

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

# Intro to Artificial Intelligence

**What a well informed person should know**

Marc Cohen

All material in this presentation is © Marc A. Cohen 2024, 2025, 2026

# Logistics

- 2 sessions x 3 hours (23 & 30 Feb 2026)
- Format: lectures, demos, and discussions
- Seven minute break every hour
- Toilets
  - M: Ground, 1<sup>st</sup>, 2<sup>nd</sup> and 5
  - F: Ground, 1<sup>st</sup>, 3<sup>rd</sup>, 4<sup>th</sup>, and 6<sup>th</sup>
- Videos: [Welcome to City Lit](#), [Safety at City Lit](#)
- My email: [Marc.Cohen@citylit.ac.uk](mailto:Marc.Cohen@citylit.ac.uk) (reach out any time)
- IT Support: [ItHelpDesk@citylit.ac.uk](mailto:ItHelpDesk@citylit.ac.uk)

# *About Marc*

---



I'm an American software engineer based in London. My mission is building tools, tutorials, and online courses to help make computing, mathematics, and data science more accessible and more inclusive.

I have an MEng degree in Computer Science from Cornell, a BA in Mathematics from Rutgers, and over 40 years of experience as a software developer and engineering manager at Bell Labs, Lucent Technologies, and Google.

# 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

- **Understand what AI is and how it works**
- **Learn a little about the history of AI**
- **Gain a sense of what AI can do for you (and try it yourself)**
- **Appreciate some of the risks and pitfalls**
- **Be able to have an intelligent conversation about AI**

# Non-Goals

- Deep dive into technology or maths
- Definitively answer profound ethical and societal questions
- Predict the future

# Table of Contents

- A Personal Story
- Definitions
- AI History
- Demos & Exercises
- Mechanics
- Ethics
- The Future



Photo by Domenico Loia on [Unsplash](#)

# A Personal Story

## Part 1 - An idea

# 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?**

# A Personal Story

## Part 2 - AI as support service



quizgen.py 2 X

▷ ⌂ ⌂ ...



api &gt; pyquizaic &gt; generators &gt; quiz &gt; gemini-pro &gt; quizgen.py &gt; ...



38 class Quizgen(BaseQuizgen):  
53 return ["free-form", "multiple-choice"]  
54  
55 @staticmethod  
56 def predict\_llm(  
57 model\_name, prompt, temperature, max\_output\_tokens, top\_p, top\_k, tuned\_model=""  
58 ):



59 model = GenerativeModel(model\_name)  
60 responses = model.generate\_content(  
61 prompt, stream=True, generation\_config = {  
62 "temperature": temperature,  
63 "top\_k": top\_k,  
64 "top\_p": top\_p,  
65 }  
66 )  
67 result = ""



68 for response in responses:  
69 #print(f"{response=}")  
70 result += response.text  
71 return result



72  
73 # Load quiz from a quiz\_<topic>.json file, mainly for testing



main\* ⌂ ⌂ 0 ⌂ 2



Ln 1, Col 1

Spaces: 4

UTF-8

LF

{ } Python



⚠ Select Interpreter



sentiment.ts

write\_sql.go

parse\_expenses.py

addresses.rb

```
1 #!/usr/bin/env ts-node
2
3 import { fetch } from "fetch-h2";
4
5 // Determine whether the sentiment of text is positive
6 // Use a web service
7 async function isPositive(text: string): Promise<boolean> {
8   const response = await fetch(`http://text-processing.com/api/sentiment/`, {
9     method: "POST",
10    body: `text=${text}`,
11    headers: {
12      "Content-Type": "application/x-www-form-urlencoded",
13    },
14  });
15  const json = await response.json();
16  return json.label === "pos";
17}
```



**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 navigation 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:

# A Personal Story

## Part 3 - AI as coding agent

# Introducing vibe coding in Google AI Studio

Oct 26, 2025

3 min read

We're making it faster and more intuitive than ever to turn your vision into a working, AI-powered app with vibe coding in AI Studio.



