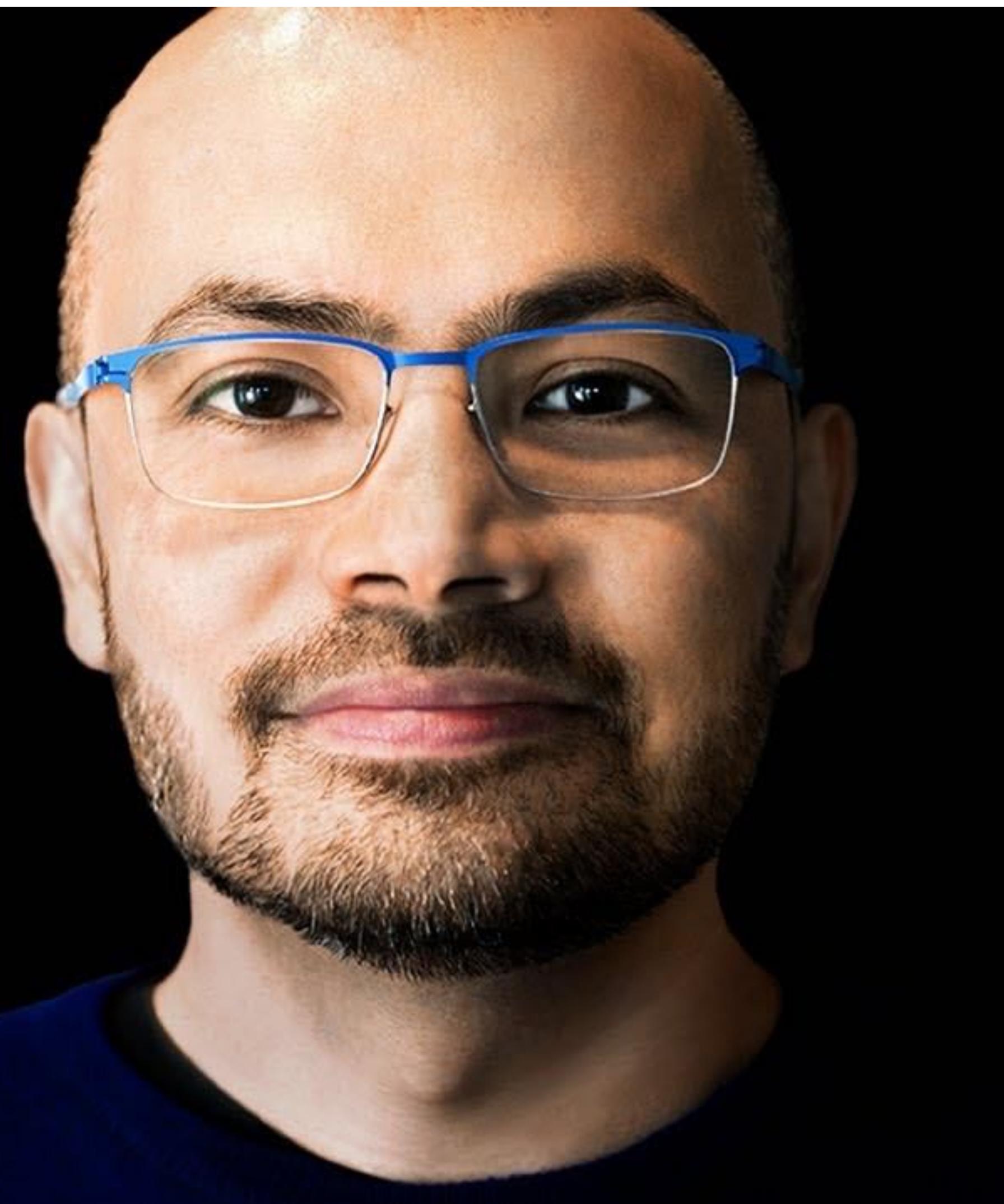


# Demis Hassabis

Demis  
Hassabis

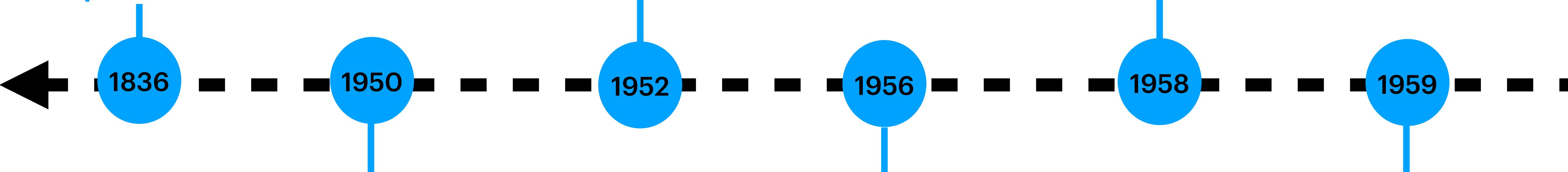
#299

Lex  
Fridman



# AI Development Timeline

Charles Babbage & Ada Lovelace design the Analytical Engine, an early mechanical general-purpose computer.



Alan Turing publishes "Computing Machinery and Intelligence," introducing the Turing Test.

Arthur Samuel developed a program to play checkers, which is the first to ever learn the game independently.

John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon organize the Dartmouth Conference, coining the term "Artificial Intelligence."

John McCarthy created LISP, the first programming language for AI research, which is still in popular use to this day.

Arthur Samuel created the term "machine learning" in a speech about teaching machines to play chess better than the humans who programmed them.

# AI Development Timeline

Edward Feigenbaum and Joshua Lederberg created the first “expert system” to replicate the decision-making abilities of human experts.

1965

1965

Joseph Weizenbaum created the first “chatterbot” (later shortened to chatbot), ELIZA, a mock psychotherapist, that used natural language processing (NLP) to converse with humans.

The first AI winter begins as funding and interest decline due to unmet expectations.

1972

1974

Marvin Minsky and Seymour Papert publish Perceptrons, highlighting limitations in neural networks.

1980

David Rumelhart, Geoffrey Hinton, and Ronald Williams publish the backpropagation algorithm, revitalizing neural networks.

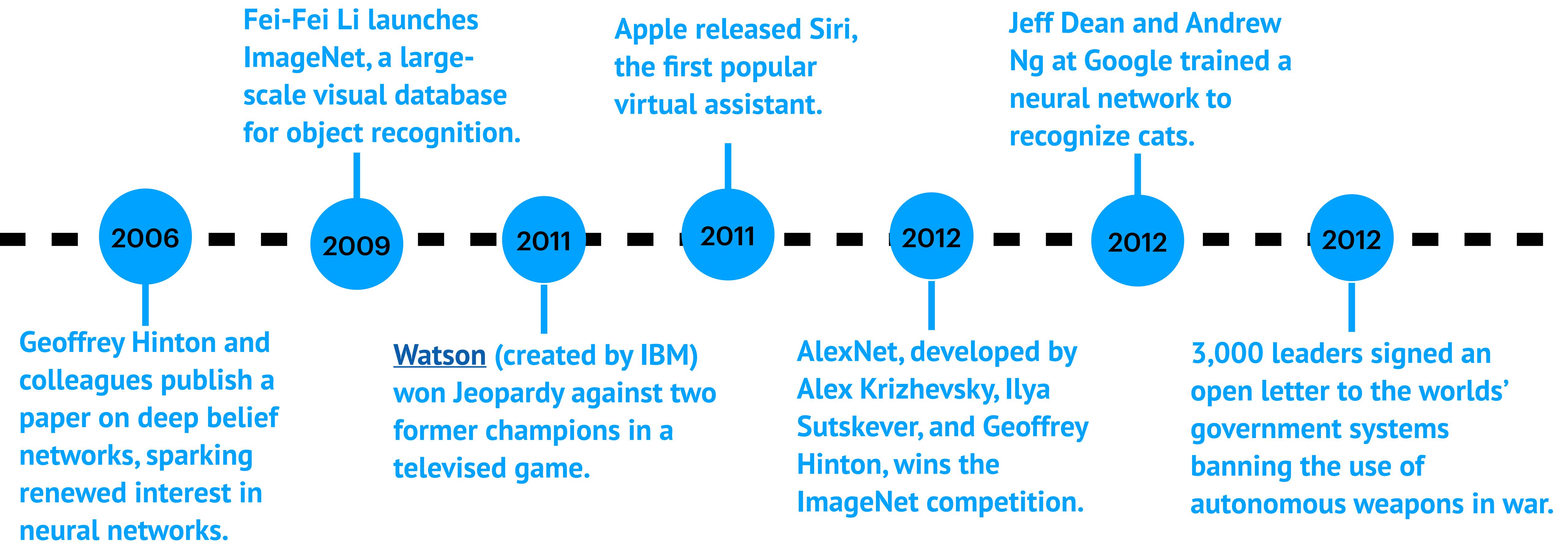
The resurgence of AI interest due to expert systems like MYCIN (medical diagnosis) and DENDRAL (chemical analysis).

1986

IBM's Deep Blue defeats world chess champion Garry Kasparov.

1997

# AI Development Timeline



# AI Development Timeline

DeepMind's AlphaGo defeats professional Go player Lee Sedol.

2015

2017

2017

2018

2020

2020

2022

2024

Two Facebook chatbots converse and learn how to negotiate, but dropped English and developed their own language, completely autonomously.

Google publishes “Attention is all you need”, unveiling the Transformer.

Chinese Alibaba beat humans on Stanford reading and comprehension test.

OpenAI releases GPT-3, a powerful language model with 175 billion parameters.

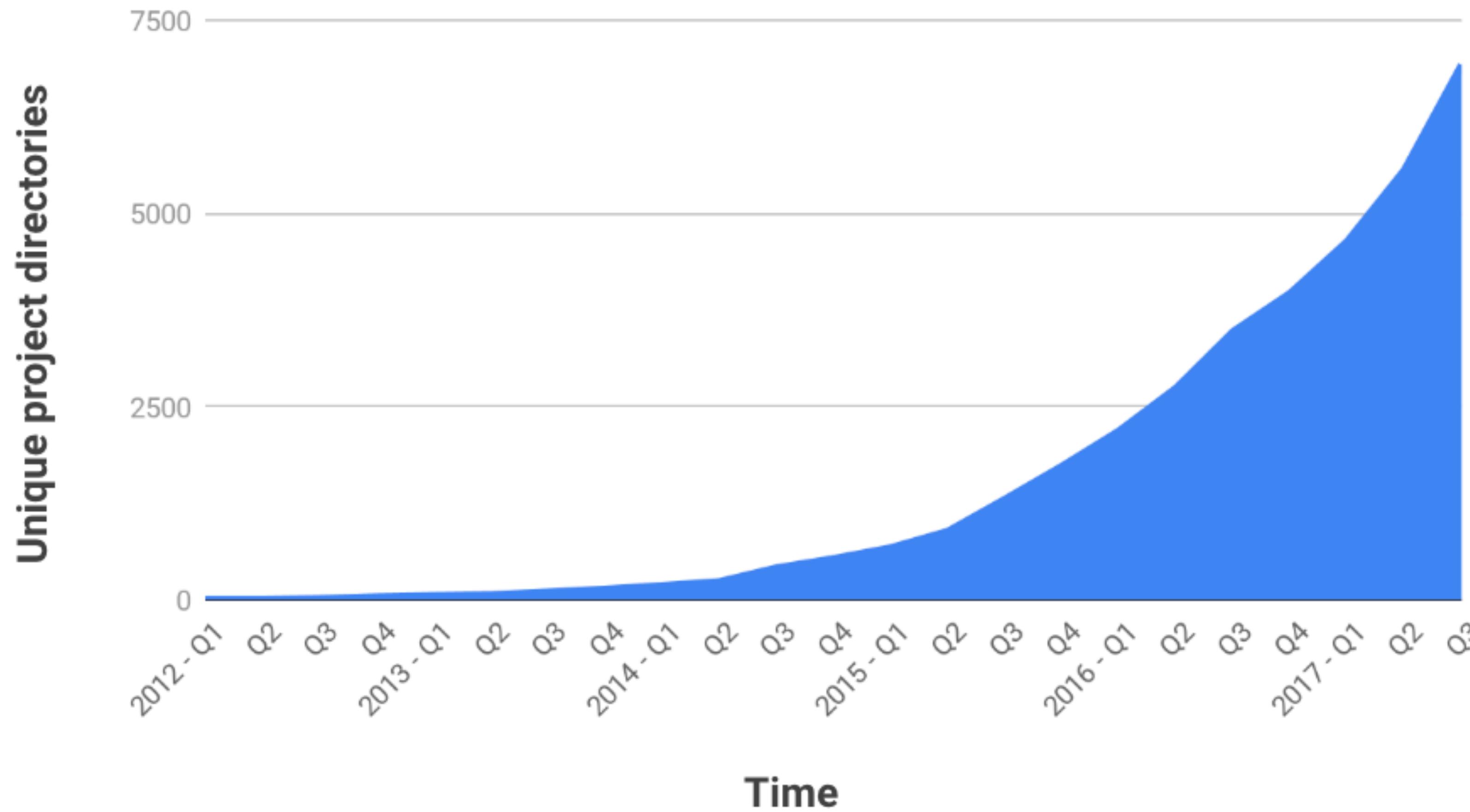
Generative AI goes mainstream with ChatGPT

Open source AI era begins

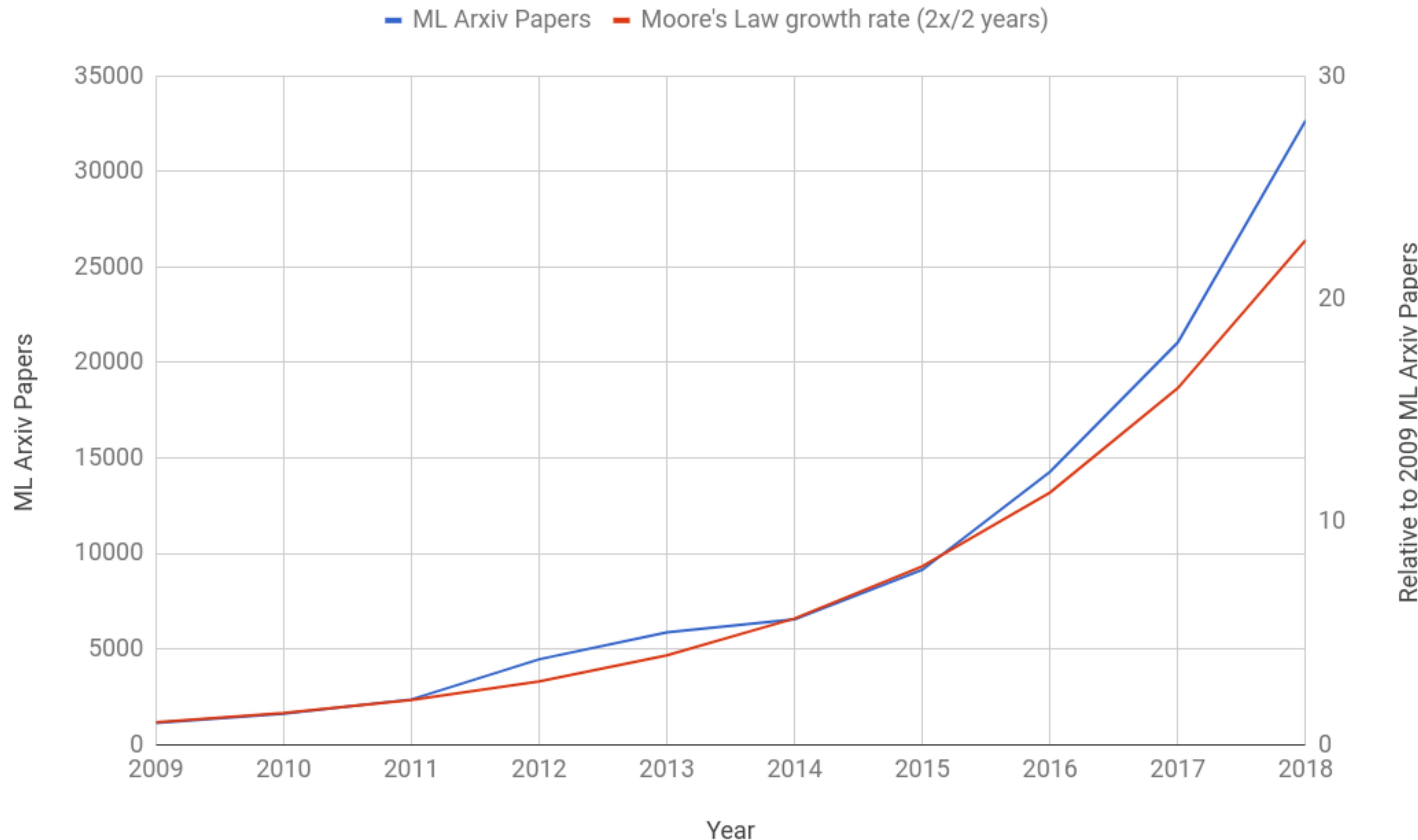
DeepMind's AlphaFold makes significant progress in protein folding.

# Use of Deep Learning at Google

source directories containing deep learning models



# Industry Adoption



# THE KURZWEIL CURVE

*Moore's Law is just the beginning: The power of technology will keep growing exponentially, says Kurzweil. By 2050, you'll be able to buy a device with the computational capacity of all mankind for the price of a nice refrigerator today.*

## Computer performance

Plotted by number of calculations per second per \$1,000

Years by which, according to  
Kurzweil, \$1,000 of computation  
will equal (or has already equaled)  
the intelligence of ...

... all human brains

... one human brain

... one mouse brain

... one insect brain

Kurzweil's  
projected  
trend line

• – COMPUTER TYPE

Hollerith Tabulator  
Bell Calculator Model I  
Univac I  
Apple Mac II

SOURCE: DATA FROM RAY KURZWEIL

$10^{40}$

$10^{30}$

$10^{20}$

$10^{10}$

1

$10^{-10}$

1900

1925

1950

1975

2001

'10

'23

2050

2075

# Why did everything change so quickly?

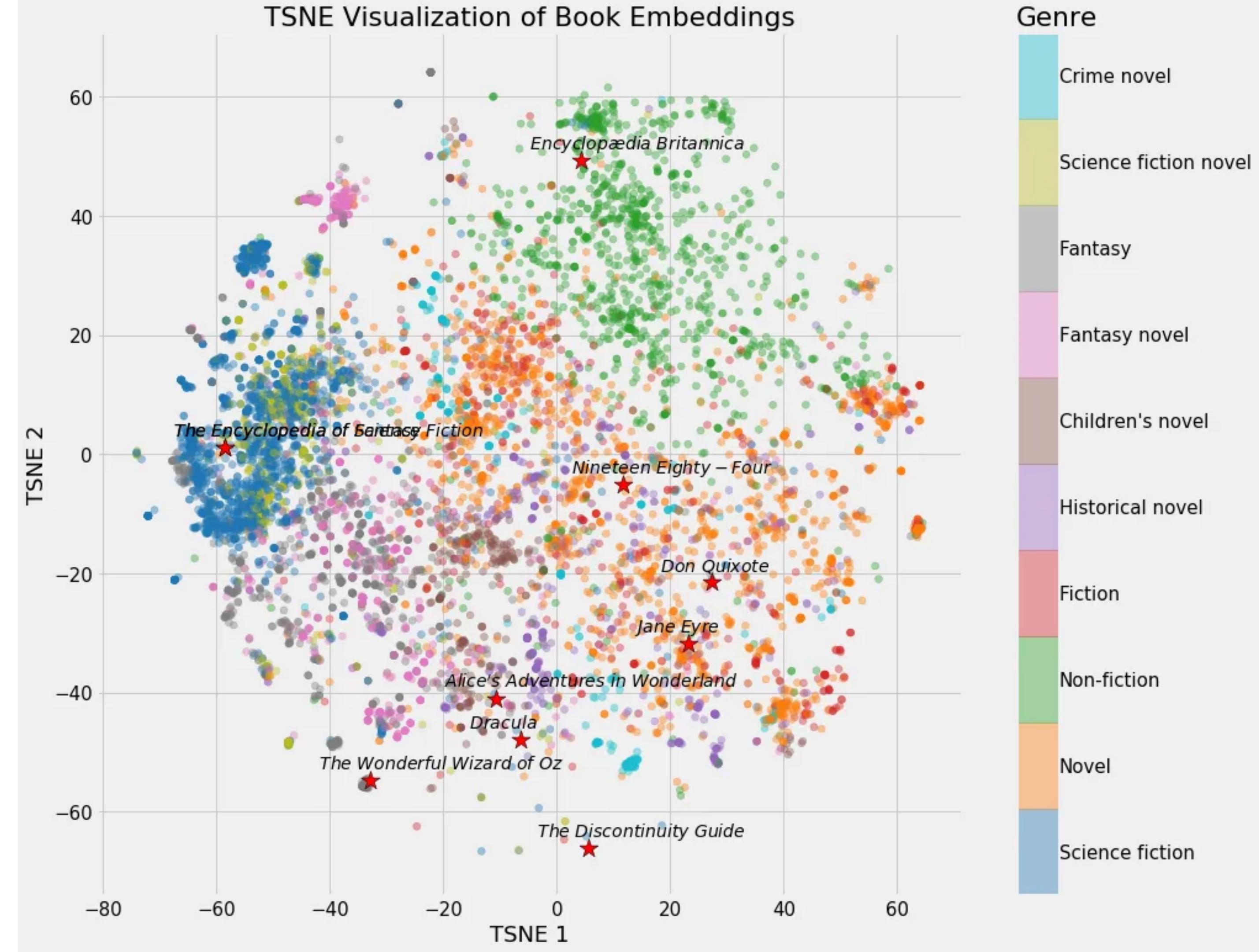
## A perfect storm:

- Moore's Law
- GPUs
- Cloud Computing
- Data abundance
- Github + Open Source
- New & improved techniques
- Transformers