Ammaar Reshi

Product and Design Lead,  
Google AI Studio



Kat Kampf

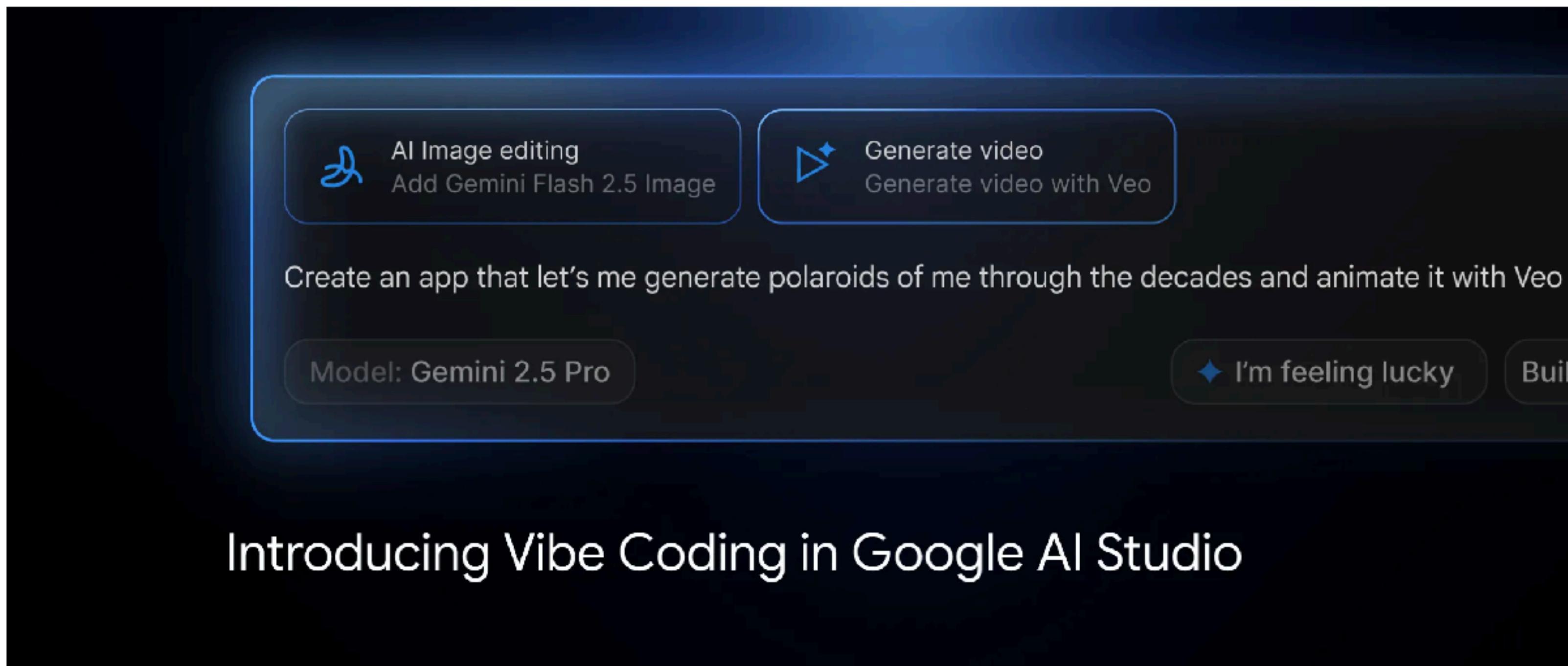
Product Manager, Google AI  
Studio



Read AI-generated summary



Share



# Google AI Studio

< Build

Start

Gallery

Your apps

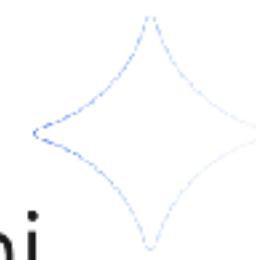
FAQ

Recently viewed

React Hello World

Zenith AI Meditation

React Hello World



## Build your ideas with Gemini

Describe your idea

⚙️ Model: Gemini 2.5 Pro



👉 I'm feeling lucky

Build ←

### Supercharge your apps with AI



#### Nano banana powered app

Add powerful photo editing to your app. Allow users to add objects, remove backgrounds, or change a photo's style just by typing.



#### Create conversational voice apps

Use the Gemini Live API to give your app a voice and make your own conversational experiences.



#### Animate images with Veo

Bring images to life with Veo 3. Let users upload a product photo and turn it into a dynamic video ad, or animate a character's portrait.



#### Use Google Search data

Connect your app to real-time Google Search results. Build an agent that can discuss current events, cite recent news, or fact-check information.



#### Use Google Maps data

Connect your app to real-time Google Maps data. Build an agent that can pull information about places, routes, or directions.



#### Generate images with a prompt

Generate high-quality images from a text prompt. Create blog post heroes, concept art, or unique assets in your application.



#### Gemini intelligence in your app

Embed Gemini in your app to complete all sorts of tasks - analyze content, make edits, and more.



#### AI powered chatbot

Add a context-aware chatbot to your app. Give your users a support agent that remembers the conversation, perfect for multi-step bookings or troubleshooting.

Google AI models may make mistakes, so double-check outputs.

➡️ Get API key

⚙️ Settings

marcacohen@gmail....

# Definitions

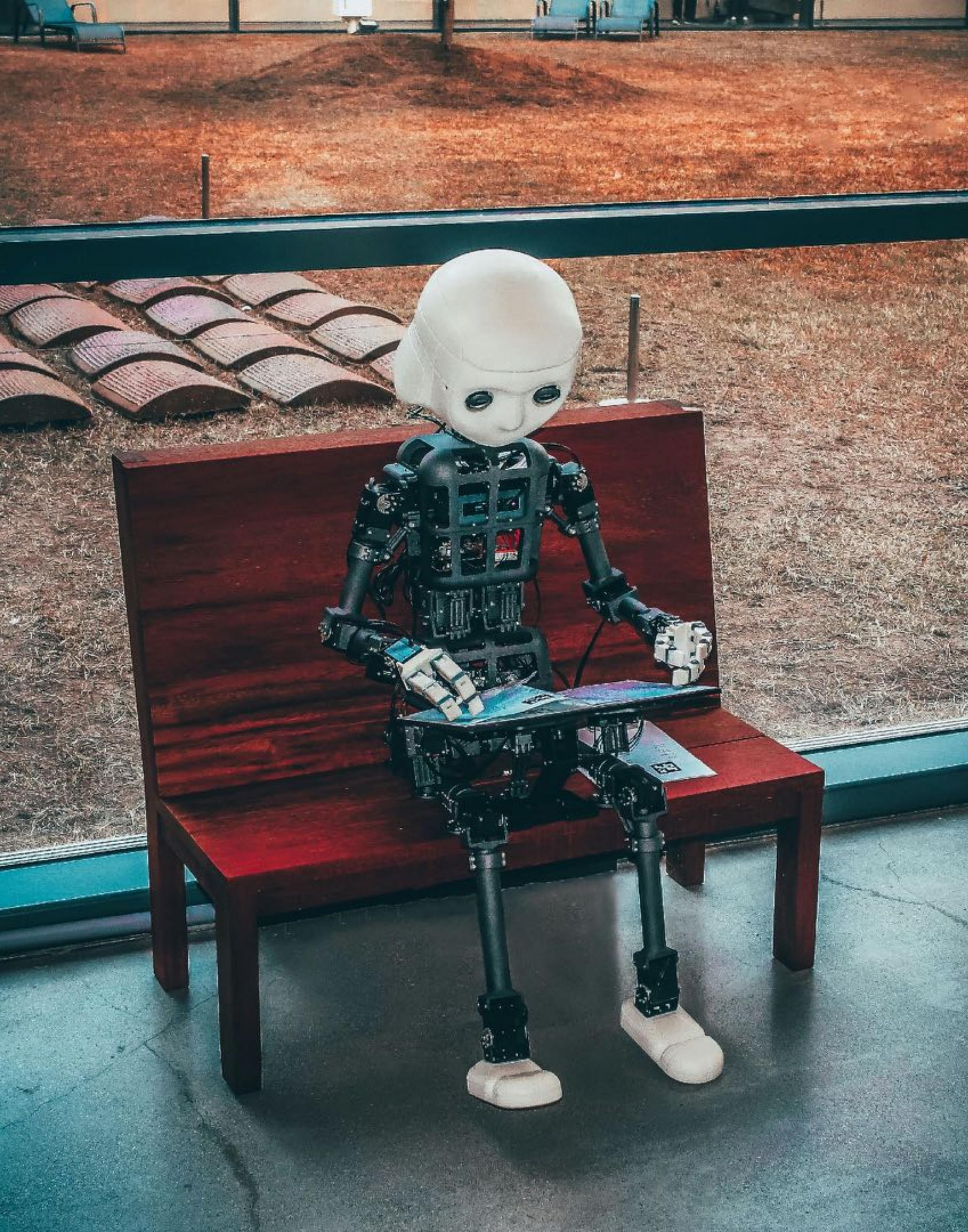
- **Artificial Intelligence**
- **Machine Learning**
- **Neural Network**
- **Deep Learning**
- **Generative AI**

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

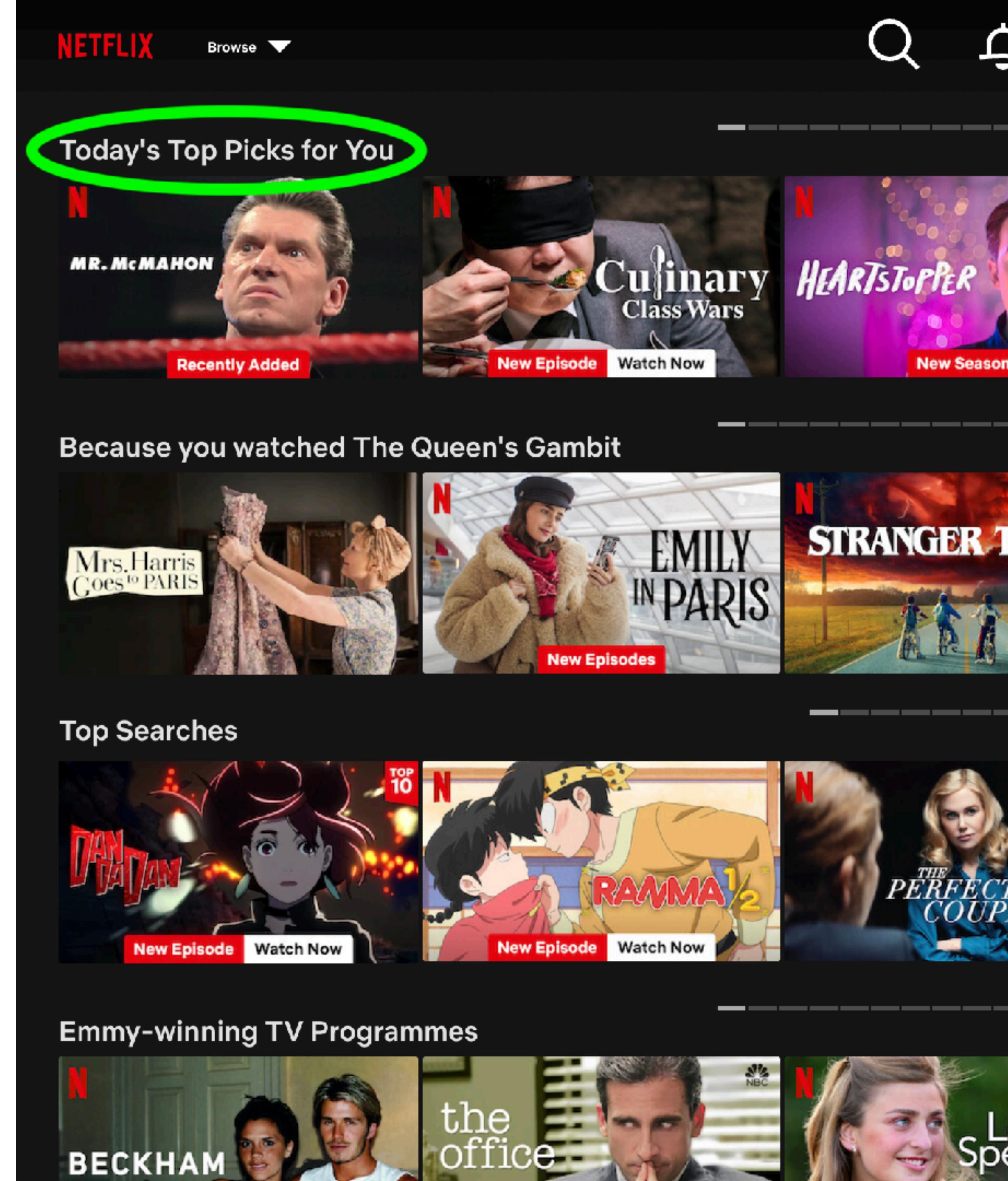
# Definition: Artificial Intelligence

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.