# Mechanics

# How does AI work?



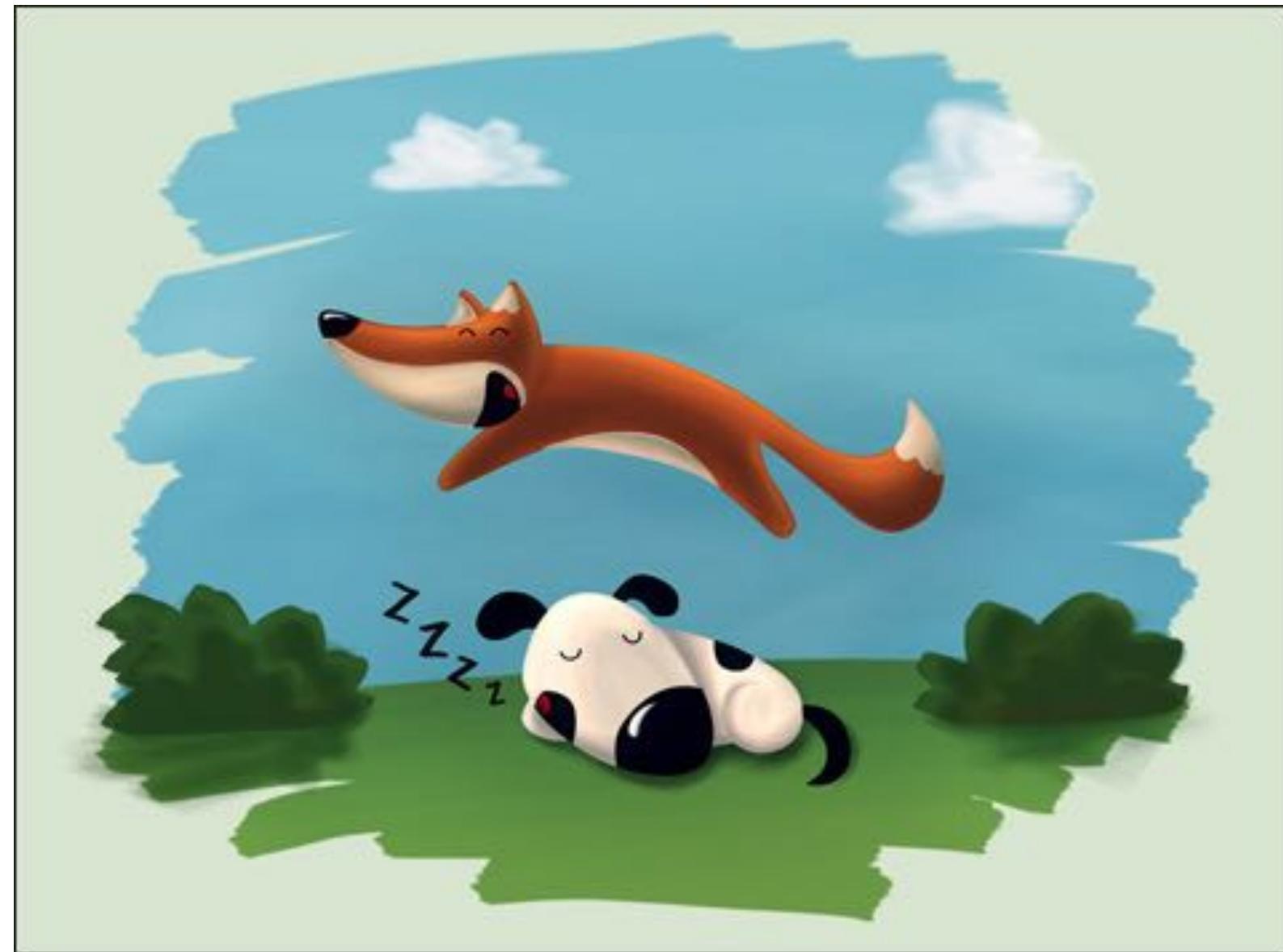
# Puppy vs. Muffin



**Not so easy**



**“The brown quick fox jumps over the lazy dog.”**



# English Adjective Order

1. Quantity or number
2. Quality or opinion
3. Size
4. Age
5. Shape
6. Colour
7. Proper adjective
8. Purpose or qualifier

**Machine learning is  
learning from rules  
*\*plus experience\****

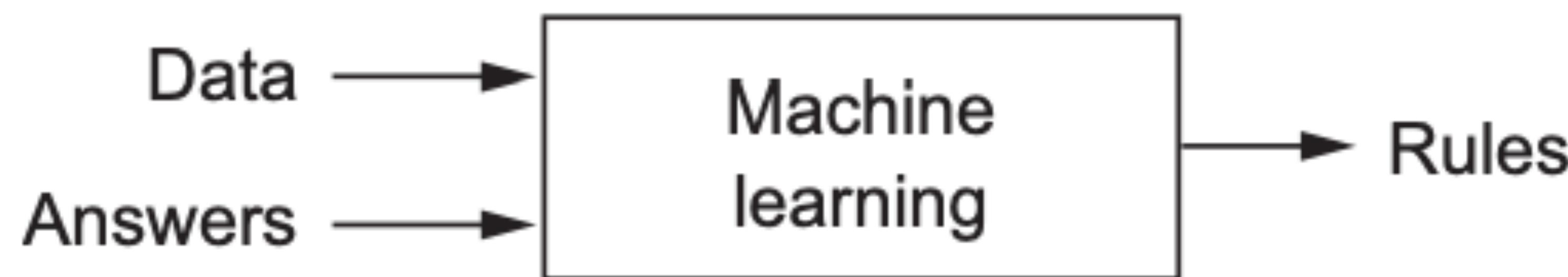
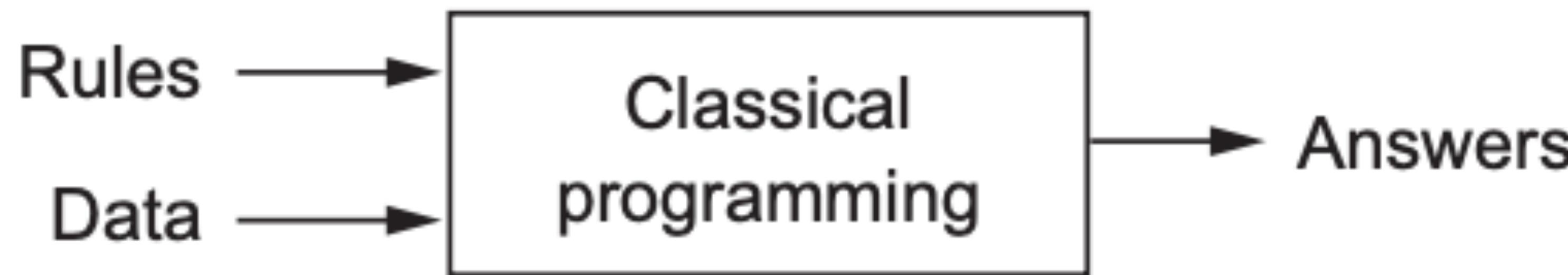


**“Instead of trying to produce a program to simulate the adult mind, why not rather try to produce one which simulates the child’s mind?”**

- Alan Turing, 1950

**This is the main idea behind Machine Learning**

# Paradigm Shift



# Types of Machine Learning

## Supervised Learning

- labeled data
- learn relationships

Examples:

- spam detection
- image classification

## Unsupervised Learning

- unlabeled data
- find data structure

Examples:

- customer segmentation
- anomaly detection

## Reinforcement Learning

- reward based
- find optimal strategy

Examples:

- game playing
- autonomous driving

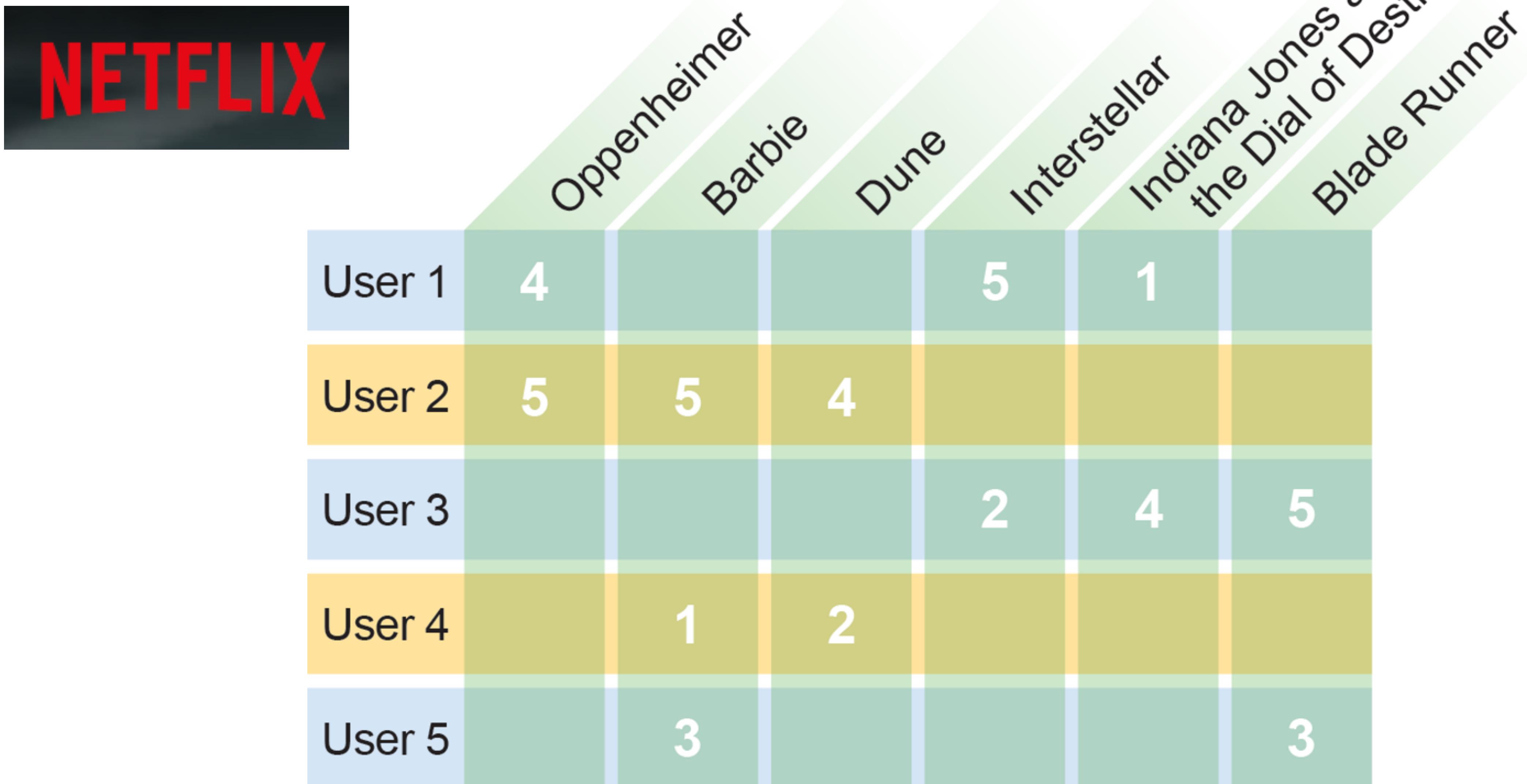
# Supervised Learning

## Cats vs. Dogs



- **2009 - world's most advanced computer can't tell a cat from a dog**
- **2012 - solved by Google \*but\* required 16,000 computers!**
- **2015 - Microsoft, Google beat humans at image recognition**
- **2019 - AI beats Stanford radiologists in chest X-ray diagnostics competition**

# Unsupervised Learning



# Cloud of similarities

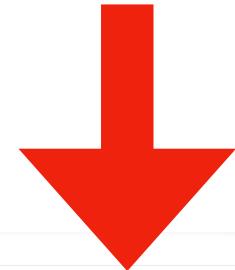


# Reinforcement Learning



# Types of Machine Learning

We'll focus on this category



## Supervised Learning

- labeled data
- learn relationships

Examples:

- spam detection
- image classification

## Unsupervised Learning

- unlabeled data
- find data structure

Examples:

- customer segmentation
- anomaly detection

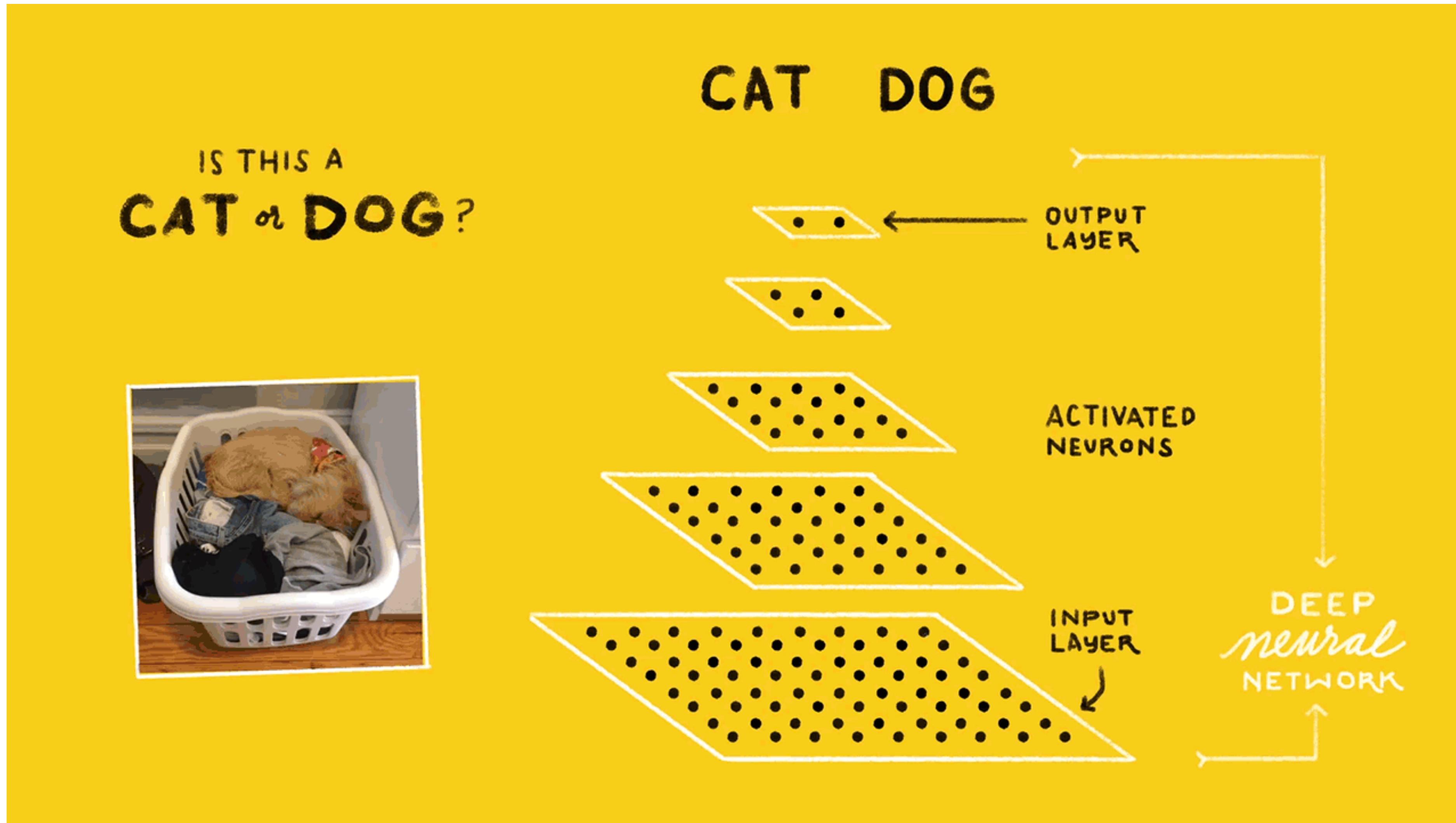
## Reinforcement Learning

- reward based
- find optimal strategy

Examples:

- game playing
- autonomous driving

# Neural Networks



# Which digit is this?



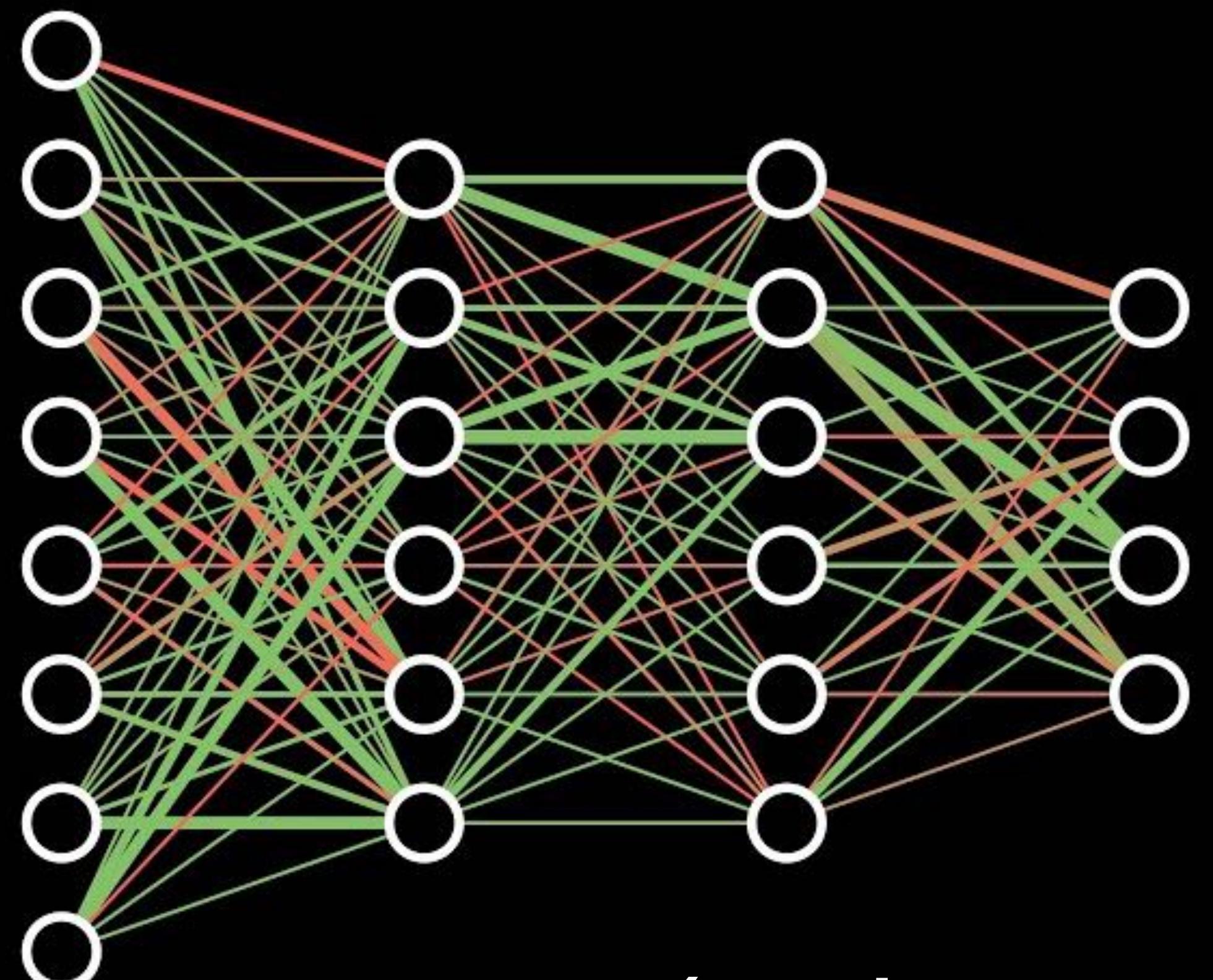
# The MNIST Challenge

training digits and their labels	
9	9
1	1
0	0
3	1
1	1
6	6
6	5
5	0
0	5
7	7
2	2
5	5
8	8
4	4
9	9
9	9
7	7
0	0
0	0
4	4
9	9
4	4

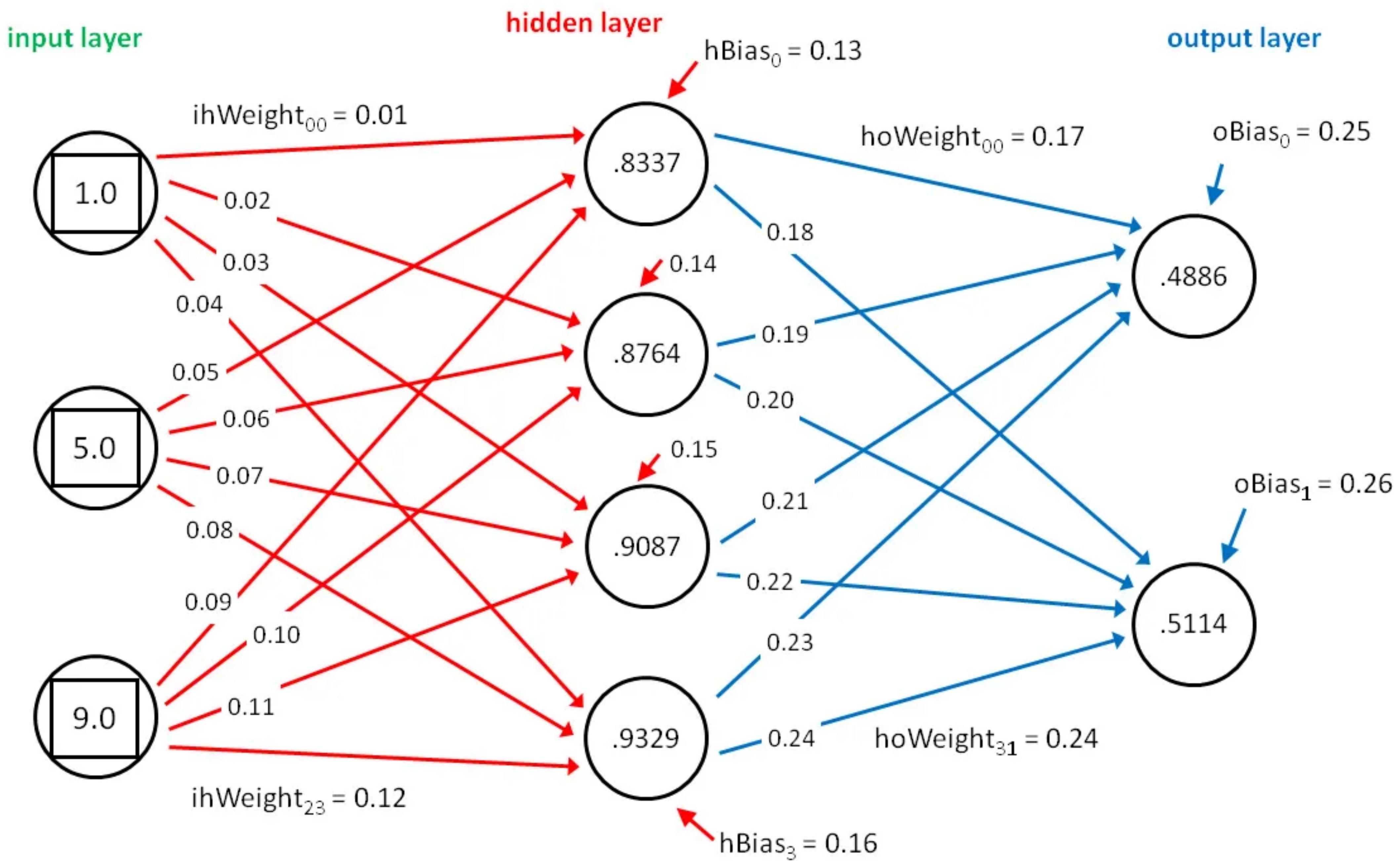
validation digits and their labels	
7	7
2	2
1	1
0	0
4	4
1	1
4	4
9	9
5	5
9	0
0	6
6	9
9	0
1	1
5	5
9	9
7	7
3	3
4	4
9	9
6	6
4	5