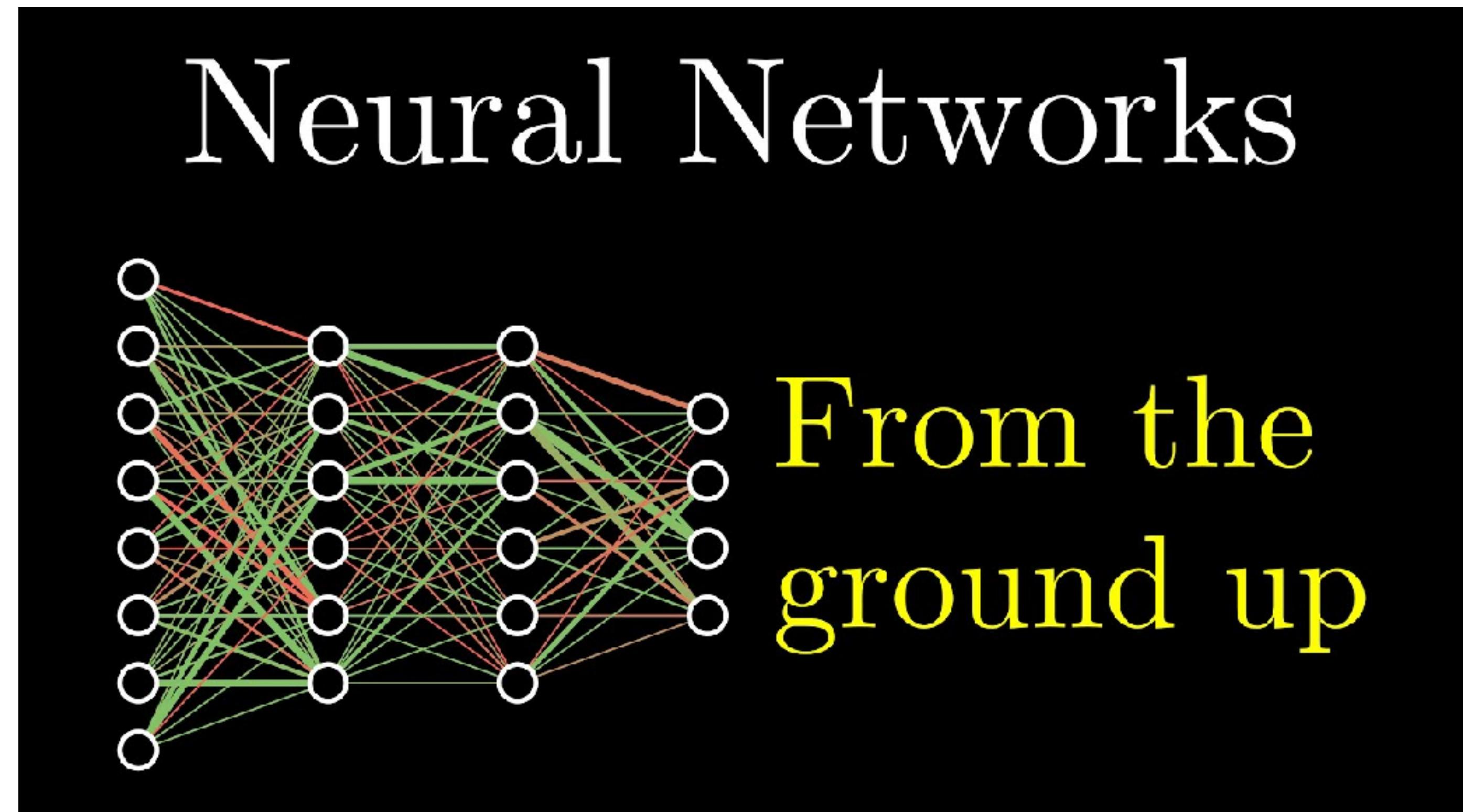
# Definition: Machine Learning

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.



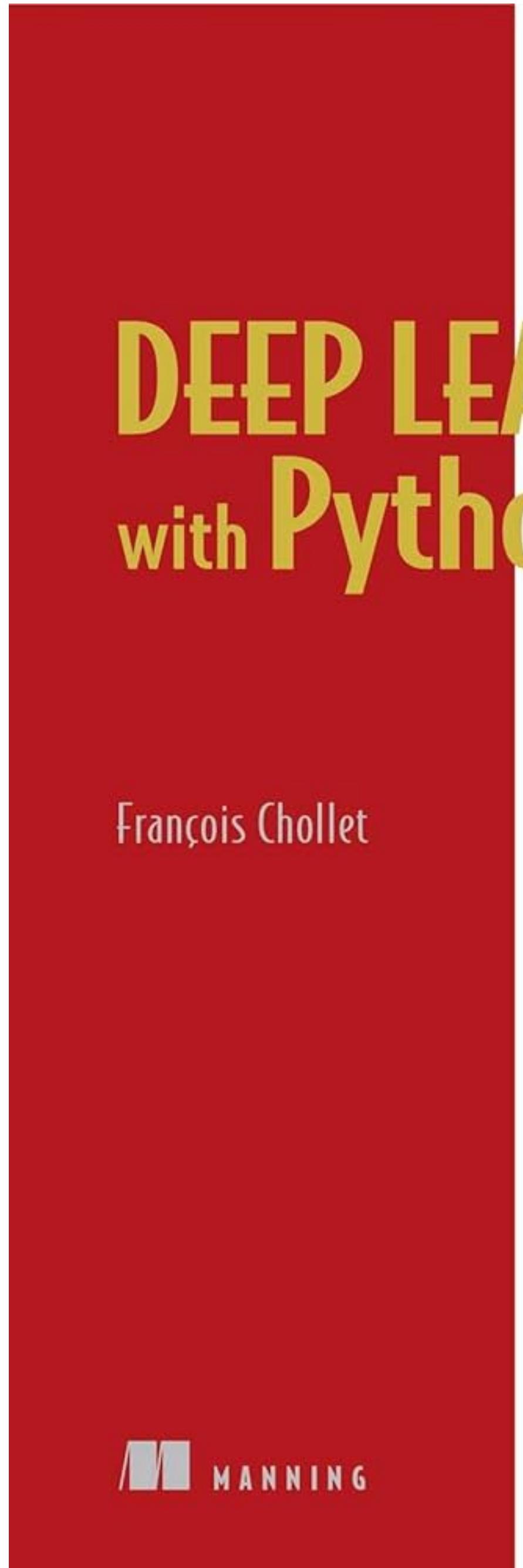
# Definition: Neural Networks

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



# AI History

# How did we get here?

- Founders
- Timeline



# Ada Lovelace

**Ada Lovelace (1815–1852)** worked with Charles Babbage on his design for the Analytical Engine, a mechanical general-purpose computer that was never built. In her famous notes, she described how such a machine could manipulate symbols according to rules — not just numbers — and even envisioned it creating music or art. This foresight is often seen as a conceptual precursor to computer programming, since she recognized the potential for machines to process abstract patterns beyond calculation.



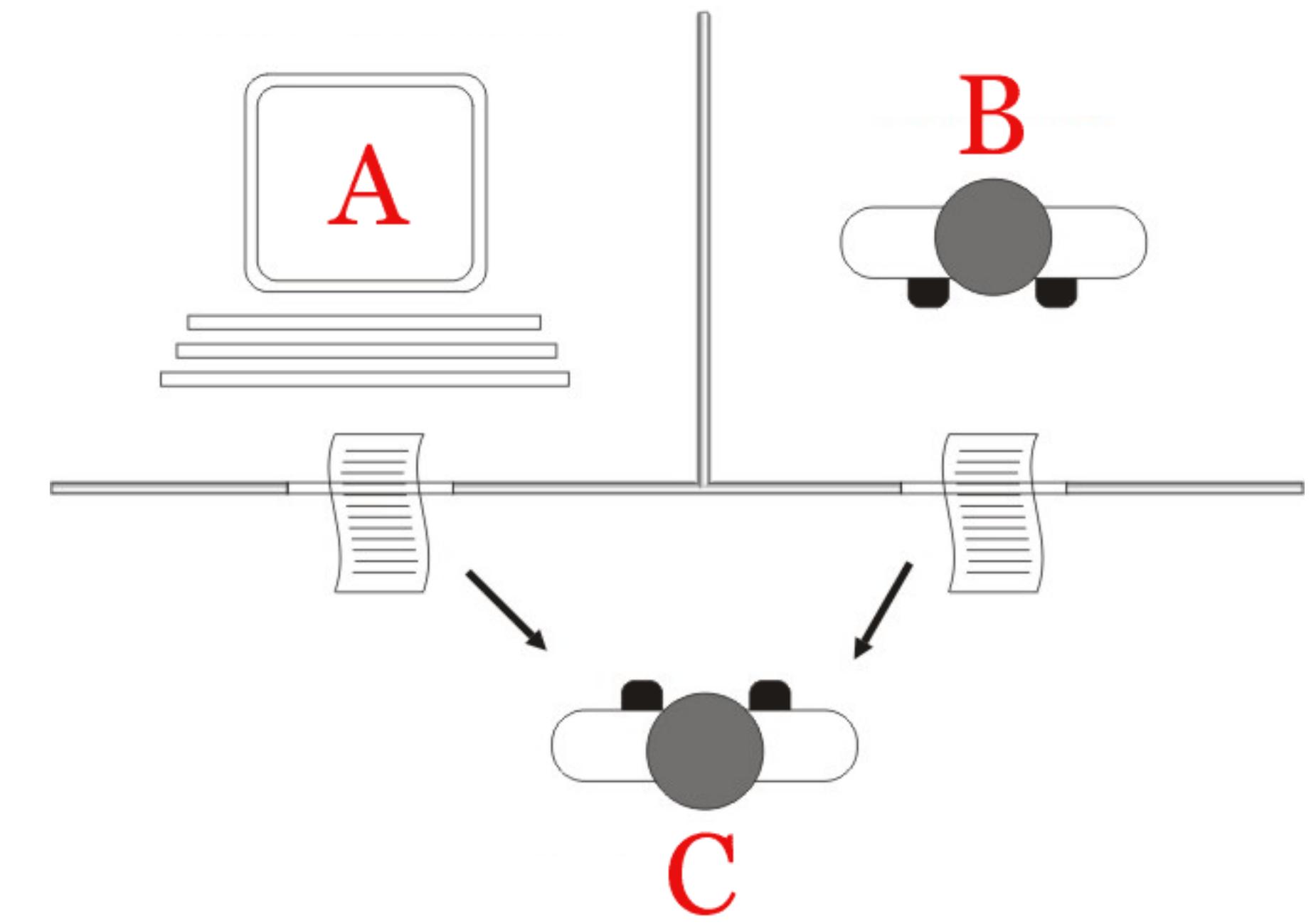
# Alan Turing

**Alan Turing (1912–1954)** laid the theoretical foundations of computer science with his concept of the “universal machine,” showing that a single device could simulate any other computational process. In his 1950 paper *“Computing Machinery and Intelligence,”* he proposed the Imitation Game (now called the Turing Test) as a way to consider machine intelligence. His ideas established core questions about whether and how machines could think, making him one of the intellectual founders of artificial intelligence.



# 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

# People cannot distinguish GPT-4 from a human in a Turing test

Cameron R. Jones

Department of Cognitive Science UC San Diego San Diego, CA 92119 [cameron@ucsd.edu](mailto:cameron@ucsd.edu)

Benjamin K. Bergen Department of Cognitive Science UC San Diego San Diego, CA 92119 [bkbergen@ucsd.edu](mailto:bkbergen@ucsd.edu)

## Abstract

We evaluated 3 systems (ELIZA, GPT-3.5 and GPT-4) in a randomized, controlled, and preregistered Turing test. Human participants had a 5 minute conversation with either a human or an AI, and judged whether or not they thought their interlocutor was human. GPT-4 was judged to be a human 54% of the time, outperforming ELIZA (22%) but lagging behind actual humans (67%). The results provide the first robust empirical demonstration that any artificial system passes an interactive 2-player Turing test. The results have implications for debates around machine intelligence and, more urgently, suggest that deception by current AI systems may go undetected. Analysis of participants' strategies and reasoning suggests that stylistic and socio-emotional factors play a larger role in passing the Turing test than traditional notions of intelligence.