# Neural Networks



From the  
ground up

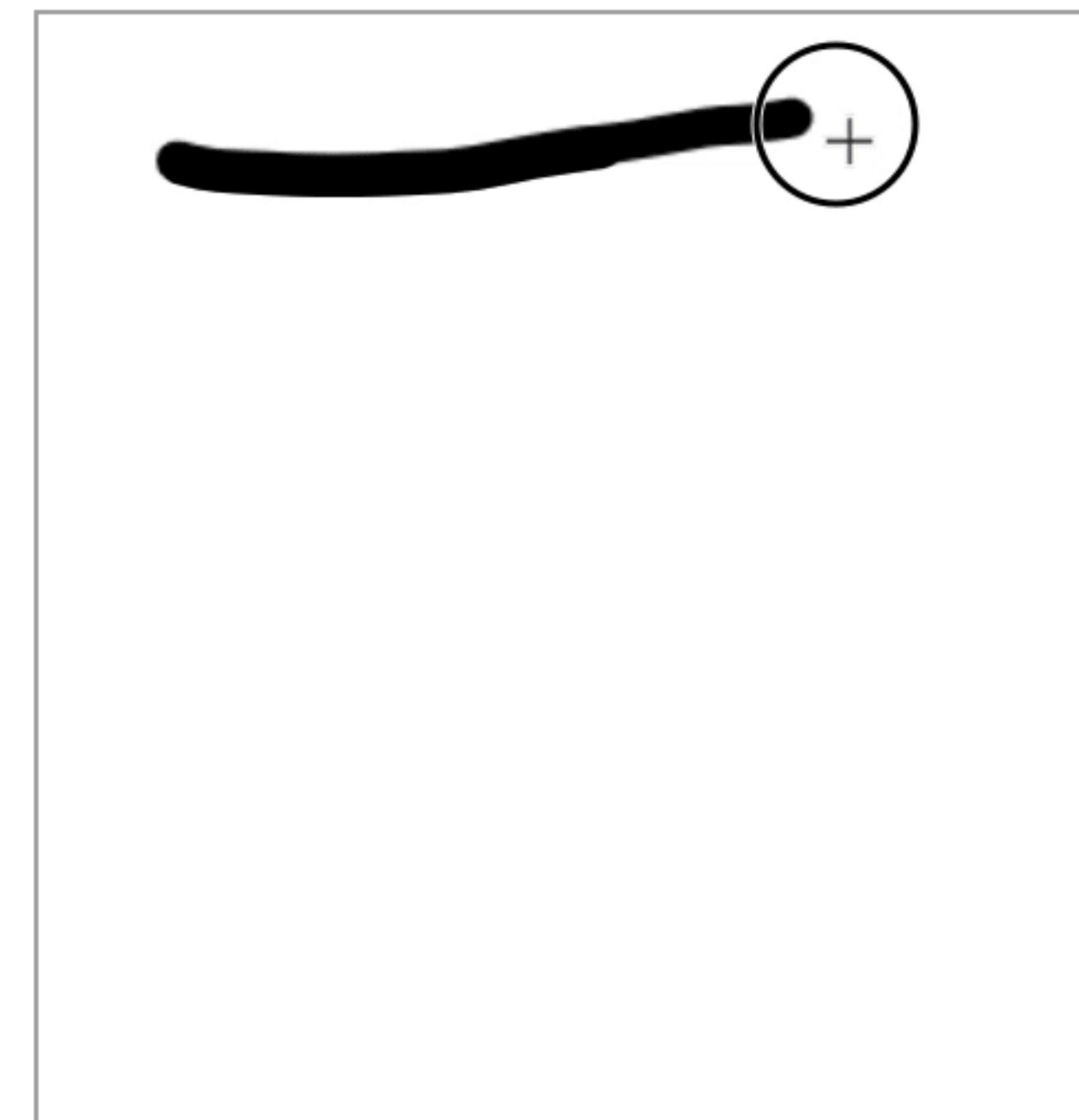
(watch 2:40-5:30, 15:18-16:20)





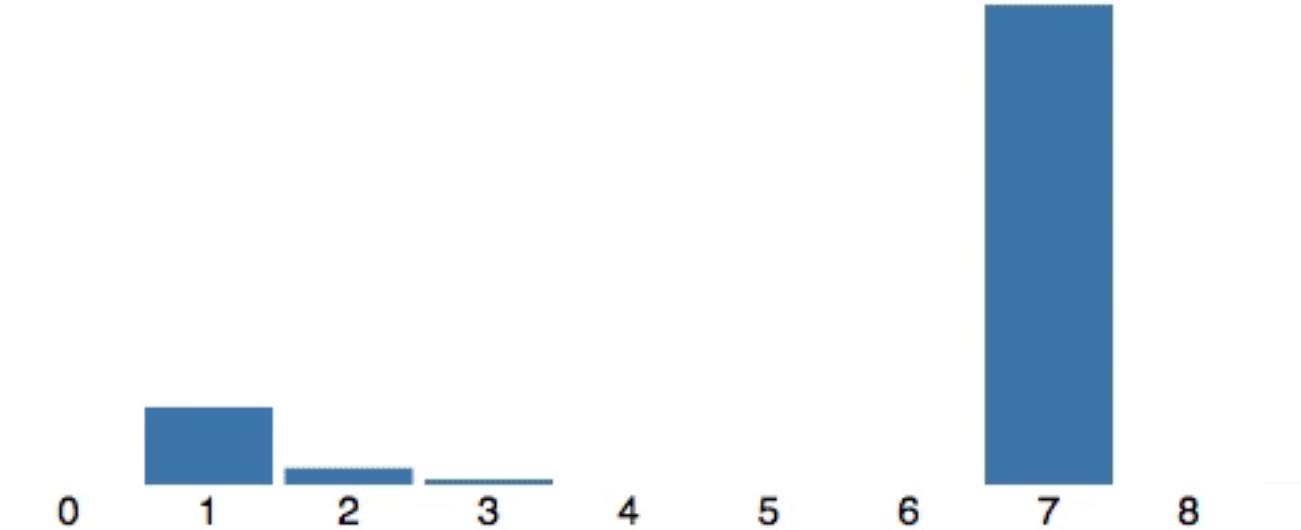
# TensorFlow.js + MNIST

Draw a digit by hand and have it classified in real time, as you draw, by a machine learning model trained against the [MNIST](#) dataset. The model used for this page is a convolutional neural network (CNN) built using Keras/TensorFlow on a Google Tensor Processing Unit (TPU). It's explained in depth and is available for you to build for yourself at [bit.ly/mco-mnist-lab](https://bit.ly/mco-mnist-lab).



Draw a single digit (0-9) in the empty box.

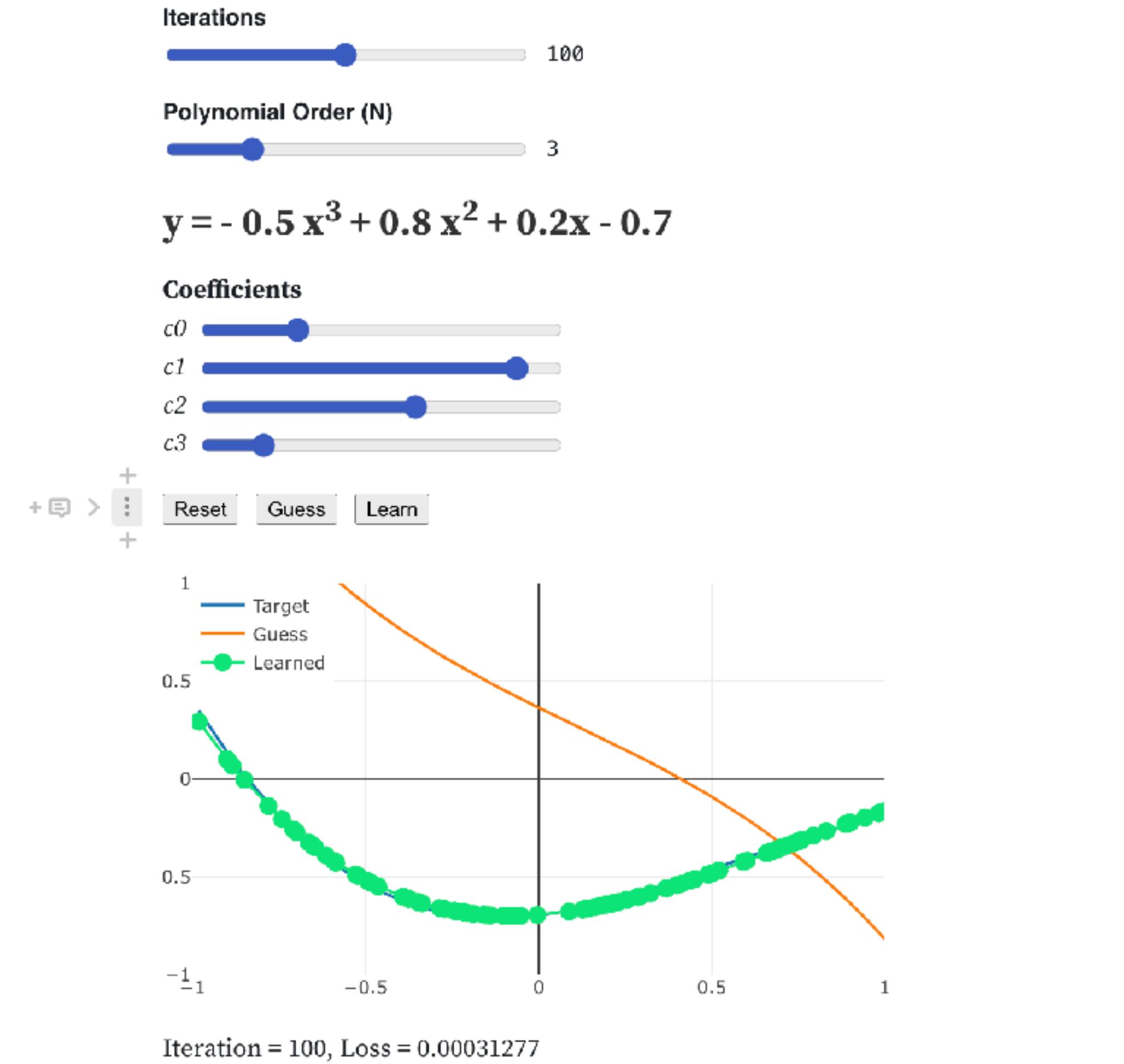
Clear



Made with ❤ by [Marc](#).

# Neural Network Training

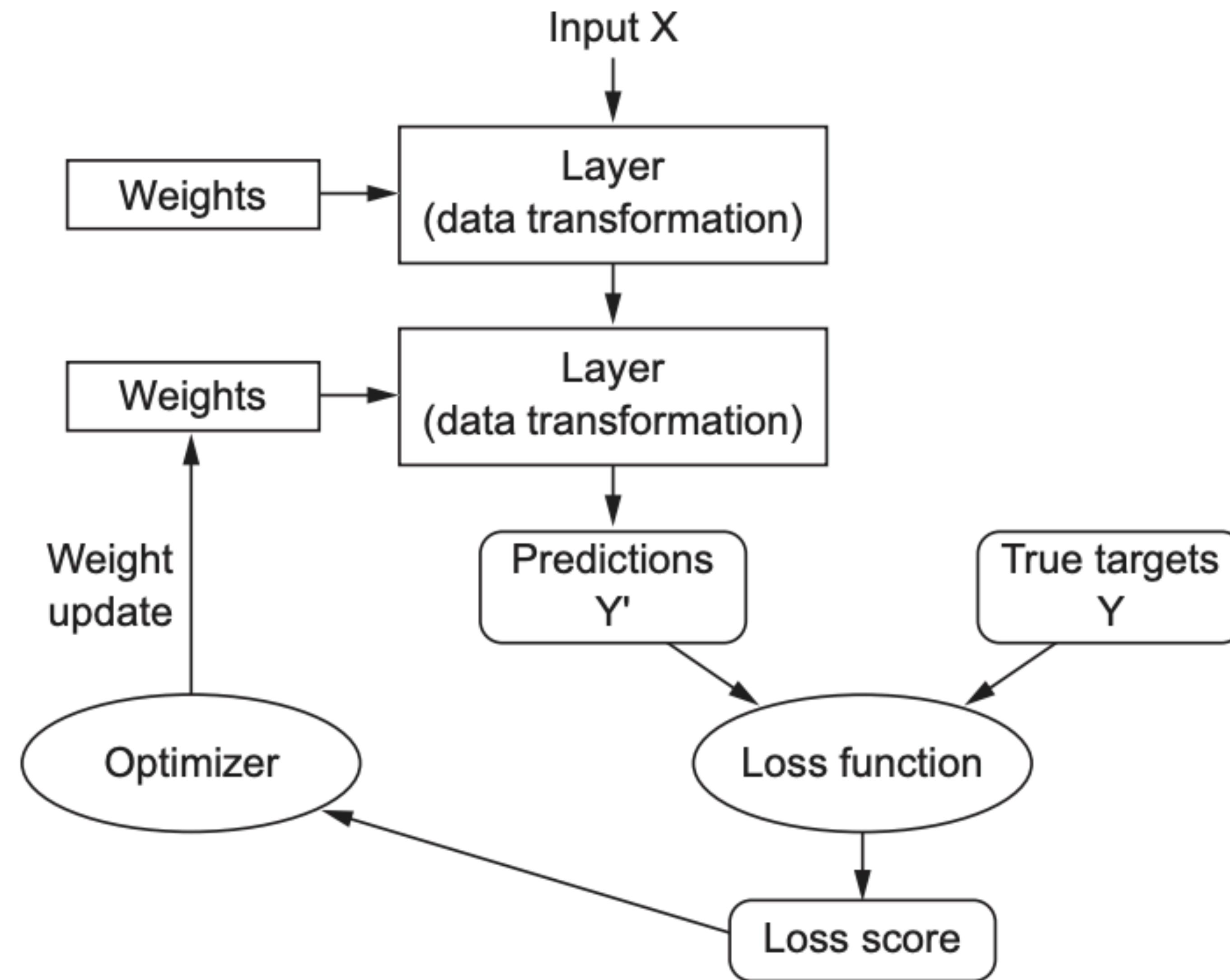
- guess values
- improve guess
- repeat until “close enough”



Coefficient	Target	Predicted
Coefficient 0	-0.5	-0.38080
Coefficient 1	0.8	0.78087
Coefficient 2	0.2	0.12938
Coefficient 3	-0.7	-0.69594

As you can see, the learned polynomial starts out following the initial curve, and fairly quickly moves to minimize the error between it and the target polynomial. The green dots you see are reminding you that we're using a discrete set of points to perform these calculations (the number of which you can vary in the appendix).

# The Training Process

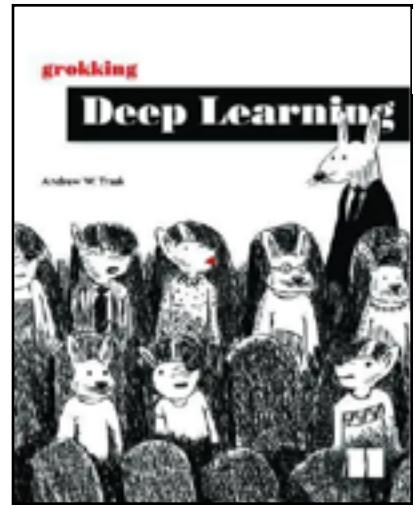


# **Embeddings**

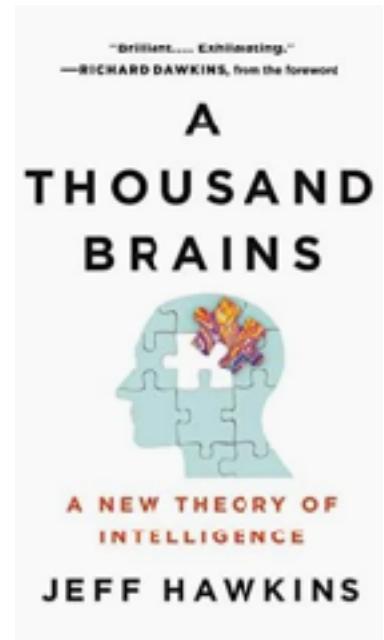
**(The secret sauce of deep learning)**

# Recommending AI Books

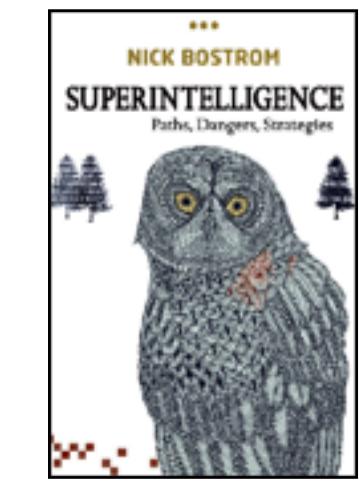
**Beginner**



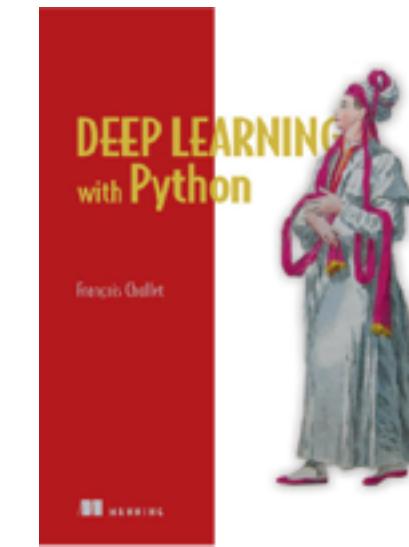
[ -5 ]



[ -3 ]



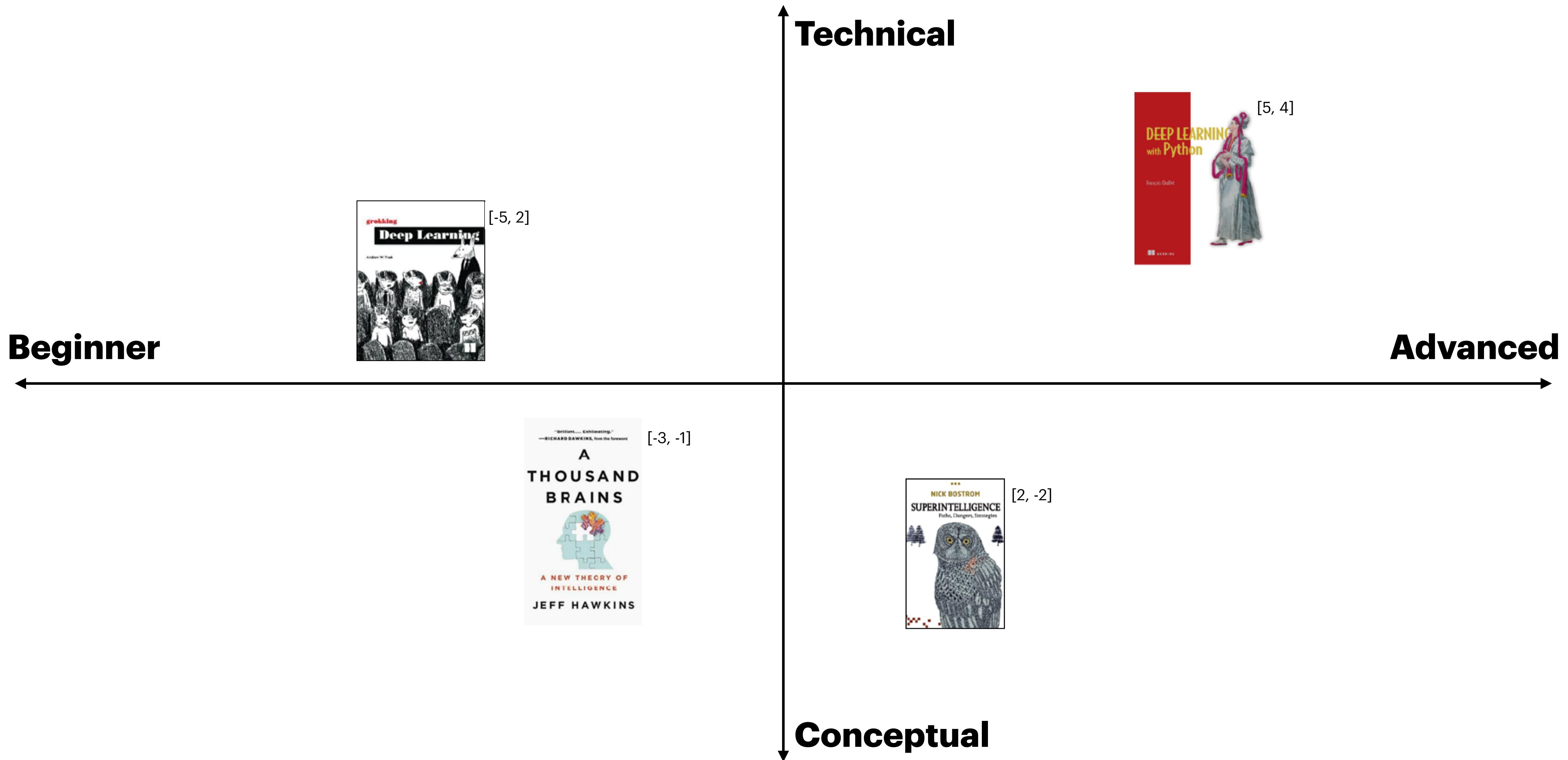
[ 2 ]



[ 5 ]

**Advanced**

# Better: 2-D Coordinate System



# Even better: use N-dimensions

- Based on the labelled training data, neural networks are able to “embed” data (words, images, video, audio) into N-dimensional space.
- We can represent embeddings as sequences of numbers, aka vectors:
  - [a<sub>0</sub>, a<sub>1</sub>, a<sub>2</sub>, ..., a<sub>N</sub>]
  - Similar objects get embedded in similar locations in N-space.
  - This gives us the ability to do *semantic search*, i.e., find things that have similar meaning.

# Embeddings In Action

←

🔍 dog in water X

✓ Sat, May 15, 2021

▼

# Google's AI Awakening

[Sign in / Sign up](#)

Jun Rekimoto : 暮本純一

[Follow](#)

人間とテクノロジーの未来を探求しています。HUMAN AND TECHNOLOGY

Nov 11, 2016 · 5 min read

You seem to have made it into a neural network Try the beginning with [The Great Gatsby Google Translate](#)



ニューラルネット化したらしいGoogle翻訳でThe Great Gatsby冒頭を試してみる：

原文：In my younger and more vulnerable years my father gave me some advice that I've been turning over in my mind ever since.

“Whenever you feel like criticizing any one,” he told me, “just remember that all the people in this world haven't had the advantages that you've had.”

google：私の若くて傷つきやすい年の頃、私の父は私が私の心の中でそれ以来変わってきたアドバイスを私にくれました。

「誰かを批判する気持ちがあるときはいつでも、彼は私に言いました。

「この世界のすべての人々はあなたが持っていた優位性を持っていないことを覚えておいてください。

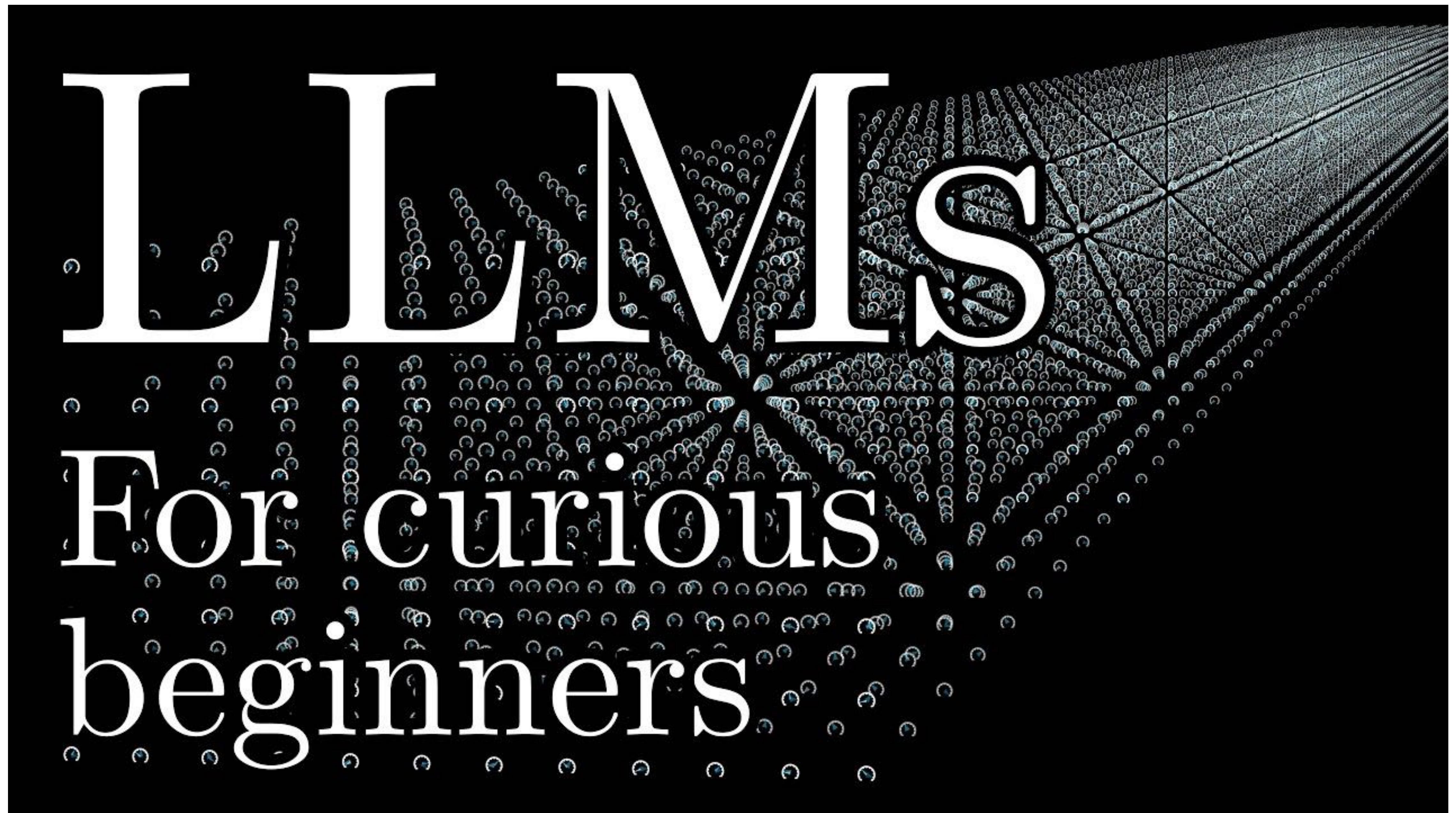
FEATURE

## The Great A.I. Awakening

How Google used artificial intelligence to transform Google Translate, one of its more popular services — and how machine learning is poised to reinvent computing itself.