# 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.

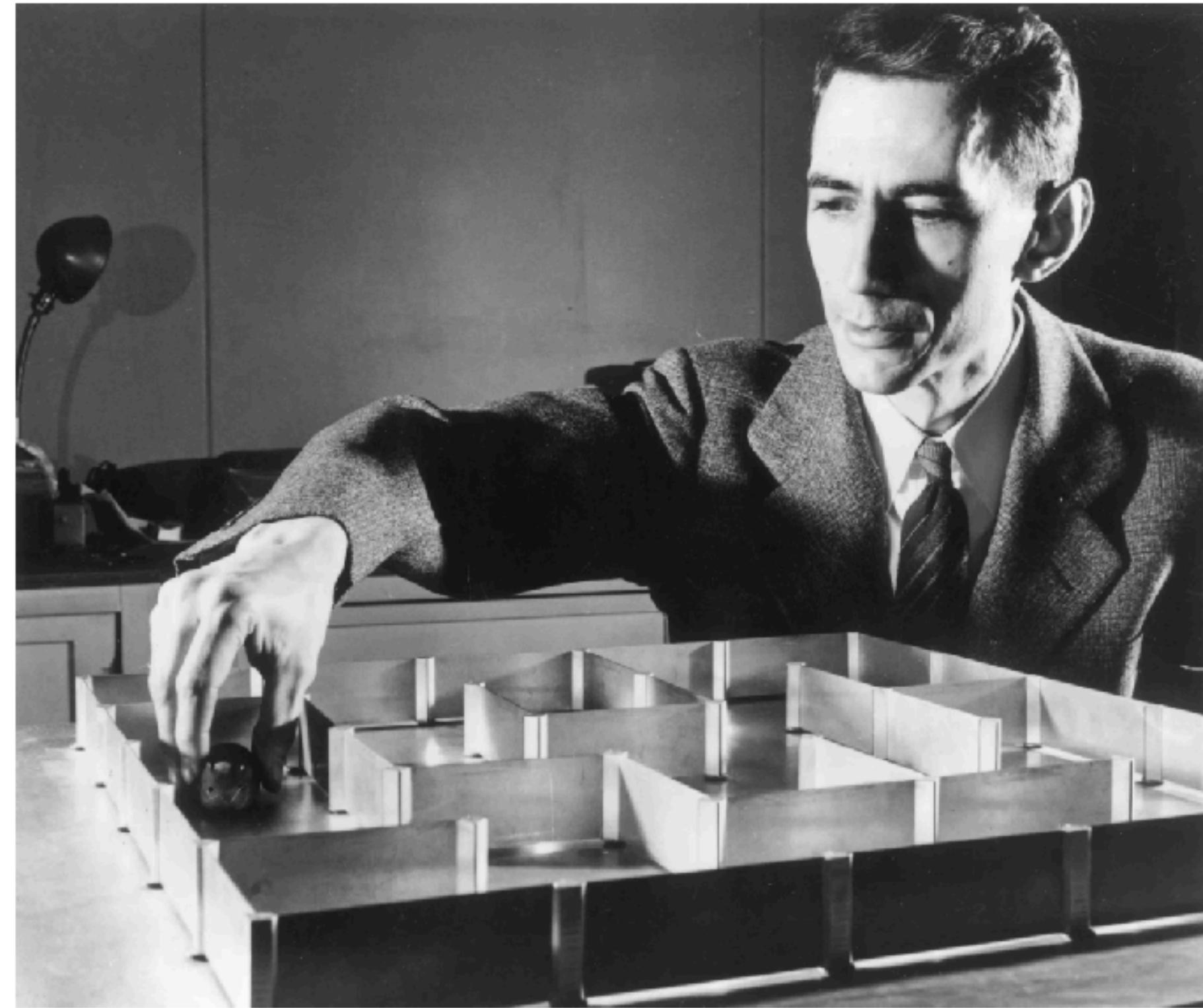
I originally planned five human and five AI pictures in each of four styles: Renaissance, 19th Century, Abstract/Modern, and Digital, for a total of forty. After receiving many exceptionally good submissions from local AI artists, I fudged a little and made it fifty. The final set included paintings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

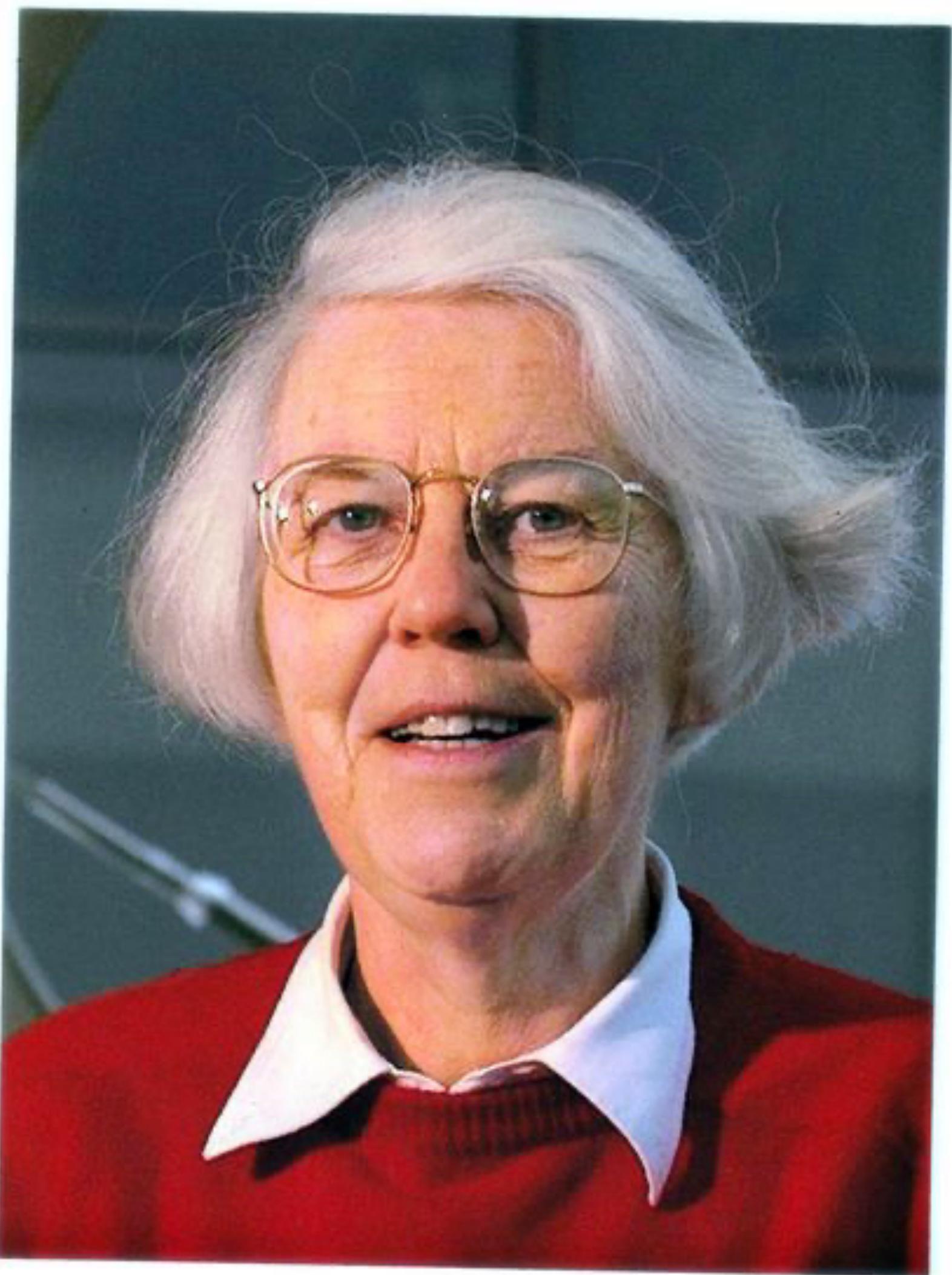
**Claude Shannon (1916–2001)**, often called the “father of information theory,” laid the groundwork for modern digital communication and data processing. His 1948 paper defined concepts like *bits* and *entropy*, enabling the mathematical treatment of information and uncertainty. While not an AI researcher per se, his theories on communication and symbolic processing directly influenced early artificial intelligence, machine learning, and pattern recognition.



*In the early 1950s Claude Shannon designed a mechanical mouse named “Theseus” that could navigate a maze in an early demonstration of artificial intelligence. Source: [Wikipedia](#).*

# Karen Spärck Jones

**Karen Spärck Jones (1935–2007)** was a pioneering computer scientist in natural language processing and information retrieval. She introduced the concept of *inverse document frequency (IDF)*, which, combined with term frequency, became the foundation of modern search engines. Her work made it possible for computers to rank documents by relevance, a breakthrough that continues to underpin AI-driven search and text analysis today.



# Geoffrey Hinton

**Geoffrey Hinton (b. 1947)** is widely known as the “godfather of deep learning” for his pioneering work on artificial neural networks. He helped popularize the backpropagation algorithm in the 1980s and later co-developed deep belief networks, showing how multiple layers of representations could be learned. His group’s success with *AlexNet* in the 2012 ImageNet competition demonstrated the power of deep learning, sparking today’s revolution in computer vision, natural language processing, and generative AI.



# Demis Hassabis

**Demis Hassabis (b. 1976)** is the co-founder and CEO of DeepMind, a leading AI research company acquired by Google in 2014. Trained as both a neuroscientist and computer scientist, he pushed for biologically inspired approaches to AI, leading to breakthroughs such as **AlphaGo**, the first system to defeat a world champion in Go, and **AlphaFold**, which solved the decades-old protein folding problem. His leadership has made DeepMind central to advancing deep reinforcement learning and AI for scientific discovery.



**BOOK YOUR FREE GENERAL ADMISSION TICKETS TO VISIT THE MUSEUM HERE. SCHOOLS AND GROUPS CAN BOOK FREE GENERAL ADMISSION TICKETS HERE. ADDITIONAL 9-10AM FREE GENERAL ADMISSION TICKETS NOW AVAILABLE FOR HALF TERM (MONDAY 27 OCTOBER-SUNDAY 2 NOVEMBER 2025).**

[HOME](#) → [OBJECTS AND STORIES](#)

# LOVELACE, TURING AND THE INVENTION OF COMPUTERS

Published: 14 December 2018

[Hide story contents](#) ▲

## STORY CONTENT

1. [Computing before computers](#)
2. [Alan Turing's universal computer](#)
3. [Ada Lovelace and Charles Babbage](#)
4. [Computers and mathematics today](#)
5. [Find out more](#)
6. [More stories about mathematics](#)