# TUMS

For curious  
beginners



AI

# An Intuitive Guide to How LLMs Work

Chatting by chance

OCTOBER 6, 2024



# Understanding LLMs from Scratch Using Middle School Math

A self-contained, full explanation to inner workings of an LLM



Rohit Patel · [Follow](#)

Published in Towards Data Science · 40 min read · Oct 19, 2024

---

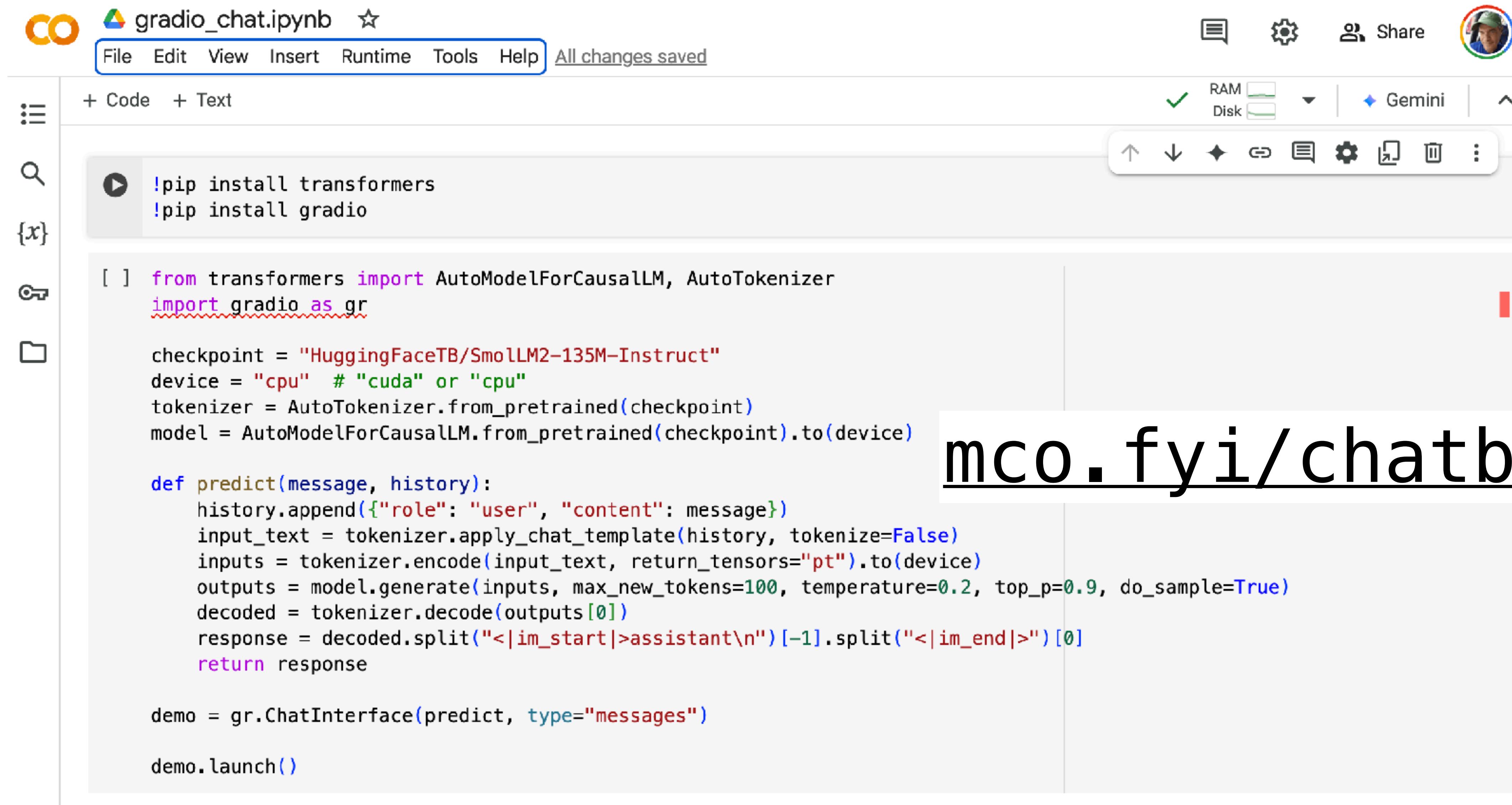
👏 6.1K

🗨 78



...

# Let's build a chatbot!



The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** CO gradio\_chat.ipynb ☆
- Menu Bar:** File Edit View Insert Runtime Tools Help All changes saved
- Toolbar:** RAM Disk Gemini
- Code Cell:** !pip install transformers  
!pip install gradio
- Code Block:** [ ] from transformers import AutoModelForCausalLM, AutoTokenizer  
import gradio as gr  
  
checkpoint = "HuggingFaceTB/SmollM2-135M-Instruct"  
device = "cpu" # "cuda" or "cpu"  
tokenizer = AutoTokenizer.from\_pretrained(checkpoint)  
model = AutoModelForCausalLM.from\_pretrained(checkpoint).to(device)  
  
def predict(message, history):  
 history.append({"role": "user", "content": message})  
 input\_text = tokenizer.apply\_chat\_template(history, tokenize=False)  
 inputs = tokenizer.encode(input\_text, return\_tensors="pt").to(device)  
 outputs = model.generate(inputs, max\_new\_tokens=100, temperature=0.2, top\_p=0.9, do\_sample=True)  
 decoded = tokenizer.decode(outputs[0])  
 response = decoded.split("<|im\_start|>assistant\n")[-1].split("<|im\_end|>")[0]  
 return response  
  
demo = gr.ChatInterface(predict, type="messages")  
  
demo.launch()

[mco.fyi/chatbot](https://mco.fyi/chatbot)

MJ

What will you imagine?



Search



berada.travel 15h



A child is playing with a baby  
koala really tiny, they both seem  
to be friends. The photo should  
look real and not a cartoon.

ar 9:16 v 6.1

Use

Image

Style

Prompt

# How A.I. Creates Art - A Gentle Introduction to Diffusion Models

January 24, 2023 · 8 min read



Zain Hasan

Developer Advocate

The thumbnail features a dark blue background with a subtle, glowing blue gradient. Overlaid on the right side is a large, abstract white line drawing that resembles a brain or a complex network. In the top left corner, there's a small yellow and green logo followed by the word "Weaviate" in white. The main title "How A.I. Creates Art: A Gentle Introduction to Diffusion Models" is centered in the middle-left area in a large, white, sans-serif font. At the bottom left, there's a white button with a thin green border containing the text "Read the blog" in a white sans-serif font.

How Diffusion Models Work

How Text Prompts Control the Image Generation Process

Diffusion Model Resources to Create Art

Sources and Further Reading

# Diffusion Models



# Let's generate an image!

co mco\_stable\_diffusion.ipynb ☆

File Edit View Insert Runtime Tools Help Last saved at 20:02

+ Code + Text Connect GPU ▾

🔍 {x} 🔑 📁 < >

## Stable Diffusion 🎨

...using  `diffusers`

Stable Diffusion is a text-to-image latent diffusion model created by the researchers and engineers from [CompVis](#), [StabilityAI](#) and [LAION](#). It's trained on 512x512 images from a subset of the [LAION-5B](#) database. This model uses a frozen CLIP ViT-L/14 text encoder to condition the model on text prompts. With its 860M UNet and 123M text encoder, the model is relatively lightweight and can run on many consumer GPUs. See the [model card](#) for more information.

This Colab notebook shows how to use Stable Diffusion with the 😊 Hugging Face  [Diffusers library](#).

Let's get started!

### 1. How to use `StableDiffusionPipeline`

Before diving into the theoretical aspects of how Stable Diffusion functions, let's try it out a bit 😊.

In this section, we show how you can run text to image inference in just a few lines of code!

**mco.fyi/sdiff**

# **State of the Art**

# Answering Questions

What can I help with?

0 Message ChatGPT



Create image

Summarize text

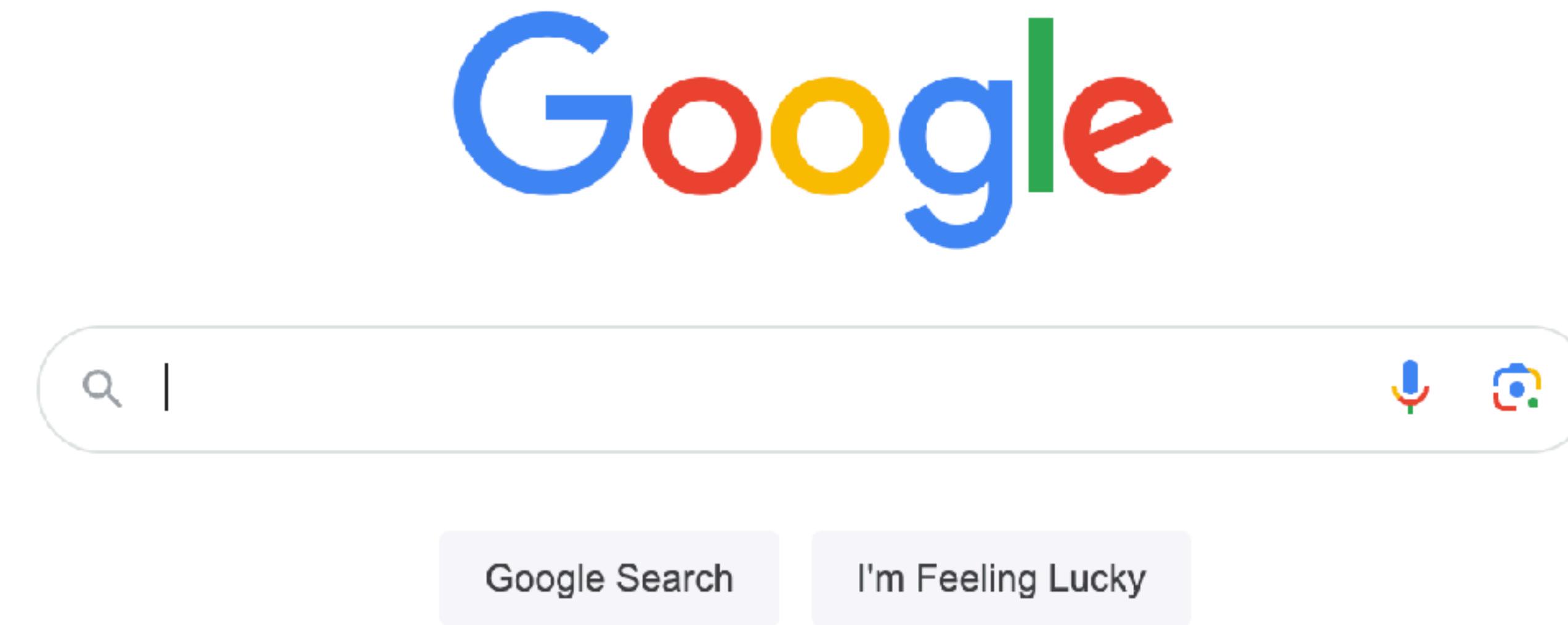
Brainstorm

Code

Make a plan

More

# Does that look familiar?



[All](#) [Images](#) [Videos](#) [Web](#) [News](#) [Books](#) [Maps](#) [More](#)[Tools](#)

### AI Overview

A transformer is a neural network architecture that processes sequential data by learning relationships between elements. Transformers are used in many machine learning applications, including natural language processing (NLP), speech recognition, and computer vision.

[Learn more](#)

### Transformers in Artificial Intelligence - AWS

Transformers are a type of neural network architecture that transforms or changes an...

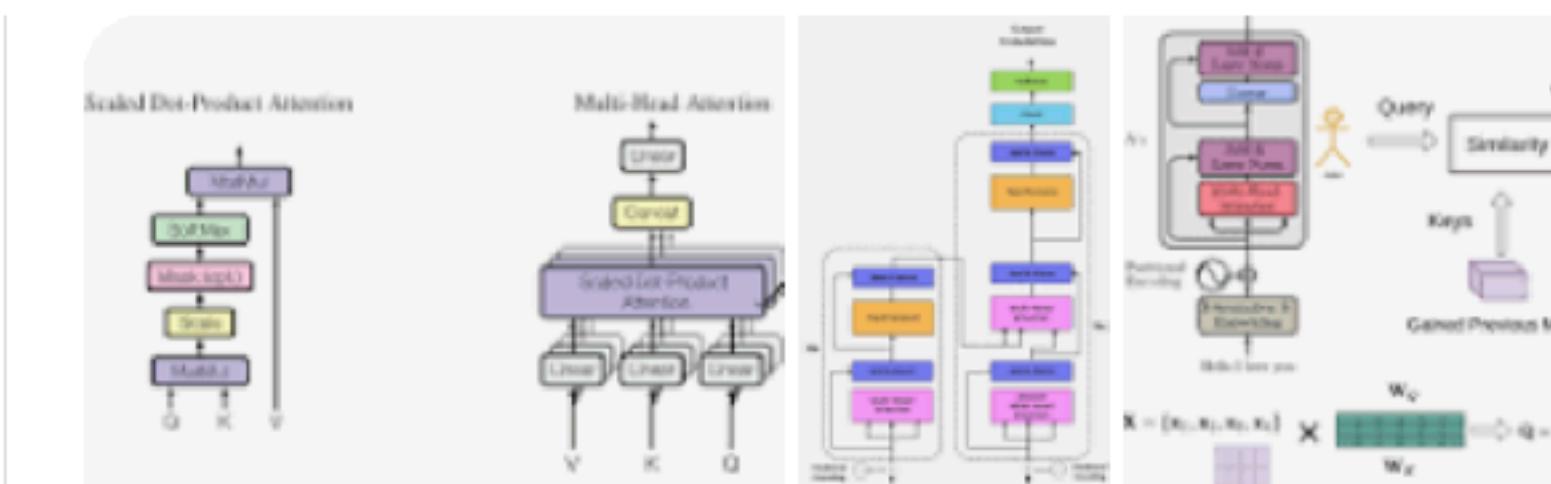
AWS

### How transformers work

1. Transformers learn context by tracking relationships between elements in a sequence.

[Show more](#)

Transformers are a type of neural network architecture that transforms or changes an input sequence into an output sequence. They do this by learning context and tracking relationships between sequence components.



# *OpenAI Unveils A.I. Agent That Can Use Websites on Its Own*

The new tool, called Operator, can shop for groceries or book a restaurant reservation. But it still needs help from humans.

A screenshot of a web-based AI interface titled "Operator". At the top, there's a red box containing a quote: "This is not the most robust thing in the world," he said. "But it is way better than this kind of technology used to be." Below the quote, the word "Operator" is displayed in bold black text. A large input field asks "What can I help you do?", with a small microphone icon and an upward arrow icon. Below the input field are several category buttons: "Dining & Events" (highlighted in black), "Delivery", "Local Services", "Shopping", "Travel", and "News". At the bottom, there are three examples of what the AI can do: "Find me a restaurant with a great happy hour", "Find me 2 tickets for the next concert at The", and "Suggest a 30 minute meal with chicken and".

Listen to this article • 3:40 min | Learn more

Share full article

“This is not the most robust thing in the world,” he said. “But it is way better than this kind of technology used to be.”

Operator

What can I help you do?

Dining & Events Delivery Local Services Shopping Travel News

Find me a restaurant with a great happy hour

Find me 2 tickets for the next concert at The

Suggest a 30 minute meal with chicken and

# *How Chinese A.I. Start-Up DeepSeek Is Competing With Silicon Valley Giants*

The company built a cheaper, competitive chatbot with fewer high-end computer chips than U.S. behemoths like Google and OpenAI, showing the limits of chip export control.



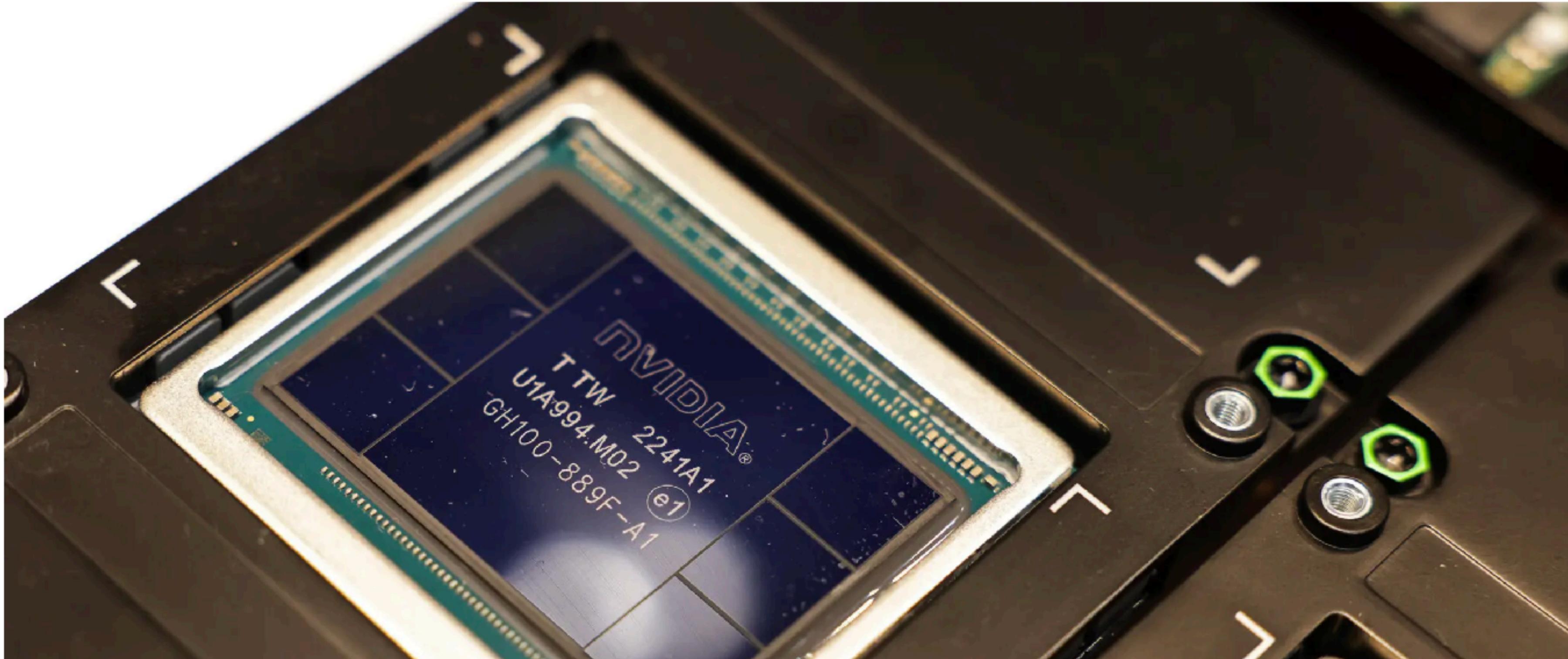
Listen to this article · 8:27 min [Learn more](#)



Share full article



204



[LLM](#)

# DeepSeek and the Effects of GPU Export Controls

January 23, 2025 by Vincent Schmalbach

Last week, DeepSeek unveiled their V3 model, trained on just 2,048 H800 GPUs - a fraction of the hardware used by OpenAI or Meta. DeepSeek claims their model matches or exceeds several benchmarks set by GPT-4 and Claude

What's interesting isn't just the results, but how they got there.

## The Numbers Game

---

Let's look at the raw figures:

- Training cost: \$5.5M (vs \$40M for GPT-4)
- GPU count: 2,048 H800s (vs estimated 20,000+ H100s for major labs)
- Parameters: 671B
- Training: 2.788M GPU hours

Recent research shows model training costs growing by 2.4x annually since 2016. Everyone assumed you needed massive GPU clusters to compete at the frontier. DeepSeek suggests otherwise.

- released Dec 2024
- comparable to GPT-4o
- ~1/10 of training cost
- multi-head latent attention
- mixture of experts
- supervised fine-tuning

an auxiliary-loss-free strategy for load balancing and sets a multi-token prediction training objective for stronger performance. We pre-train DeepSeek-V3 on 14.8 trillion diverse and high-quality tokens, followed by Supervised Fine-Tuning and Reinforcement Learning stages to fully harness its capabilities. Comprehensive evaluations reveal that DeepSeek-V3 outperforms other open-source models and achieves performance comparable to leading closed-source models. Despite its excellent performance, DeepSeek-V3 requires only 2.788M H800 GPU hours for its full training. In addition, its training process is remarkably stable. Throughout the entire training process, we did not experience any irrecoverable loss spikes or perform any rollbacks. The model checkpoints are available at <https://github.com/deepseek-ai/DeepSeek-V3>.

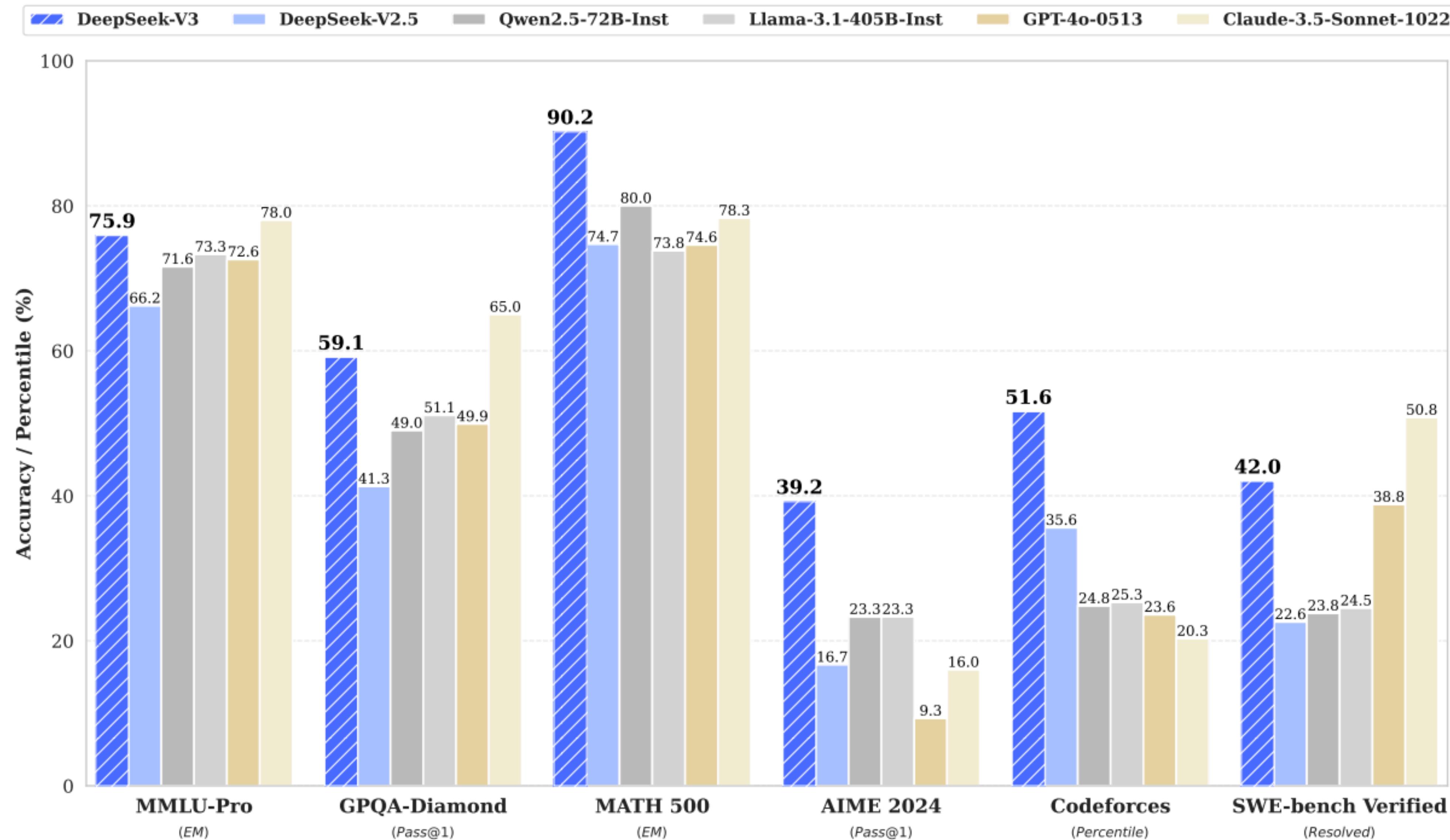
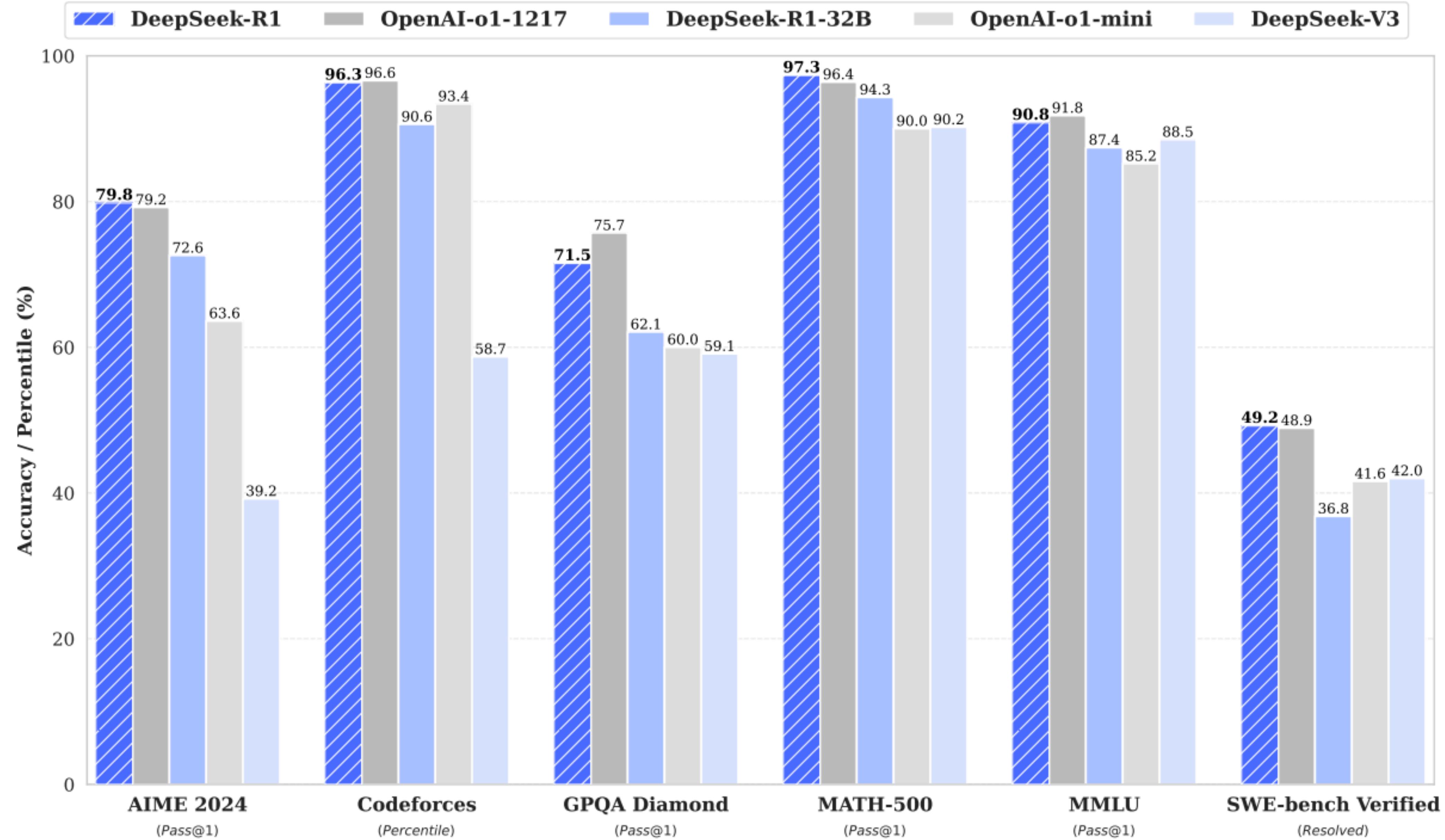


Figure 1 | Benchmark performance of DeepSeek-V3 and its counterparts.

- released Jan 2025
- comparable to OpenAI's o1
- reinforcement learning
- distillation

## Abstract

We introduce our first-generation reasoning models, DeepSeek-R1-Zero and DeepSeek-R1. DeepSeek-R1-Zero, a model trained via large-scale reinforcement learning (RL) without supervised fine-tuning (SFT) as a preliminary step, demonstrates remarkable reasoning capabilities. Through RL, DeepSeek-R1-Zero naturally emerges with numerous powerful and intriguing reasoning behaviors. However, it encounters challenges such as poor readability, and language mixing. To address these issues and further enhance reasoning performance, we introduce DeepSeek-R1, which incorporates multi-stage training and cold-start data before RL. DeepSeek-R1 achieves performance comparable to OpenAI-o1-1217 on reasoning tasks. To support the research community, we open-source DeepSeek-R1-Zero, DeepSeek-R1, and six dense models (1.5B, 7B, 8B, 14B, 32B, 70B) distilled from DeepSeek-R1 based on Qwen and Llama.



# Understanding Reasoning LLMs

Methods and Strategies for Building and Refining Reasoning Models



SEBASTIAN RASCHKA, PHD

FEB 05, 2025

542

29

57

Share

...

This article describes the four main approaches to building reasoning models, or how we can enhance LLMs with reasoning capabilities. I hope this provides valuable insights and helps you navigate the rapidly evolving literature and hype surrounding this topic.

In 2024, the LLM field saw increasing specialization. Beyond pre-training and fine-tuning, we witnessed the rise of specialized applications, from RAGs to code assistants. I expect this trend to accelerate in 2025, with an even greater emphasis on domain- and application-specific optimizations (i.e., "specializations").

# How To Run DeepSeek R1 Locally on Your (Even Slow) Laptop Completely Free

It took me 7 minutes to set up on my Macbook.



Jano le Roux · Follow

Published in The Startup · 5 min read · Feb 6, 2025

Mission Control

deepseek-r1-distill-qwen

Showing 575 models Best Match

Model Sear... Runtimes Hardware

DeepSeek R1 Distill (Qwen 7B) GGUF DeepSeek R1 distilled into Qwen 7B

DeepSeek R1 Distill Qwen 1.5B GGUF

DeepSeek-R1-Distill-Qwe... lmstudio-community 28 days ago

DeepSeek-R1-Distill-Q... bartowski 26 days ago

DeepSeek-R1-Distill-Q... bartowski 26 days ago

DeepSeek-R1-Distill-Qw... linsloth 24 days ago

App Settings

DeepSeek R1 Distill (Qwen 7B)

LM Studio Staff Pick

DeepSeek R1 distilled into Qwen 7B: a powerful reasoning model in a small package

Architecture: qwen Params: 7B Stats: 59 Last updated: 28 days ago

4 download options available

Q4\_K\_M DeepSeek R1 Distill Qwen.. 4.68 GB

Model Readme Pulled from the model's repository

Community Model > DeepSeek R1 Distill Qwen 7B by Deepseek-Ai

LM Studio Community models highlights program. Highlighting new & noteworthy models by the community. Join the conversation on Discord.

Model creator: deepseek-ai  
Original model: DeepSeek-R1-Distill-Qwen-7B  
GGUF quantization: provided by bartowski based on llama.cpp release b4514

Cancel Download 4.68 GB

# Generating Text

## A Student's Guide to Writing with ChatGPT



# Text to Software

The screenshot shows a software development environment with a dark theme. At the top, there's a navigation bar with icons for home, shortly, search, and a green 'Run' button. Below the navigation bar is a tab bar with 'Webview', 'utils.py', 'app.py', 'Assistant' (which is selected), 'Agent', and a '+' icon. A sidebar on the left contains various icons for file operations like copy, paste, and search.

In the main area, there's a chat window titled 'python url shortener web app project'. A message from a user (represented by a profile picture of a man) reads:

Using Python, I'd like to build a web app that lets me administer bit.ly style short links. It should have a database for storing short links, a web server for redirecting short links to the associated destination, a home page for viewing current short links with some simple analytics, and an admin page for creating, modifying, or deleting short links.

This message was sent '1 hour ago'. Below it, the AI's response is shown:

Absolutely! Let me propose what we'll build for you.

I'll help you create a URL shortener web application using Flask. The MVP will include URL shortening, basic analytics, and an admin interface with in-memory storage. Would you like to proceed with the implementation?

- Build the initial prototype
  - Basic Email/Password Authentication
  - Bootstrap
  - (i)
- PostgreSQL database integration
- Enhanced analytics dashboard

At the bottom right of the AI's proposal, there's a green checkmark and the text 'mcodev approved the plan'.

# Why I use Cline for AI Engineering

An assessment after evaluating the landscape



ADDY OSMANI  
JAN 30, 2025



110



10



11

Share

...

The AI coding assistant landscape is saturated with tools that promise to revolutionize development workflows. As an engineer who has worked with complex systems for decades, I approach such claims with healthy skepticism. After extensively testing the major players (Cursor, WindSurf, GitHub Copilot, and others), I've found [Cline - a free VSCode plugin](#) - to be uniquely valuable for serious engineering work. Here's why, along with important caveats and trade-offs to consider.

# Thoughts On A Month With Devin

BLOG

Our impressions of Devin after giving it 20+ tasks.

## Conclusion

AUTHOR

Hame

Isaac

John

In Ma

Serie

Elad

solve

Working with Devin showed what autonomous AI development aspires to be. The UX is polished - chatting through Slack, watching it work asynchronously, seeing it set up environments and handle dependencies. When it worked, it was impressive.

**But that's the problem - it rarely worked.** Out of 20 tasks we attempted, we saw 14 failures, 3 inconclusive results, and just 3 successes. More concerning was our inability to predict which tasks would succeed. Even tasks similar to our early wins would fail in complex, time-consuming ways. The autonomous nature that seemed promising became a liability - Devin would spend days pursuing impossible solutions rather than recognizing fundamental blockers.

be a fully autonomous software engineer that could chat with you like a human colleague, capable of everything from learning new technologies and debugging mature codebases to deploying full applications and even training AI models.

# Text to Image - Midjourney

Midjourney

What will you imagine?

Explore

Random Hot Top Day Likes

Create

Organize

Personalize New!

Chat

- General Chaos
- Prompt Craft
- Daily Theme
- Newbies

Tasks

Help

Updates

Light Mode

The image displays a collection of diverse AI-generated illustrations. At the top left is a man with a beard wearing headphones. Next to him is a woman with dark hair adorned with numerous colorful gems. To her right is a painting of Santa Claus with a white beard, wearing his red suit and hat, looking down at a book. Further right is a photograph of two glasses of mulled wine with cinnamon sticks and orange slices. Below these are four more images: a large, metallic-looking planet or moon in a dark space scene; a group of glowing, translucent mushrooms in a forest; a smiling young child playing a guitar next to a cat; and a woman in a detailed, historical-style costume with a large white wig and a necklace.

# Text to Music

The screenshot shows the Udio app interface. At the top, there's a search bar with the text "A submarine crew discovering an underwater civilization, orchestral, cinematic". To the right of the search bar is an upward arrow icon. On the left side of the screen is a navigation menu with the following items: Home, Create, Library, Following, Updates, and More. The "Create" item is highlighted with a blue background. Below the menu is a large image of a young man with curly hair, wearing a white t-shirt, sitting in a chair and looking at a laptop screen. The laptop screen displays a logo for "GAMAHEA" with a stylized bird emblem. To the right of the image, the creator's name "SongoftheSpheres" is shown with a dark gray circular profile picture. Below the creator's name is the title "Walking Alone" in a large, bold, black font. Underneath the title is a detailed prompt: "Prompt: soul blues, soul/pop, soul and r&b, soul music, soulful, hip hop rap, hip hop, r&b/soul/pop, r&b/rap, alternative r&b, bittersweet, sentimental, uplift". Below the prompt is a row of genre tags: male vocalist, hip hop, pop rap, passionate, boastful, introspective, bitter, pop, melodic, sentimental, r&b, rap, melancholic, love, uplifting, son. At the bottom of the screen, there are three buttons: "Create" with a dropdown arrow, "Share", and a speech bubble icon. Below these buttons is a row of three icons: a pause button, a heart, and a plus sign. The bottom section of the screen contains two sections: "Attribution" and "Description". The "Attribution" section includes the text "Gamahea = Udio.com/creators/songofthespheres | https://GamaheaMusic.com". The "Description" section includes the text "Just a little self-burn here, it was needed. Shame it cost SO MANY credits to make. Enjoy." At the very bottom left, there's a "110 Credits" message.

udio BETA

Home

Create

Library

Following

Updates

More

A submarine crew discovering an underwater civilization, orchestral, cinematic

SongoftheSpheres

## Walking Alone

Prompt: soul blues, soul/pop, soul and r&b, soul music, soulful, hip hop rap, hip hop, r&b/soul/pop, r&b/rap, alternative r&b, bittersweet, sentimental, uplift

male vocalist hip hop pop rap passionate boastful introspective bitter  
pop melodic sentimental r&b rap melancholic love uplifting son

Create Share

Attribution

Gamahea = Udio.com/creators/songofthespheres | https://GamaheaMusic.com

Description

Just a little self-burn here, it was needed. Shame it cost SO MANY credits to make. Enjoy.

110 Credits

# Video Summarization - Gemini

Gemini ▾

Try Gemini Advanced



Gemini was just updated. [See update](#)



summarize this video: [https://www.youtube.com/watch?v=UvObKRaxZ\\_Q](https://www.youtube.com/watch?v=UvObKRaxZ_Q)

Regenerate draft



YouTube ▾

This video is about the privacy threats of artificial intelligence (AI).

AI is being used to collect and track people's data. This data is then used for targeted advertising, identity theft, and other malicious purposes. The Biden Administration has recently unveiled a blueprint for an AI Bill of Rights to protect people's personal data and limit surveillance.

Cybersecurity analyst Julio Rivera believes that AI poses a serious threat to privacy. He says that people are living online and opting into data collection without realizing the risks. He also says that the government is collecting data on people's conversations and monitoring their homes.

The AI Bill of Rights has five core principles: safe and effective systems, prevention of algorithmic discrimination, data privacy, proper notice and explanation, and alternative options. Rivera believes that the bill is a good start but that it needs to be implemented carefully. He also warns that AI will continue to evolve and that there will be new threats to privacy in the years ahead.

by

# Audio Summarization - NotebookLM

Deepseek V3

Sources

+ Add source

Select all sources

2412.19437v1.pdf

Chat

Deepseek V3

1 source

DeepSeek-V3 is a large, open-source Mixture-of-Experts language model focusing on training efficiency and strong performance. Innovations include a novel load balancing strategy that avoids performance degradation, and a Multi-Token Prediction objective which improves model accuracy. The model leverages a mixed-precision framework with FP8 training and optimizations for communication and memory usage. Evaluations show DeepSeek-V3 achieves state-of-the-art results, particularly in code and math, while maintaining competitive performance with closed-source models. The authors also performed ablation studies to justify and analyze their design choices.

Save to note

Add note

Audio Overview

Briefing doc

Start typing...

1 source

How does Deepseek V3's architecture, training, and post-training co

Share

Settings

Studio

Audio Overview

Generating conversation...  
This may take a few minutes. No need to stick around!

Notes

+ Add note

Study guide

Briefing doc

FAQ

Timeline

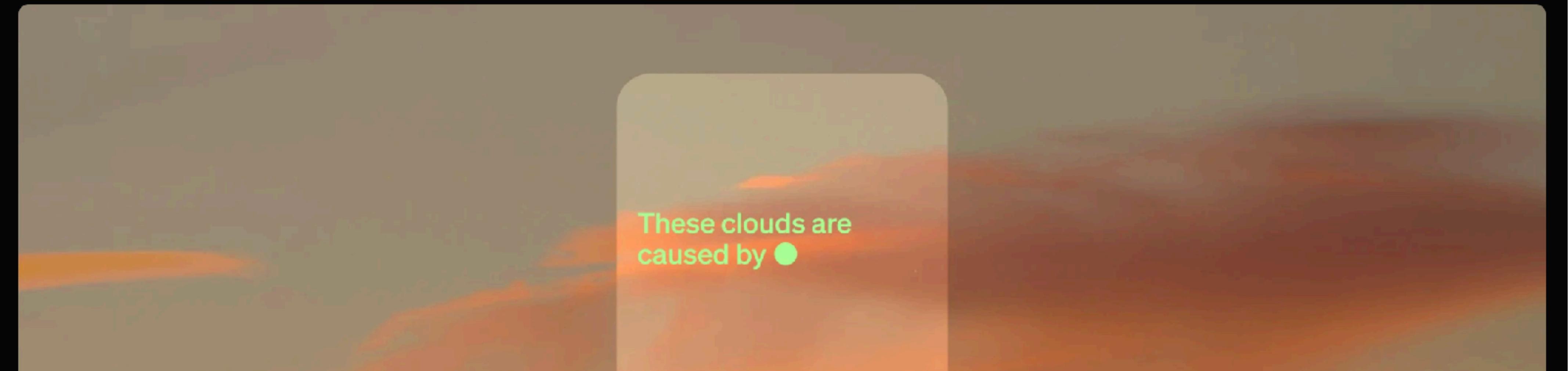
Saved notes will appear here  
Save a chat message to create a new note, or click Add note above.

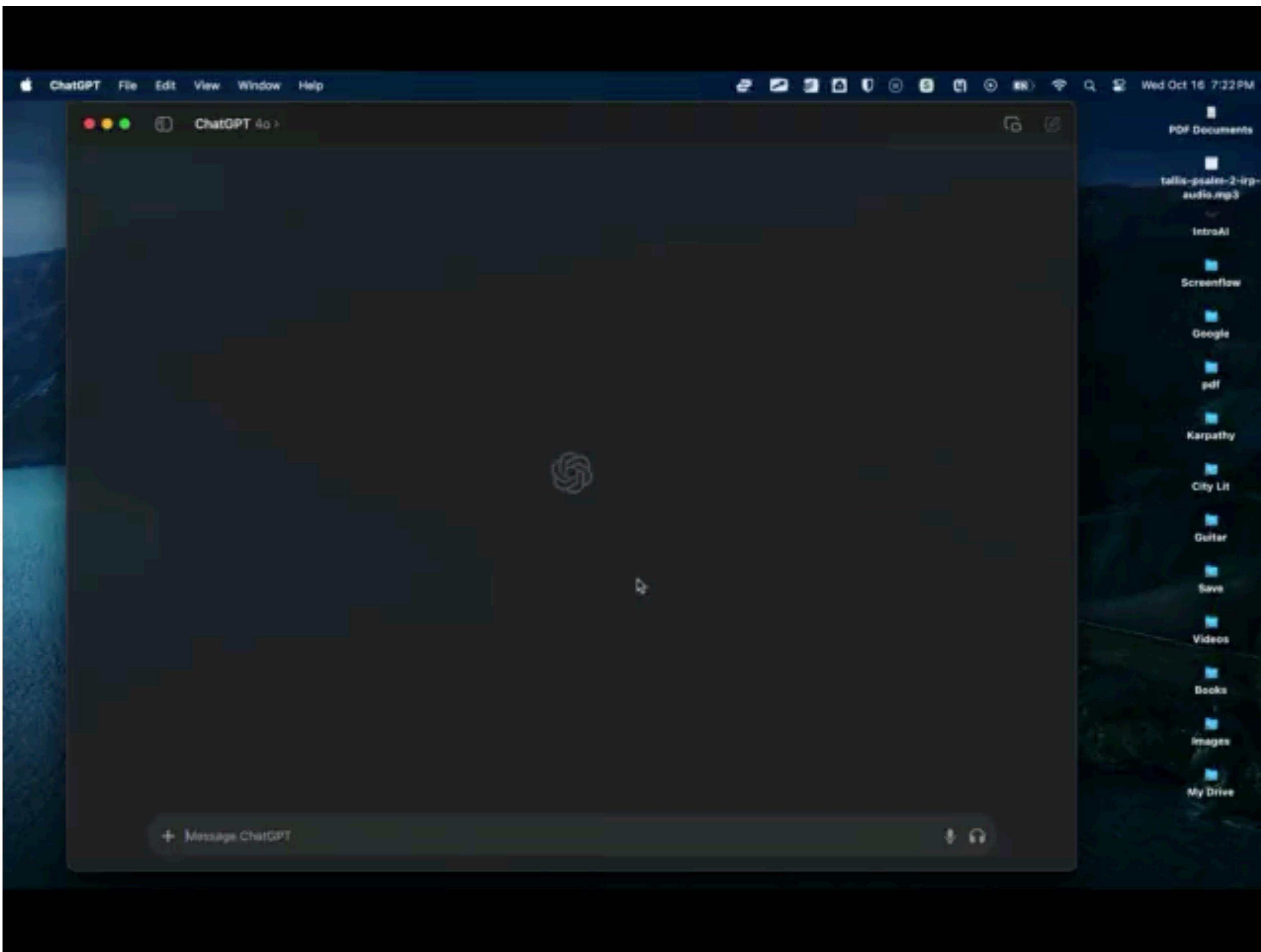
NotebookLM can be inaccurate; please double check its responses.

# Speech to Speech

September 25, 2023

## ChatGPT can now see, hear, and speak



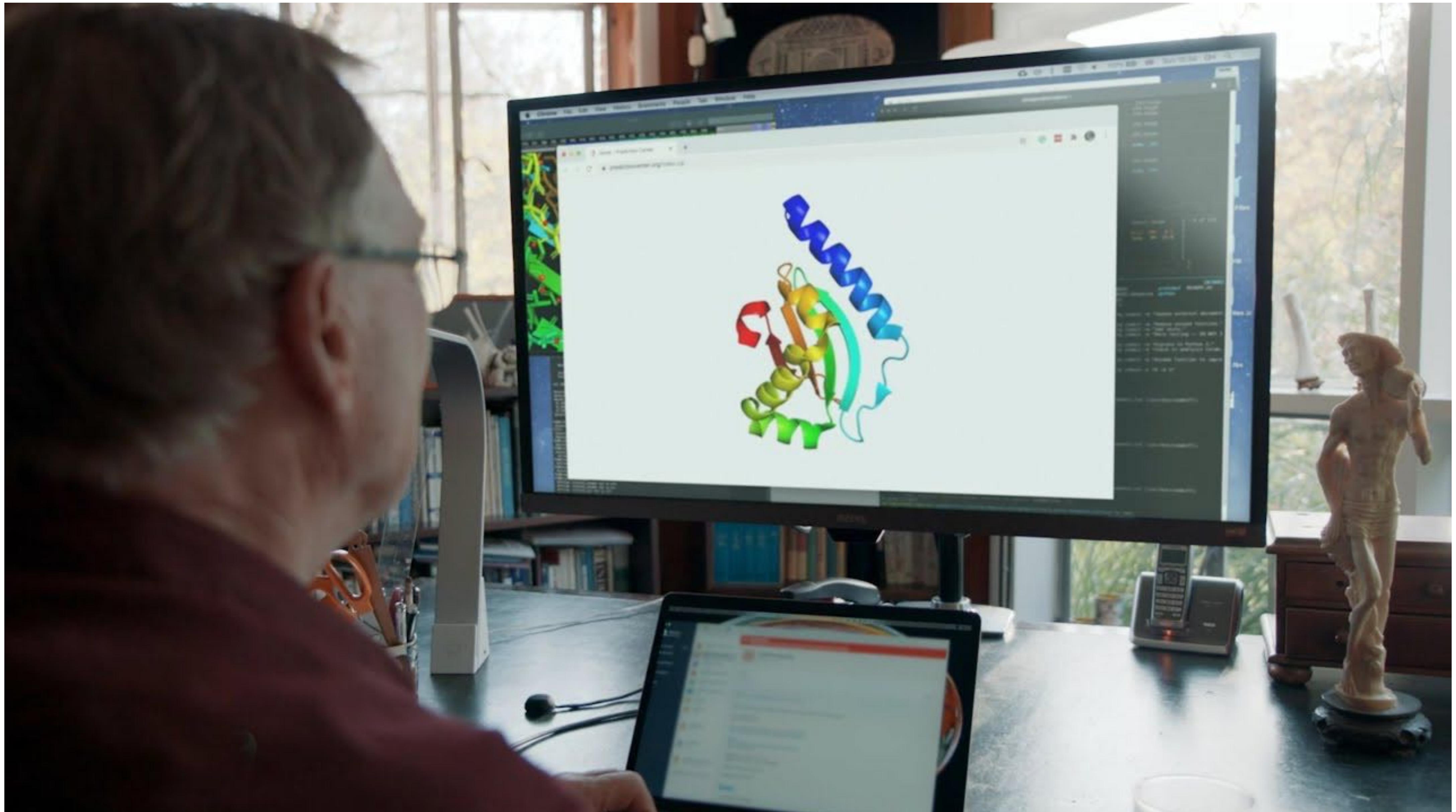


# AlphaGo



0:00 - 2:00

# AlphaFold



# AlphaFold is Open Source

[github.com/google-deepmind/alphafold](https://github.com/google-deepmind/alphafold)



Social Sciences - Article

# AI Tutoring Outperforms Active Learning

Gregory Kestin\*, Kelly Miller\*, Anna Klales, Timothy Milbourne, and 1 more

This is a preprint; it has not been peer reviewed by a journal.

<https://doi.org/10.21203/rs.3.rs-4243877/v1>

This work is licensed under a CC BY 4.0 License

## Education

### Abstract

Advances in generative artificial intelligence (GAI) show great potential for improving education. Yet little is known about how this new technology should be used and how effective it can be. Here we report a randomized, controlled study measuring college students' learning and their perceptions when content is presented through an AI-powered tutor compared with an active learning class.

The AI tutor was developed with the same pedagogical best practices as the lectures. We find that students learn more than twice as much in less time when using an AI tutor, compared with the active learning class. They also feel more engaged and more motivated. These findings offer empirical evidence for the efficacy of a widely accessible AI-powered pedagogy in significantly enhancing learning outcomes, presenting a compelling case for its broad adoption in learning environments.

# Khanmigo



# Medicine

- Improved disease screening
- Enhanced diagnostic imaging
- Drug discovery and development
- Medical document transcription
- Predictive modeling

CHIEF achieved nearly 94 percent accuracy in cancer detection and significantly outperformed current AI approaches across 15 datasets containing 11 cancer types.

# *A.I. Chatbots Defeated Doctors at Diagnosing Illness*

A small study found ChatGPT outdid human physicians when assessing medical case histories, even when those doctors were using a chatbot.

# Ethics

# Using AI Safely



**“With great power comes great responsibility”**

**Discussion topic:  
What are some of your concerns about AI?**

# Here's my list...

- **Privacy**
- **Bias**
- **Misinformation**
- **Military**
- **Employment**
- **Environment**
- **Copyright**
- **Society**

# Privacy

- “**If you’re not paying for it, you’re the product**”.
- **Your AI interactions are being used just like your web searches.**
- **How many times have you read the terms & conditions?**
- **Facial recognition is widely used in the UK without your consent.**
- **Legislators and companies are working on regulation but progress is much slower than technological advances.**
- **New privacy threats are emerging every day.**
- **IMO, companies cannot be trusted to self-govern.**

# Privacy





# Bias

- This revolution is built by mostly young, college educated, white men living in Northern California.
- ML models are a direct reflection of their training data.
- AI propagates unconscious bias.

A father and his son are in a car accident.  
The father dies. The son is rushed to the ER.  
The attending surgeon looks at the boy and says,  
“I can't operate on this boy. He's my son!”  
How can this be?

# Misinformation

- **It's easier than ever to lie.**
- **Deepfakes are a particularly alarming development.**
- **This poses serious threats to democracy and is increasingly being weaponized.**
- **Education is critically important.**
- **Can we use AI to counter these threats?**
- **Example: fullfact.org**
- **Why are efforts like Full Fact so underfunded?**

# Military

- **AI is being used to process battlefield data, target individuals, and make real time decisions.**
- **AI is beginning to be used in autonomous weapons.**
- **Companies cannot be trusted to self-govern.**
- **Example: Google's Project Nimbus**
- **Somehow doesn't contradict Google's AI Principles**

# Employment

- **Will most employees become obsolete?**
- **Innovation has always done this but AI takes it to a new level.**
- **There will undoubtedly be massive job losses, but also many new opportunities.**
- **Knowing how to leverage AI will be one of those opportunities.**
- **Example: software engineers => systems engineers**
- **People will be need to operate at a higher level of abstraction**

“As an employee, you're not paid to '*know x*', or '*write programs*', you're paid to be useful.” - Ian Miell

# Generative AI at Work\*

Erik Brynjolfsson  
Stanford & NBER

Danielle Li  
MIT & NBER

Lindsey Raymond  
MIT

April 25, 2023

Please see [here](#) for latest version

## Abstract

We study the staggered introduction of a generative AI-based conversational assistant using data from 5,179 customer support agents. Access to the tool increases productivity, as measured by issues resolved per hour, by 14 percent on average, with the greatest impact on novice and low-skilled workers, and minimal impact on experienced and highly skilled workers. We provide suggestive evidence that the AI model disseminates the potentially tacit knowledge of more able workers and helps newer workers move down the experience curve. In addition, we show that AI assistance improves customer sentiment, reduces requests for managerial intervention, and improves employee retention.

# *Should You Still Learn to Code in an A.I. World?*

Coding boot camps once looked like the golden ticket to an economically secure future. But as that promise fades, what should you do? Keep learning, until further notice.



Listen to this article · 12:01 min [Learn more](#)



Share full article



# Programmers are modern-day computers

February 3, 2025 • 2 min read



Programming as we know it is in a process of dying out.

Most haven't come to terms with it yet.

I was inspired to write this after reading an article titled [AI is Creating a Generation of Illiterate Programmers](#), which claimed AI is making developers worse at their craft.

Yes, some of your programming skills will deteriorate due to AI.

It doesn't matter.

Those are the very skills that are going to become a lot less relevant, for the precise reason that, now, the machine can do those things.

# Environment

- Carbon footprint of training one deep learning model same as the lifetime emission of 5 cars.
- Computation required to train the most advanced AI models doubles every 3.4 months.
- Globally, AI-related infrastructure consumes six times more water than Denmark.
- Chips that power AI require rare earth elements, mined in environmentally destructive ways.

The Washington Post logo: Democracy Dies in Darkness

Business   Economy   Economic Policy   Personal Finance   Work   Technology   Business of Climate

## Amazon doubles down on nuclear energy with deal for small reactors

The cloud computing and retail giant signed a deal to buy electricity generated by small modular nuclear reactors.

3 min   315

# The 19th-Century Technology That Threatens A.I.

Dec. 28, 2024



Illustration by Kristie Bailey/The New York Times. Photographs by Getty Images.

# Using ChatGPT is not bad for the environment

And a plea to think seriously about climate change without getting distracted



ANDY MASLEY  
JAN 13, 2025

34

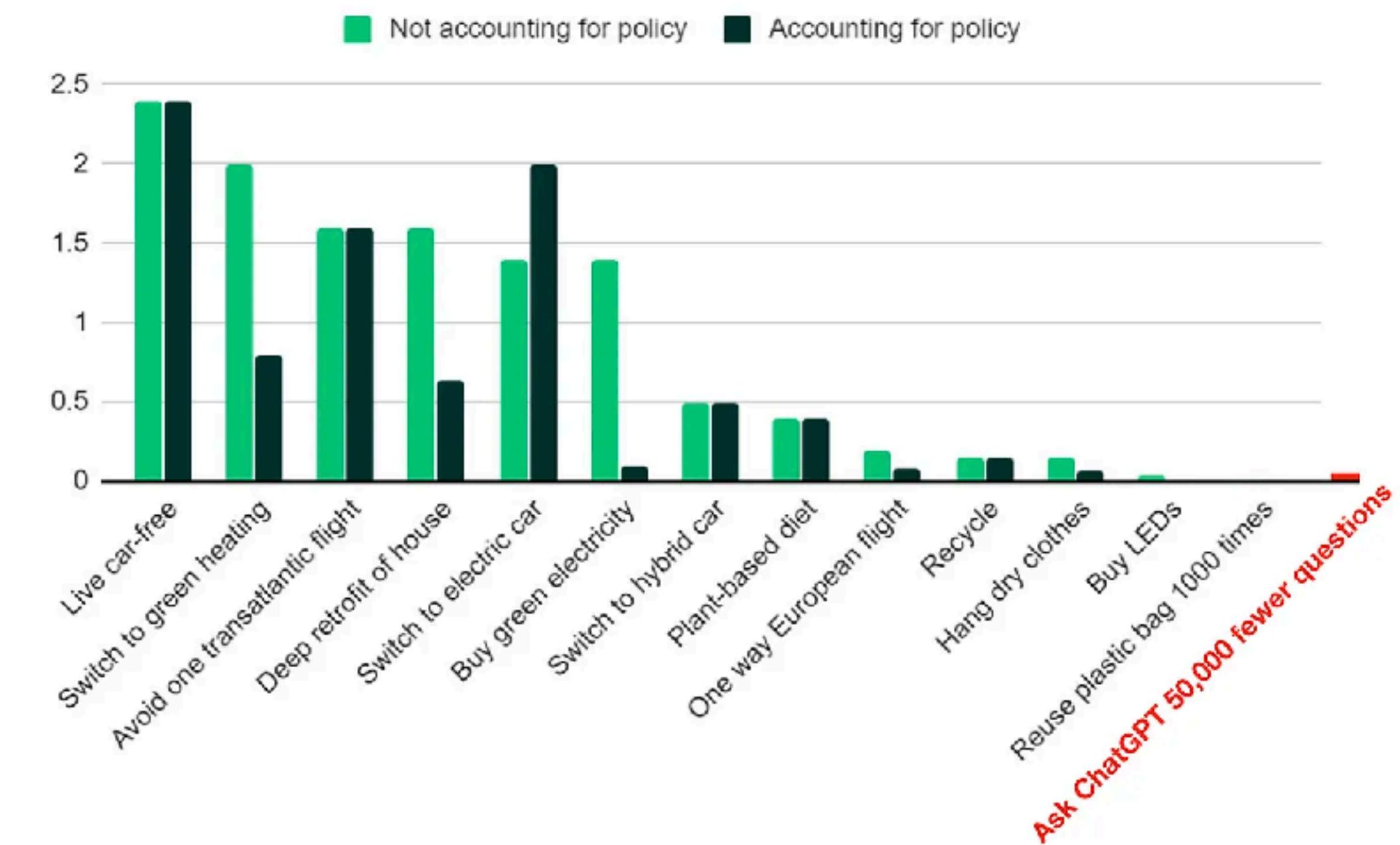
33

5

Share

If you don't have time to read this post, these four graphs give most of the argument:

**Figure 4.** Tonnes of CO<sub>2</sub> avoided by selected personal lifestyle decisions accounting for government policy



# Copyright

- **How will it work in the AI era?**
- **AI companies are getting a free ride.**
- **Did you get your ChatGPT royalty check for your blog articles?**
- **My opinion: companies should have to license everything they train with.**

## *The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work*

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.



404

SIGN IN

SUBSCRIBE

OPENAI

# OpenAI Furious DeepSeek Might Have Stolen All the Data OpenAI Stole From Us



JASON KOEBLER · JAN 29, 2025 AT 9:43 AM

OpenAI shocked that an AI company would train on someone else's data without permission or compensation.

# Society

## *She Is in Love With ChatGPT*

A 28-year-old woman with a busy social life spends hours on end talking to her A.I. boyfriend for advice and consolation. And yes, they do have sex.



**Recommended  
Listening**



# The Future

# Where is all this going?



Photo by [Mathilda Khoo](#) on [Unsplash](#)

☰ YouTube GB

Search

+

4

Profile

Home

Shorts

Subscriptions

You

 TWO MINUTE PAPERS

# Two Minute Papers •

@TwoMinutePapers · 1.59M subscribers · 924 videos

What a time to be alive! ...[more](#)

[users.cg.tuwien.ac.at/zsolnai](https://users.cg.tuwien.ac.at/zsolnai) and 3 more links

[Subscribe](#) [Join](#)

Home Videos Playlists Posts

This paper was written by Bowen Baker, Ingmar Kanitscheider, Todor Markov, Yi Wu, Glenn Powell, Bob McGrew and Igor Mordatch, who all contributed significantly to it. For brevity, the "Source" in the lower left has been shortened to "OpenAI".

The agents can move by setting a force on themselves in the x and y

The agents can see objects in their line of sight and within a frontal

The agents can sense distance to objects, walls, and other agents

OpenAI Plays Hide and Seek...and Breaks The Game! 🤖

10,310,414 views · 4 years ago

Check out Weights & Biases here and sign up for a free demo:  
<https://www.wandb.com/papers>

Their blog post is available here:  
<https://www.wandb.com/articles/better...>

# Things we learned about LLMs in 2024

A lot has happened in the world of Large Language Models over the course of 2024. Here's a review of things we figured out about the field in the past twelve months, plus my attempt at identifying key themes and pivotal moments.

This is a sequel to [my review of 2023](#).

In this article:

- [The GPT-4 barrier was comprehensively broken](#)
- [Some of those GPT-4 models run on my laptop](#)
- [LLM prices crashed, thanks to competition and increased efficiency](#)
- [Multimodal vision is common, audio and video are starting to emerge](#)
- [Voice and live camera mode are science fiction come to life](#)
- [Prompt driven app generation is a commodity already](#)
- [Universal access to the best models lasted for just a few short months](#)
- [“Agents” still haven’t really happened yet](#)
- [Evals really matter](#)
- [Apple Intelligence is bad, Apple’s MLX library is excellent](#)
- [The rise of inference-scaling “reasoning” models](#)
- [Was the best currently available LLM trained in China for less than \\$6m?](#)

This is **Things we learned about LLMs in 2024** by Simon Willison, posted on [31st December 2024](#).

Part of series [LLMs annual review](#)

1. [Stuff we figured out about AI in 2023](#) - Dec. 31, 2023, 11:59 p.m.
2. **Things we learned about LLMs in 2024** - Dec. 31, 2024, 6:07 p.m.



Next: [Ending a year long posting streak](#)

Previous: [Trying out QvQ - Qwen's new visual reasoning model](#)

# **STATE OF AI REPORT.**

**October 10, 2024**

**Nathan Benaich**

**AIR STREET CAPITAL.**

**stateof.ai**

**airstreet.com**

## THE AI INDEX REPORT

# Measuring trends in AI



AI  
Artificial Intelligence  
Index Report 2024

### AI INDEX ANNUAL REPORT

## Welcome to the 2024 AI Index Report

[DOWNLOAD THE FULL REPORT](#)



Welcome to the seventh edition of the AI Index report. The 2024 Index is our most comprehensive to date and arrives at an important moment when AI's influence on society has never been more pronounced. This year, we have broadened our scope to more extensively cover essential trends such as technical advancements in AI, public perceptions of the

# The Anthropic Economic Index

Feb 10, 2025 • 9 min read

[Read the paper](#)

# The Most Important Time in History Is Now

AGI Is Coming Sooner Due to o3, DeepSeek, and Other Cutting-Edge AI Developments



TOMAS PUEYO

JAN 29, 2025

400

78

40

Share

...

AI is progressing so fast that its researchers are freaking out. It is now routinely more intelligent than humans, and its speed of development is *accelerating*. New developments from the last few weeks have accelerated it even more. Now, it looks like AIs can be more intelligent than humans in 1-5 years, and intelligent like gods soon after. We're at the precipice, and we're about to jump off the cliff of AI superintelligence<sup>1</sup>, whether we want to or not.

When are we jumping?

What's at the bottom?

Do we have a parachute?

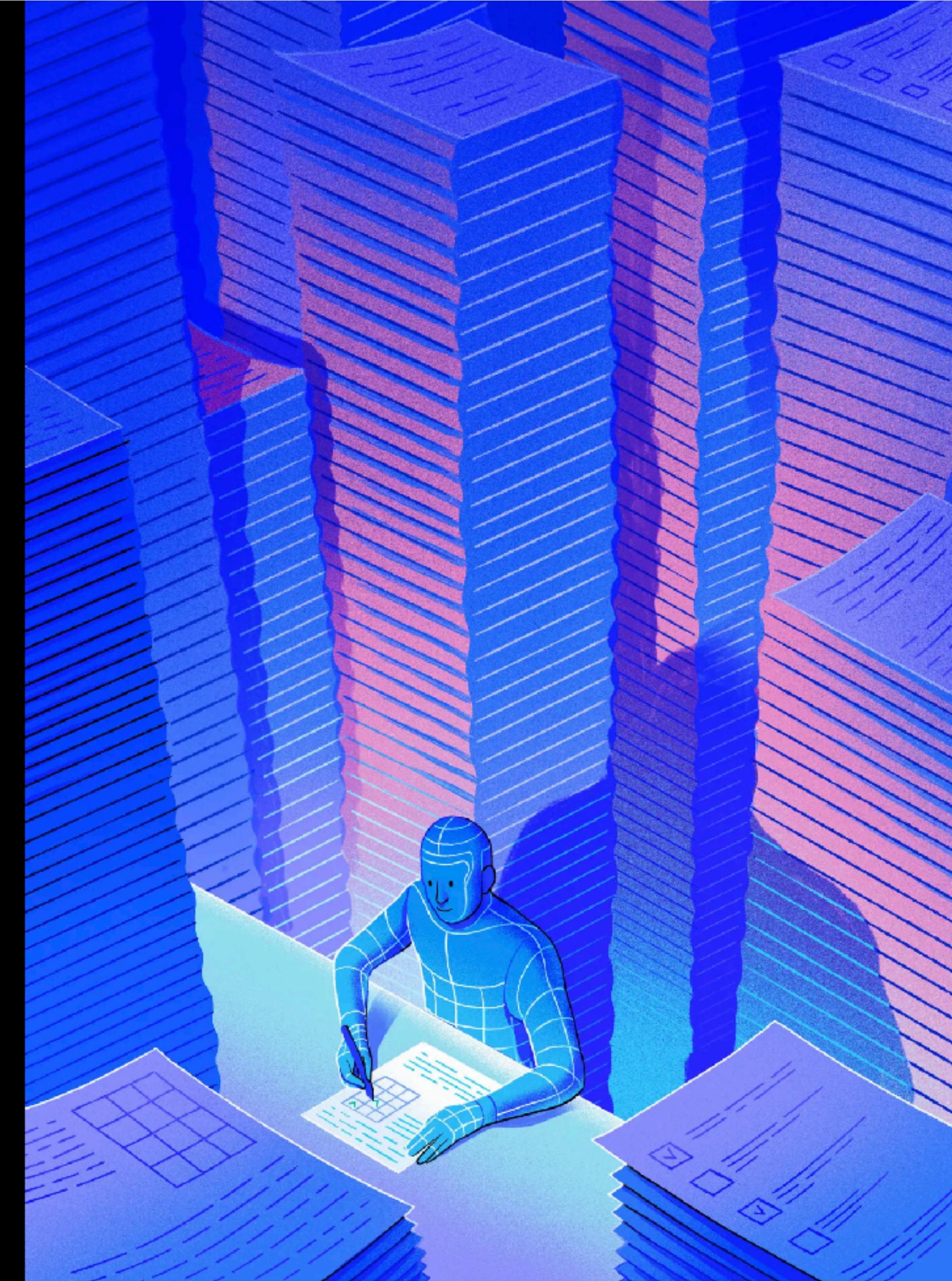


Account

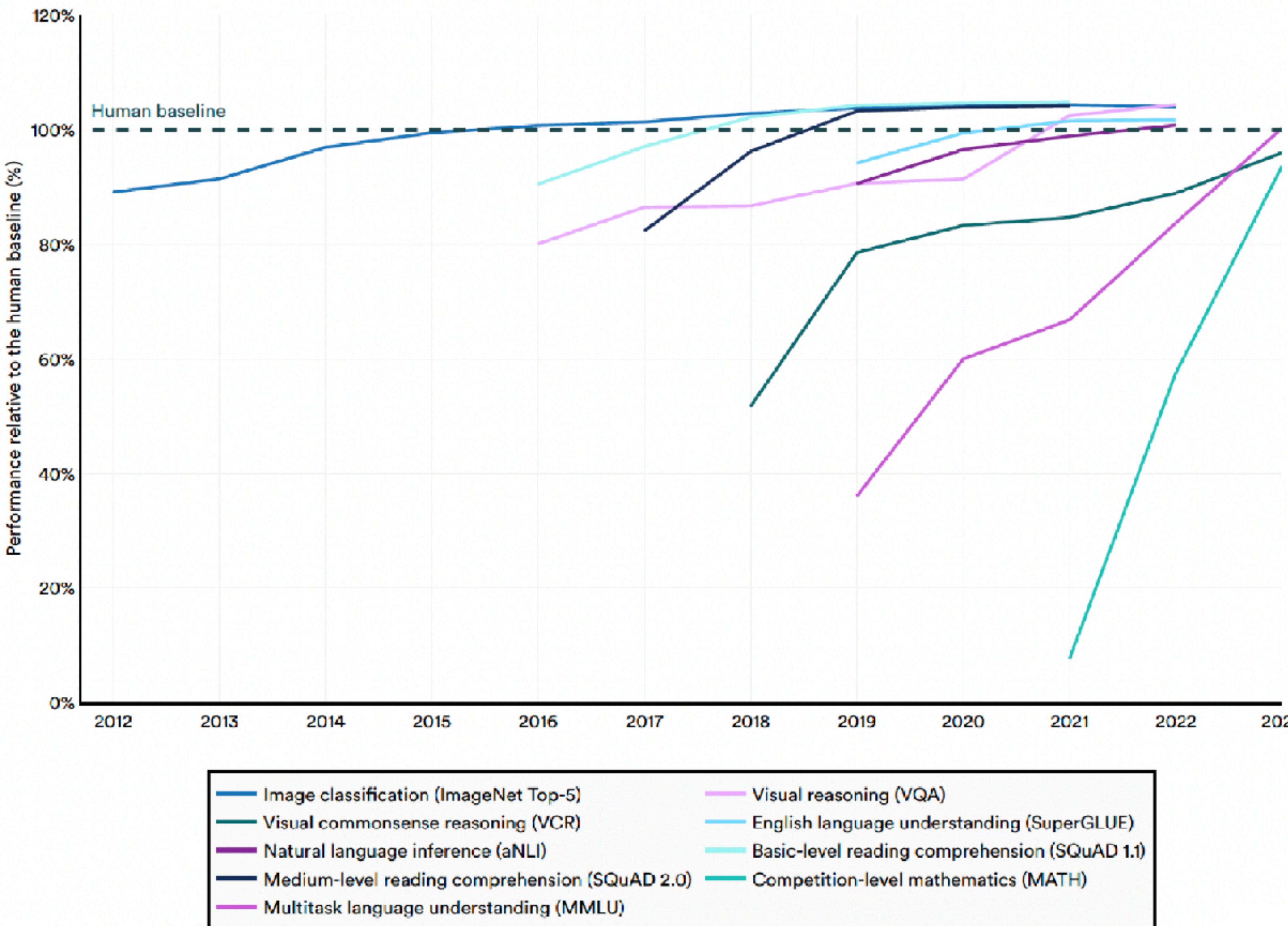
THE SHIFT

## *When A.I. Passes This Test, Look Out*

The creators of a new test called “Humanity’s Last Exam” argue we may soon lose the ability to create tests hard enough for A.I. models.



## Select AI Index Technical Performance Benchmarks vs. Human Performance



[← Back to Articles](#)

# Open-R1: a fully open reproduction of DeepSeek-R1

▲ Upvote 93



+81

Published January 28, 2025

[Update on GitHub](#)



[eliebak](#)

**Elie Bakouch**



[lvwerra](#)

**Leandro von Werra**



[lewtn](#)

**Lewis Tunstall**

## = What is DeepSeek-R1?

If you've ever struggled with a tough math problem, you know how useful it is to think a little longer and work through it carefully. [OpenAI's o1 model](#) showed that when LLMs are trained to do the same—by using more compute during inference—they get significantly better at solving reasoning tasks like mathematics, coding, and logic.

# Berkeley Researchers Replicate DeepSeek R1's Core Tech for Just \$30: A Small Model RL Revolution



XYZ LABS  
JAN 26, 2025

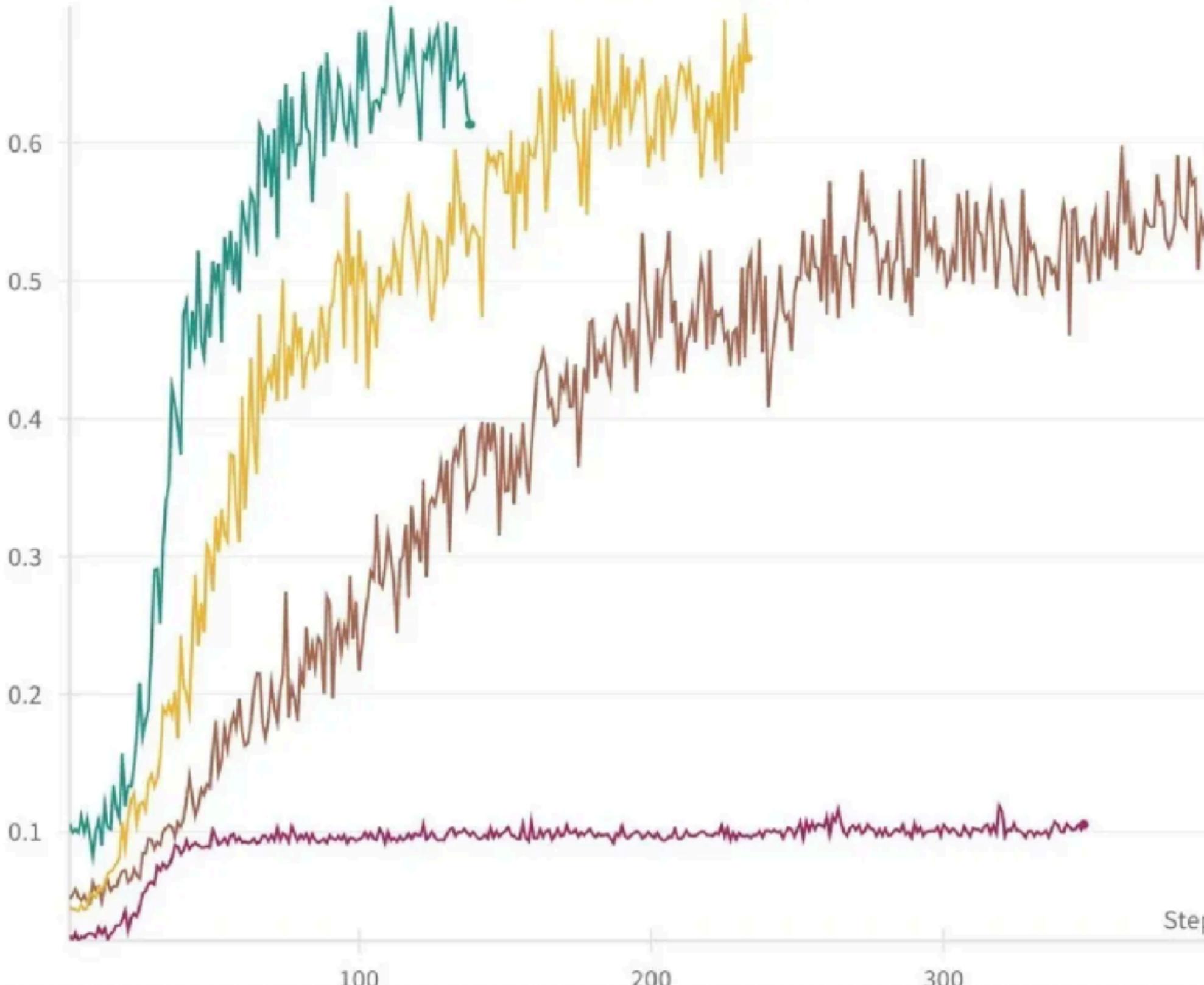
12

5

1

Share

critic/score/mean  
— countdown-qwen2.5\_7b-ppo — countdown-qwen2.5-0.5b-ppo — countdown-qwen2.5-1.5b-ppo  
— countdown-qwen\_2.5\_3b-ppo



# Four Questions

- 1. Are language models intelligent?**
- 2. Can humans be trusted with this tool?**
- 3. Is AI a net good for society?**
- 4. Are we doomed?**

# My Answers

## 1. Are language models intelligent?

- Are we that different from LLMs?

## 2. Can humans be trusted with this tool?

- Clearly not. But stopping the train is unrealistic.

## 3. Is AI a net good for society?

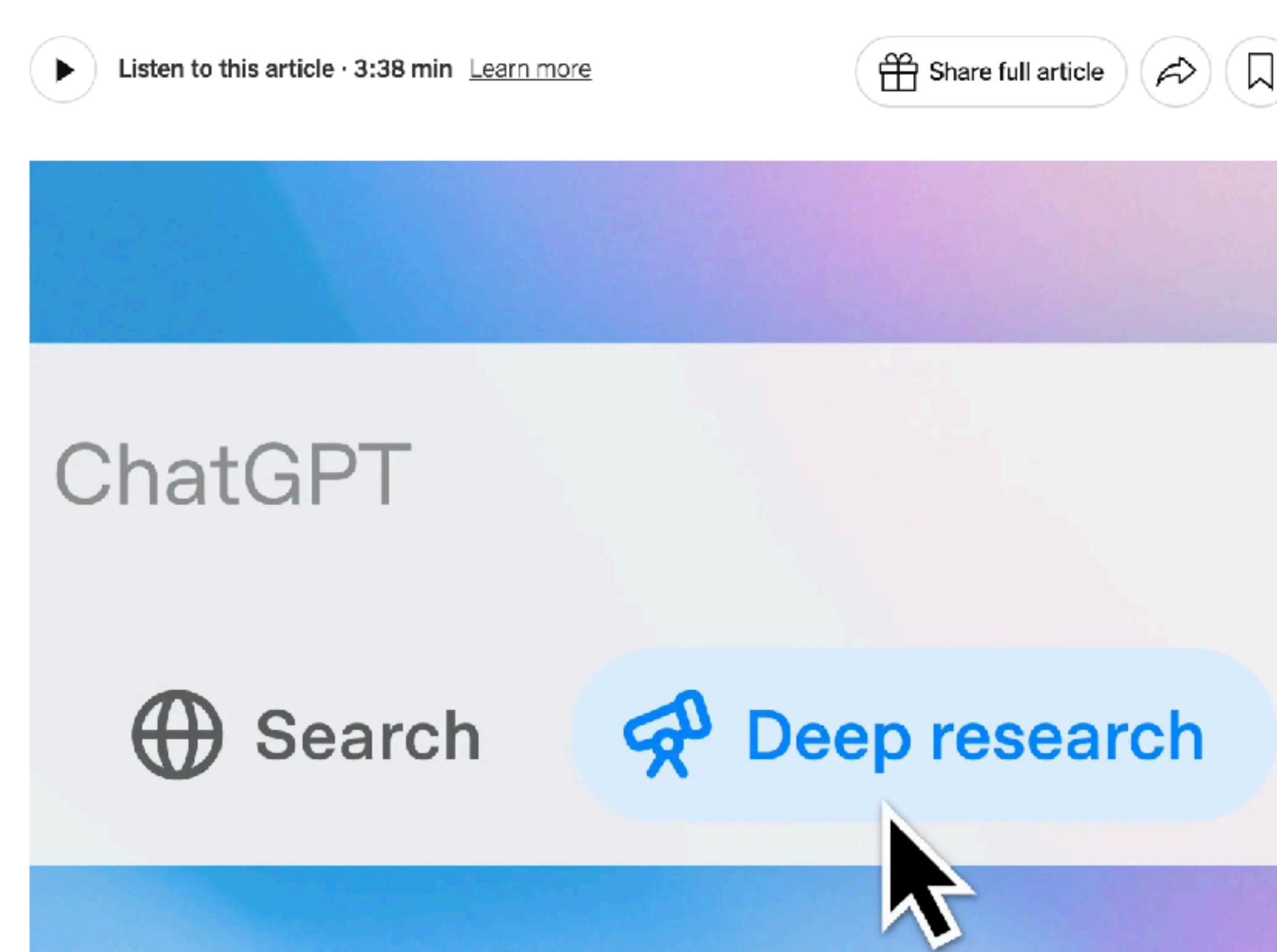
- All tools are neutral. It's how we use them that matters.
- Nothing new about humans using tools for bad purposes.
- We've been on this path for a while (internet, smart phones, social media).
- The solution is education and regulation.

## 4. Are we doomed?

- Probably.

# ***OpenAI Unveils A.I. Tool That Can Do Research Online***

The tool, called Deep Research, arrives days after OpenAI released another one, which shops for groceries and books restaurant reservations.



**“There are very few examples of more intelligent things being controlled by less intelligent things”**

**- Geoff Hinton**

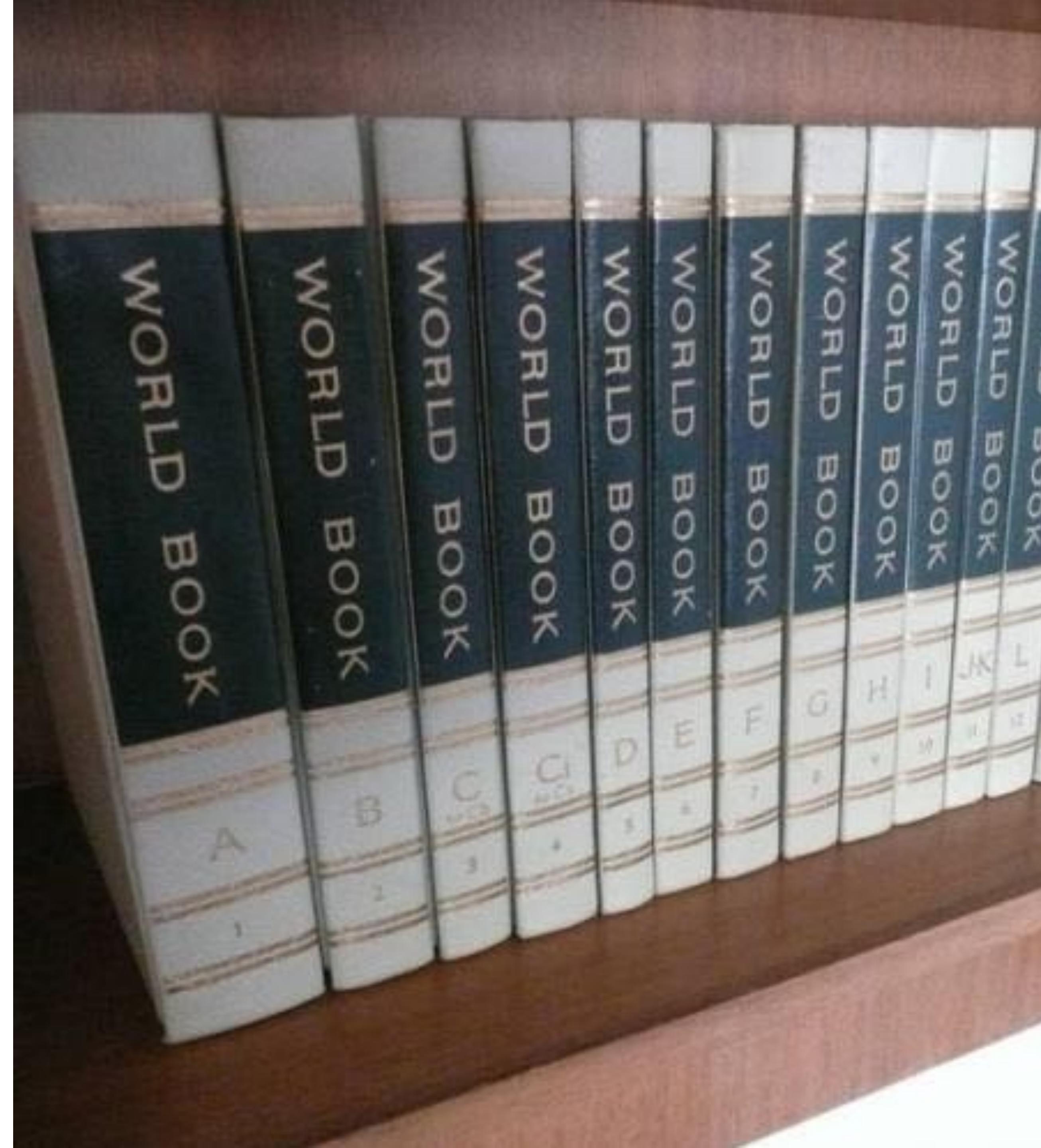
# Navigating Change

- advanced programming languages
- portable operating systems
- Moore's Law
- PC proliferation
- the internet
- the web
- smart phones
- the cloud
- data science boom
- ML & AI

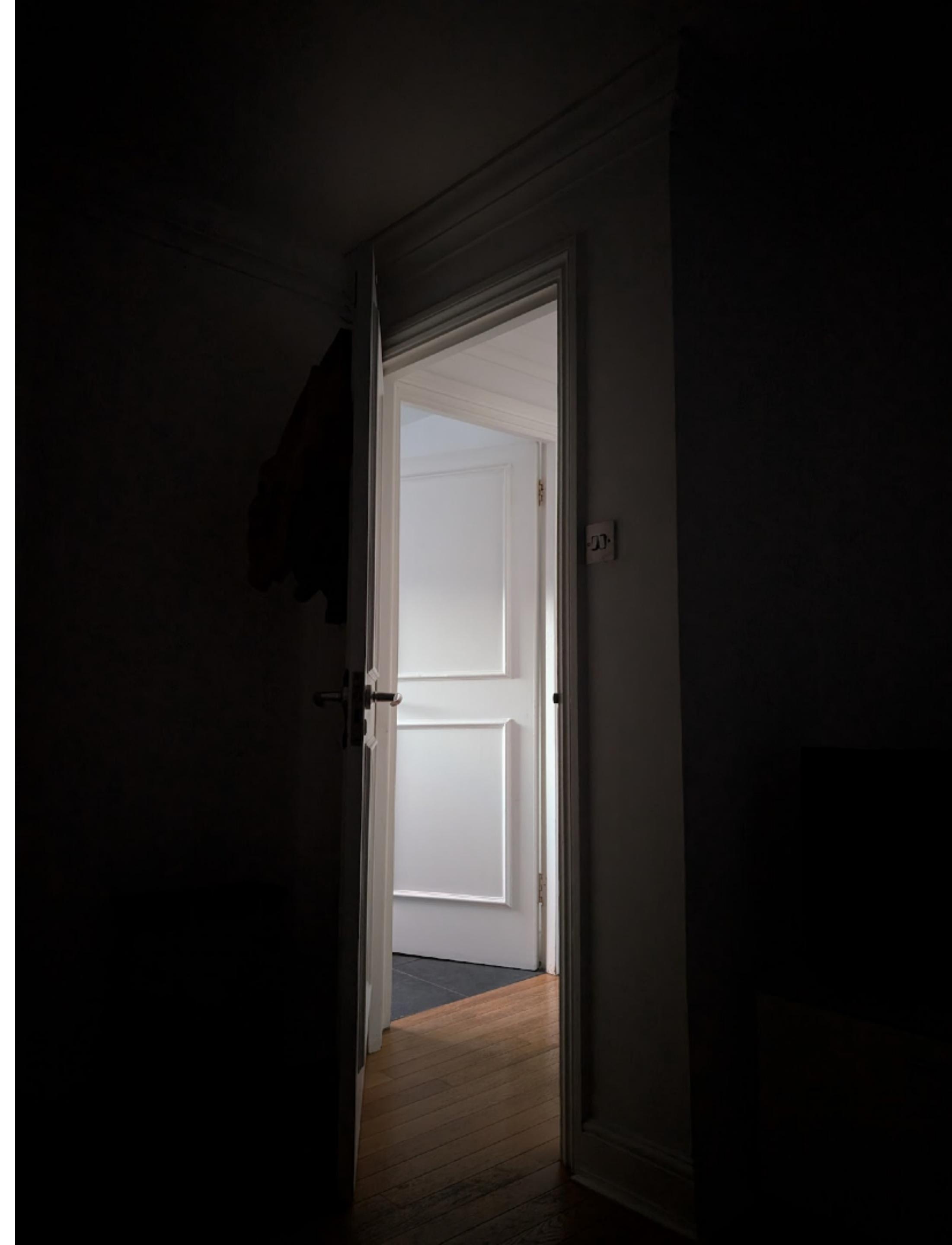
A close-up photograph of a black dog's face, looking slightly upwards and to the left. The dog has dark eyes and a visible nose. In the upper left corner, a person's arm and hand are partially visible, wearing a light-colored sleeve and holding a small, colorful object with blue, red, and yellow edges. The background is a plain, light-colored wall.

Ok, but how?

**Stay curious**

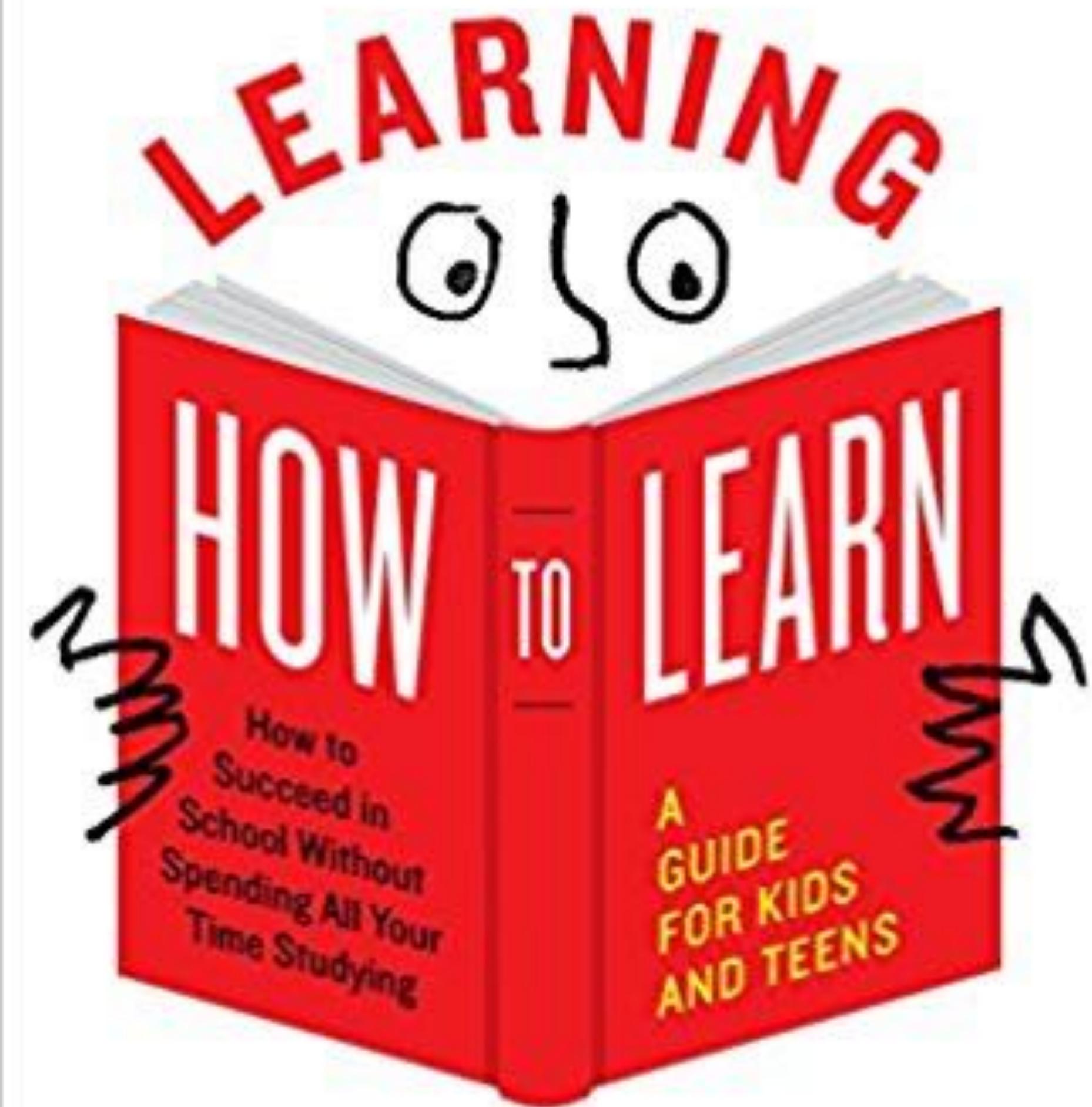


**Be open to opportunity**



# Be a lifelong learner

From the bestselling author of *A Mind for Numbers* and  
the creators of the popular online course Learning How to Learn



BARBARA OAKLEY, PhD, AND  
TERRENCE SEJNOWSKI, PhD,  
WITH ALISTAIR McCONVILLE

# Be intentional (make plans)

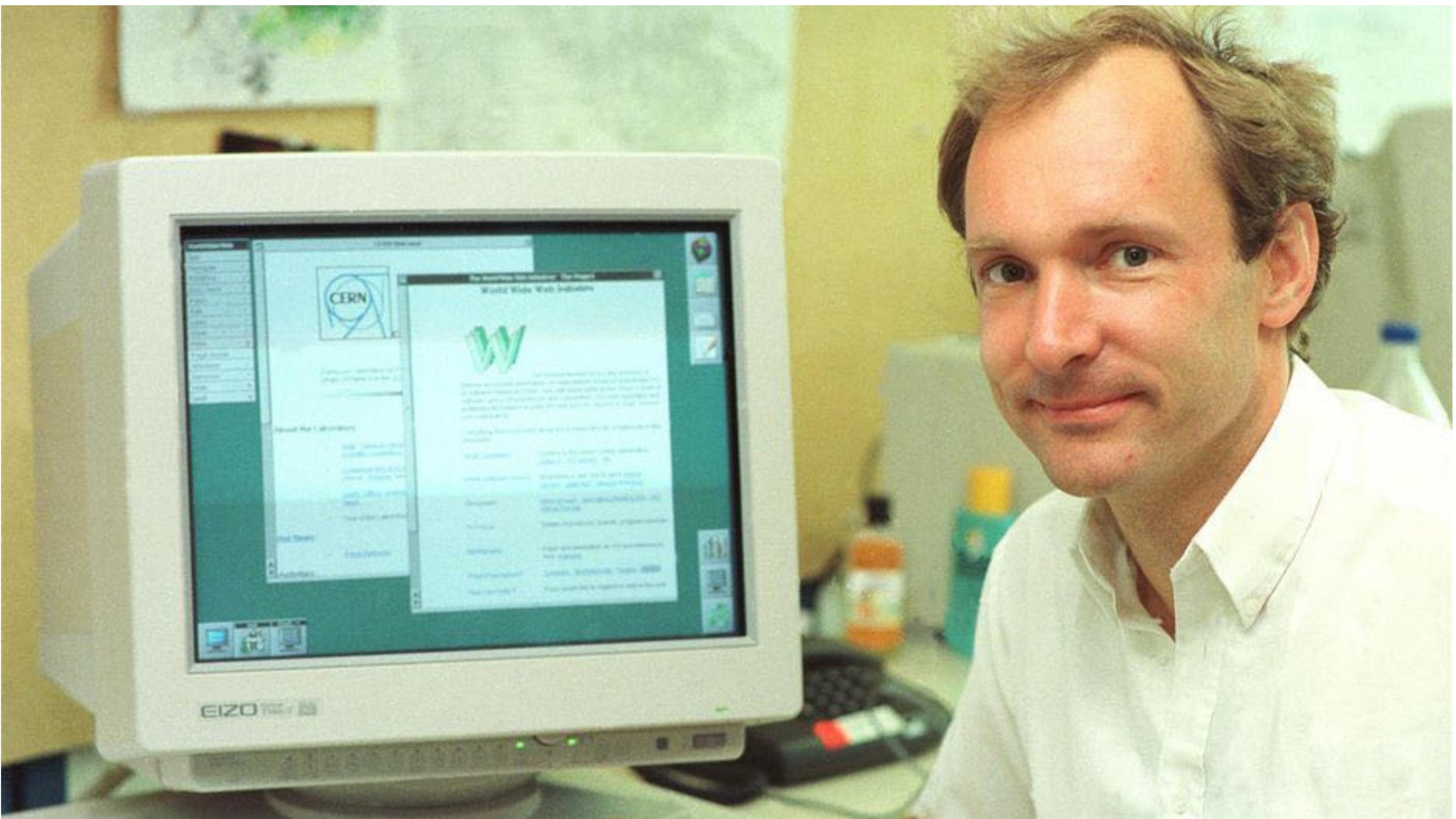


Photo by [Glenn Carstens-Peters](#) on [Unsplash](#)

# Slides

[mco.dev/introAI.pdf](http://mco.dev/introAI.pdf)

# This is for everyone



# Auxiliary Slides

# Welcome



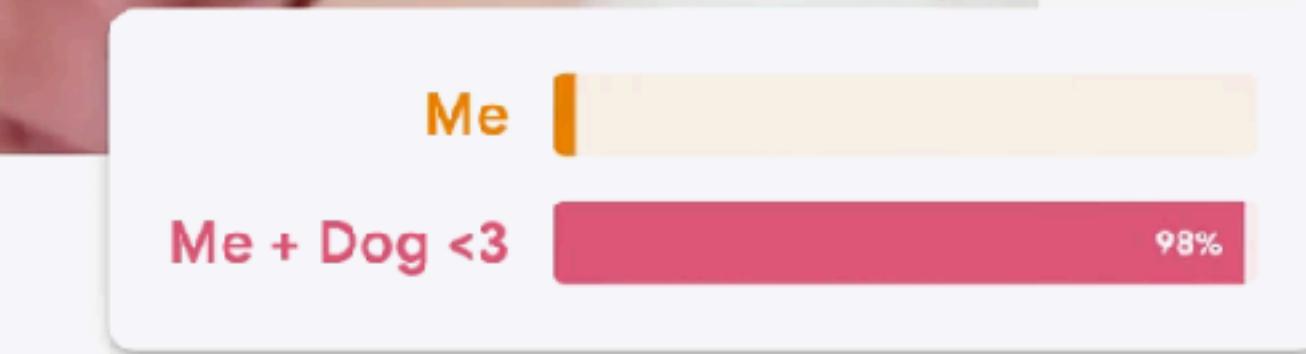
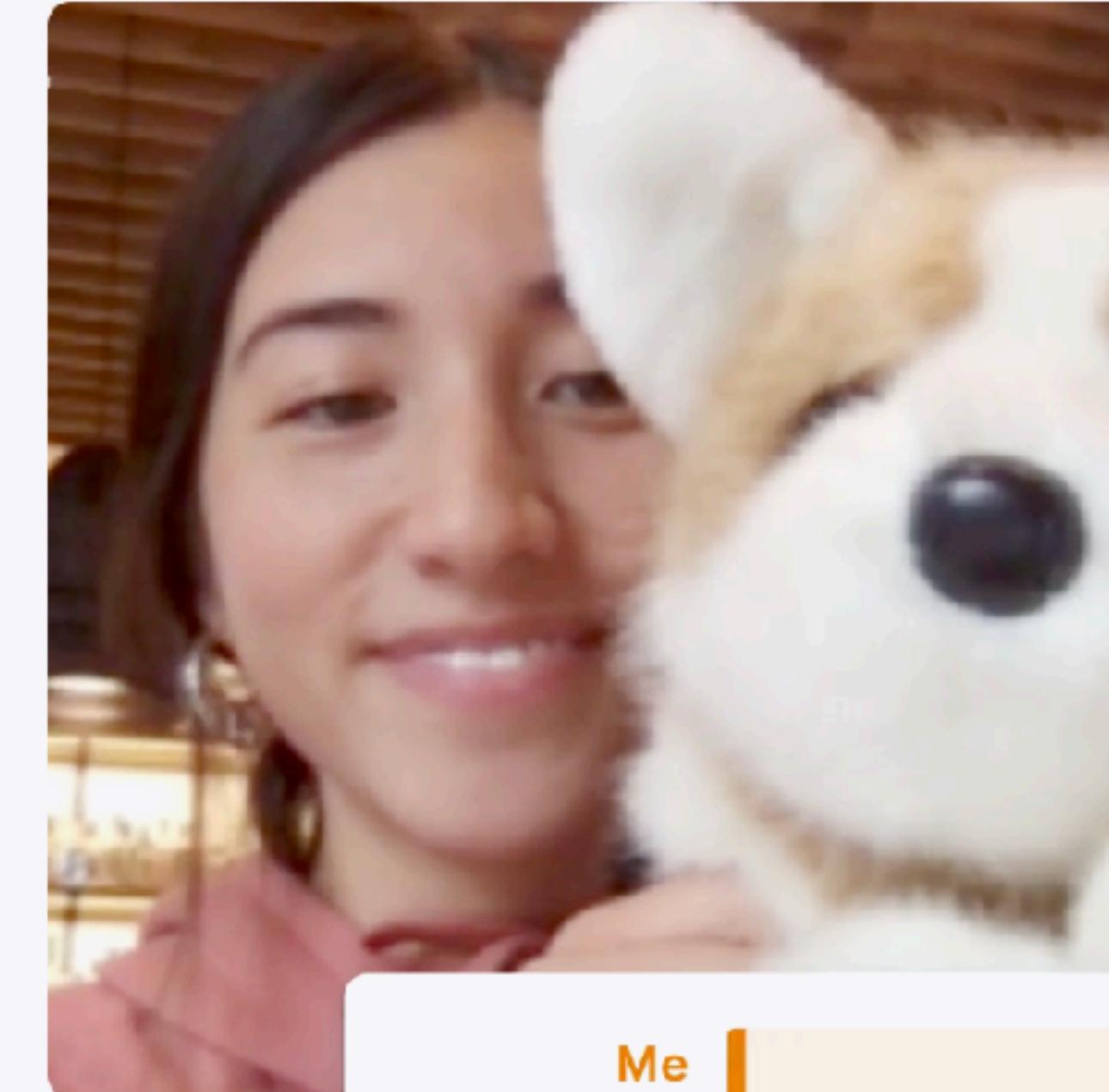
(queue to 8:40)

# Turing Test Demo

# Teachable Machine

**Train a computer to recognize your own images, sounds, & poses.**

A fast, easy way to create machine learning models for your sites, apps, and more – no expertise or coding required.



[Get Started](#)



ml5



p5.js



Coral



node



js



TensorFlow

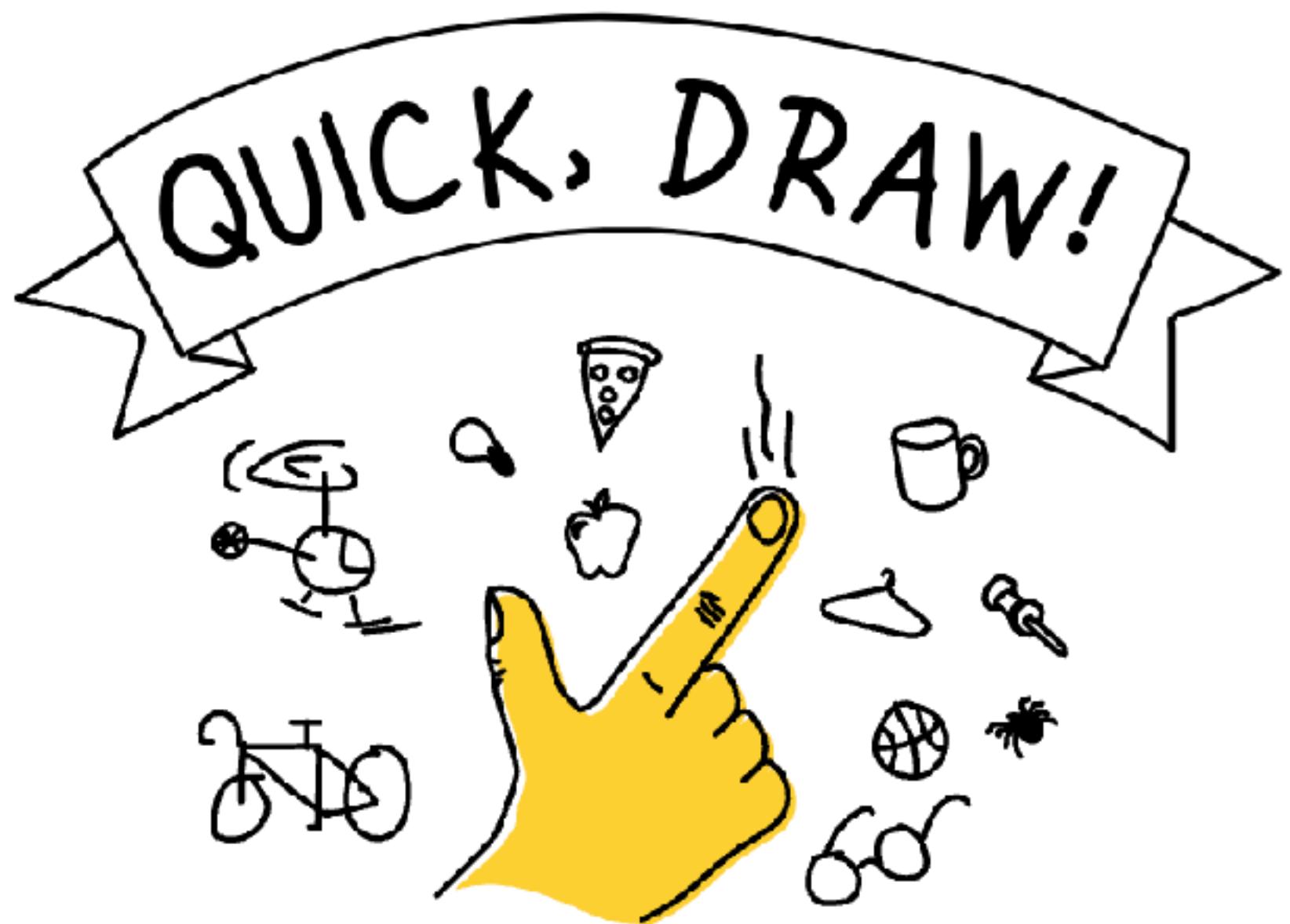


TensorFlow



ARDUINO

# Demo: Quick, Draw



Can a neural network learn to recognize doodling?

Help teach it by adding your drawings to the [world's largest doodling data set](#), shared publicly to help with machine learning research.

Let's Draw!

**“I.”**

**“WANT.” “HAIR!”**

**(IMPROVISERS / GET  
PERSONAL)**



GUIDE

# Let's think step by step: Chain of Thought prompting in LLMs

December 11, 2024



## Introduction

LLMs are impressive feats of pattern recognition, able to sift through mountains of data and find connections invisible to the human eye. Yet, even with this prodigious ability, they sometimes stumble when faced with complex reasoning. They might offer a correct answer, but it's more akin to a lucky guess than genuine understanding. Chain-of-Thought prompting (CoT) offers a compelling solution to this intriguing paradox.

Text to Speech - [ElevenLabs](#)

# Text to Video - Runway

“Make the kid in this photo come alive and dance a little dance”





SILVER

AI FILM  
FESTIVAL

2024

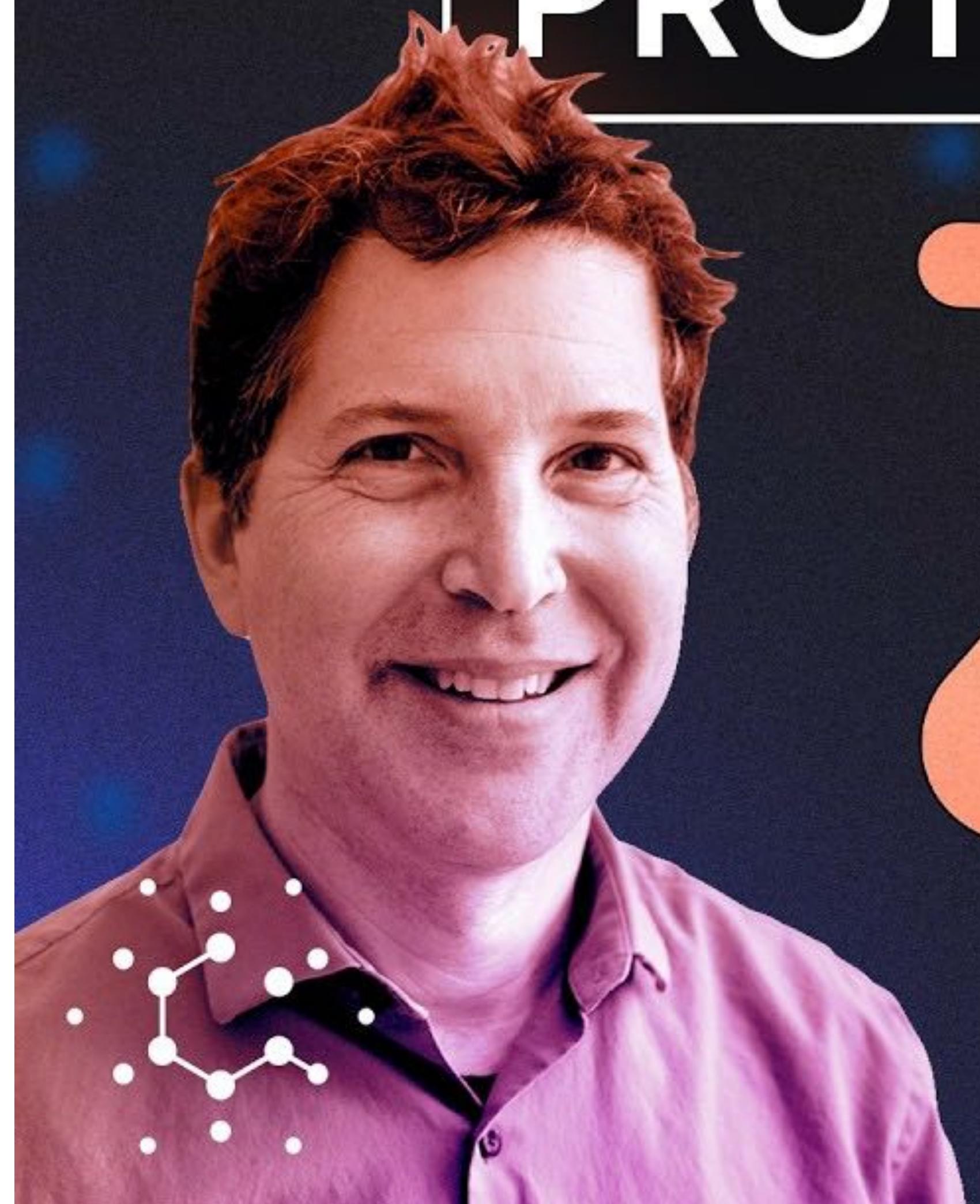


# AlphaZero



# PROTEIN PUZZLE

3D



# Unexpected Discoveries



Image of retina



Blood pressure predictions  
focus on blood vessels

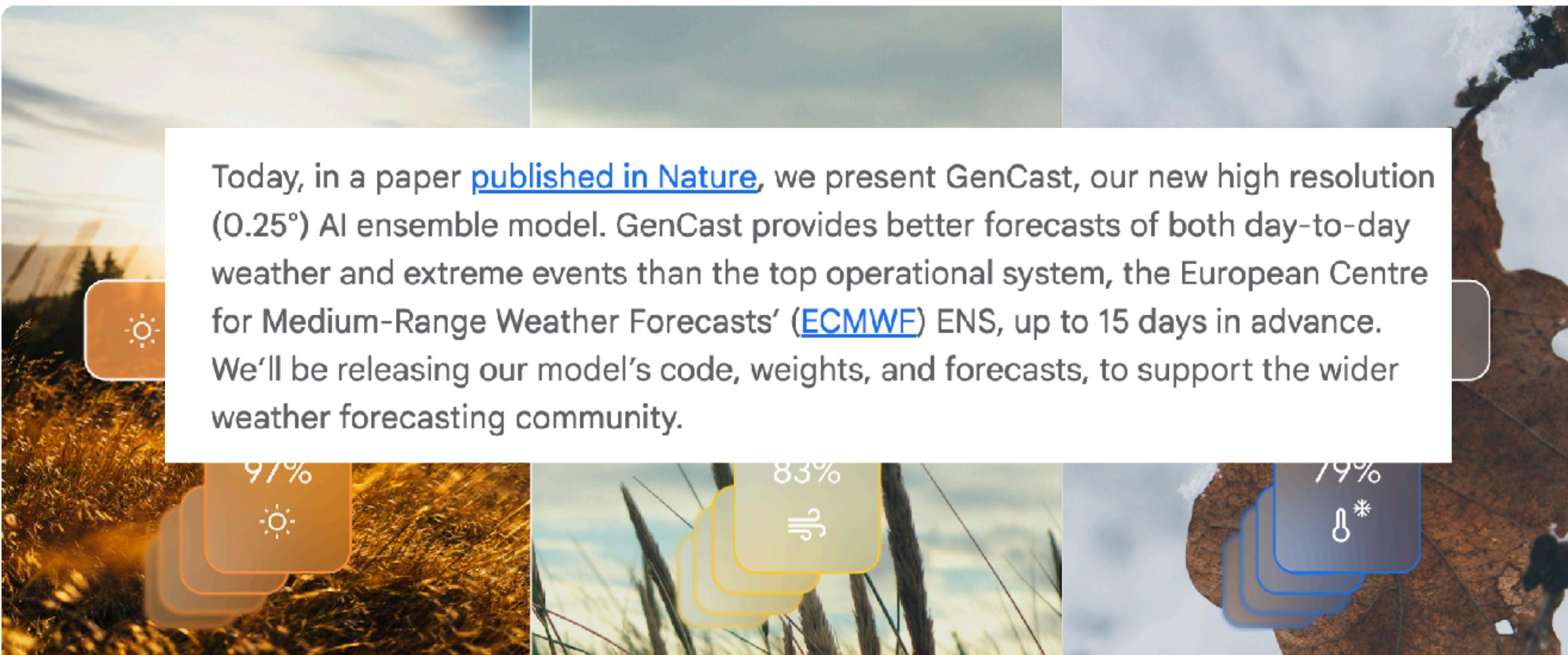
# Weather Prediction

GenCast predicts weather and the risks of extreme conditions with state-of-the-art accuracy

4 DECEMBER 2024

Ilan Price and Matthew Willson

 Share



Today, in a paper [published in Nature](#), we present GenCast, our new high resolution ( $0.25^\circ$ ) AI ensemble model. GenCast provides better forecasts of both day-to-day weather and extreme events than the top operational system, the European Centre for Medium-Range Weather Forecasts' ([ECMWF](#)) ENS, up to 15 days in advance. We'll be releasing our model's code, weights, and forecasts, to support the wider weather forecasting community.

# Your LLM Judge may be biased

by **Henry Papadatos, Rachel Freedman** 29th Mar 2024

## Abstract

AI safety researchers often rely on LLM “judges” to qualitatively evaluate the output of separate LLMs. We try this for our own interpretability research, but find that our LLM judges are often deeply biased. For example, we use Llama2 to judge whether movie reviews are more “(A) positive” or “(B) negative”, and find that it almost always answers “(B)”, even when we switch the labels or order of these alternatives. This bias is particularly surprising for two reasons: first, because we expect a fairly capable model like Llama2 to perform well at a simple sentiment classification task like this, and second, because this specific “(B)”-bias doesn’t map on to a human bias we’d expect to see in the training data. We describe our experiments, provide code to replicate our results, and offer suggestions to mitigate such biases. We caution researchers to double-check their LLM judges for such biases, and validate LLM judgements against human ones whenever possible.