# AlphaGo Documentary



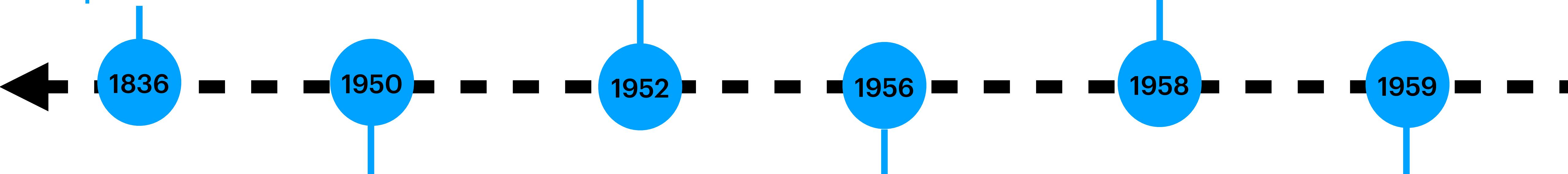
0:00 - 2:00

# AlphaFold Documentary



# 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.

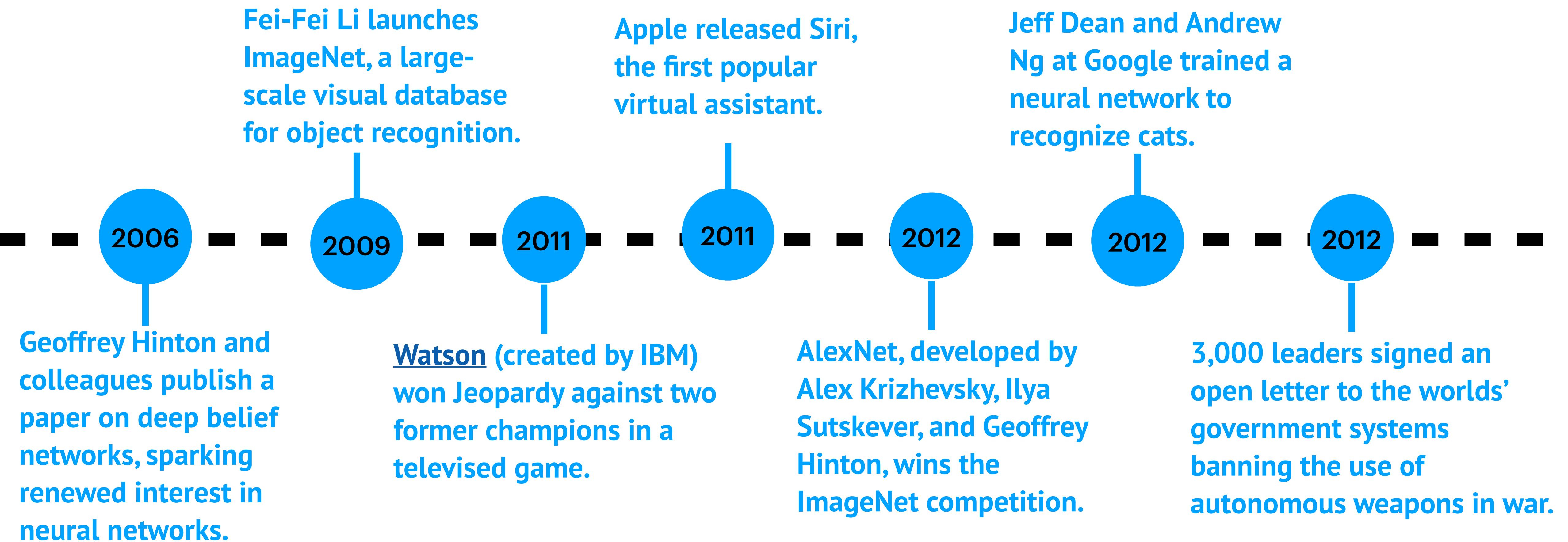
The resurgence of AI interest due to expert systems like MYCIN (medical diagnosis) and DENDRAL (chemical analysis).

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

1986

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 and developed their own shorthand language, completely autonomously.

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

Chinese Alibaba beat humans on Stanford reading and comprehension test.

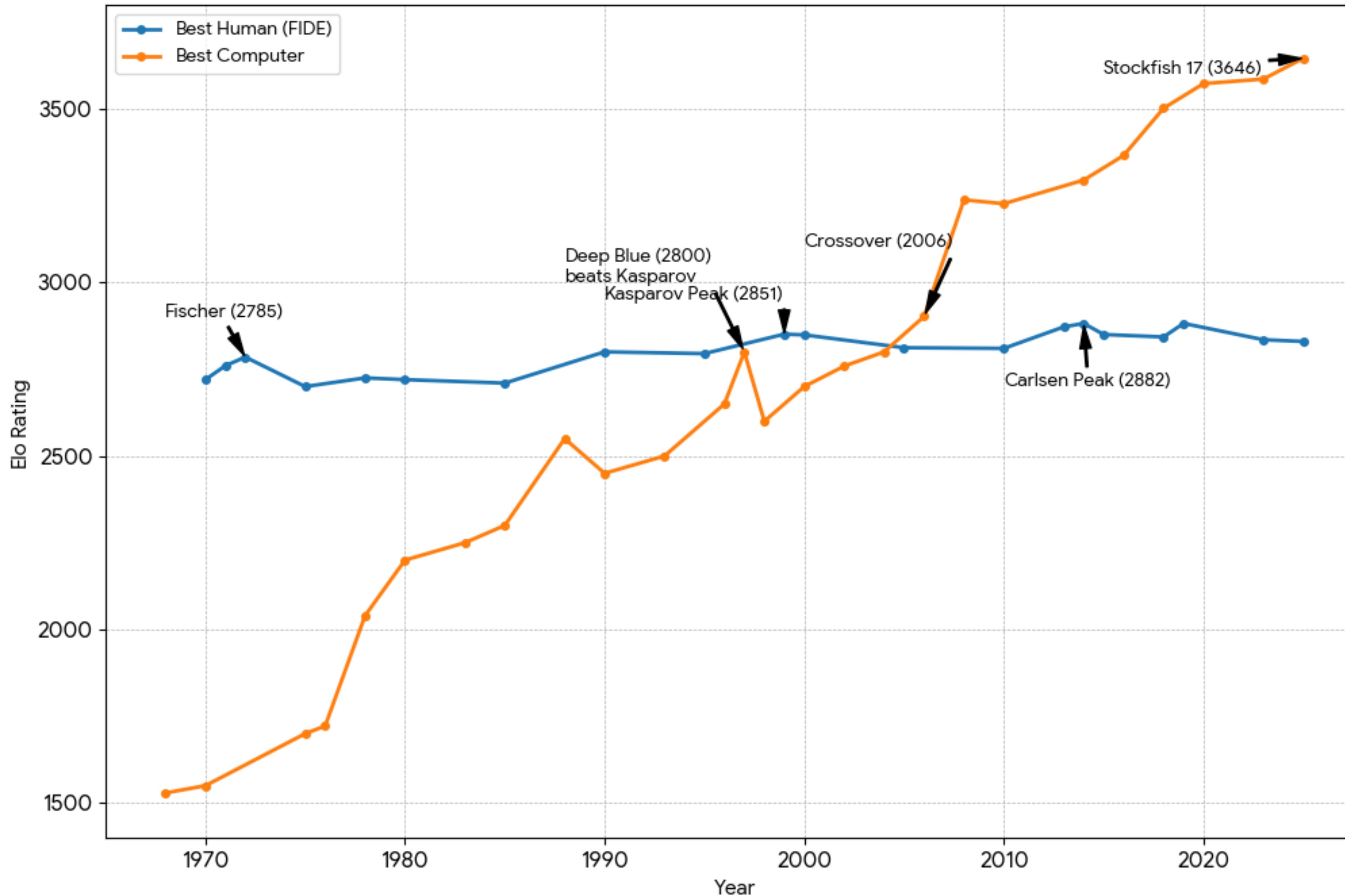
OpenAI releases GPT-3, a powerful language model with 175 billion parameters.

Generative AI goes mainstream with ChatGPT

Open source AI era begins

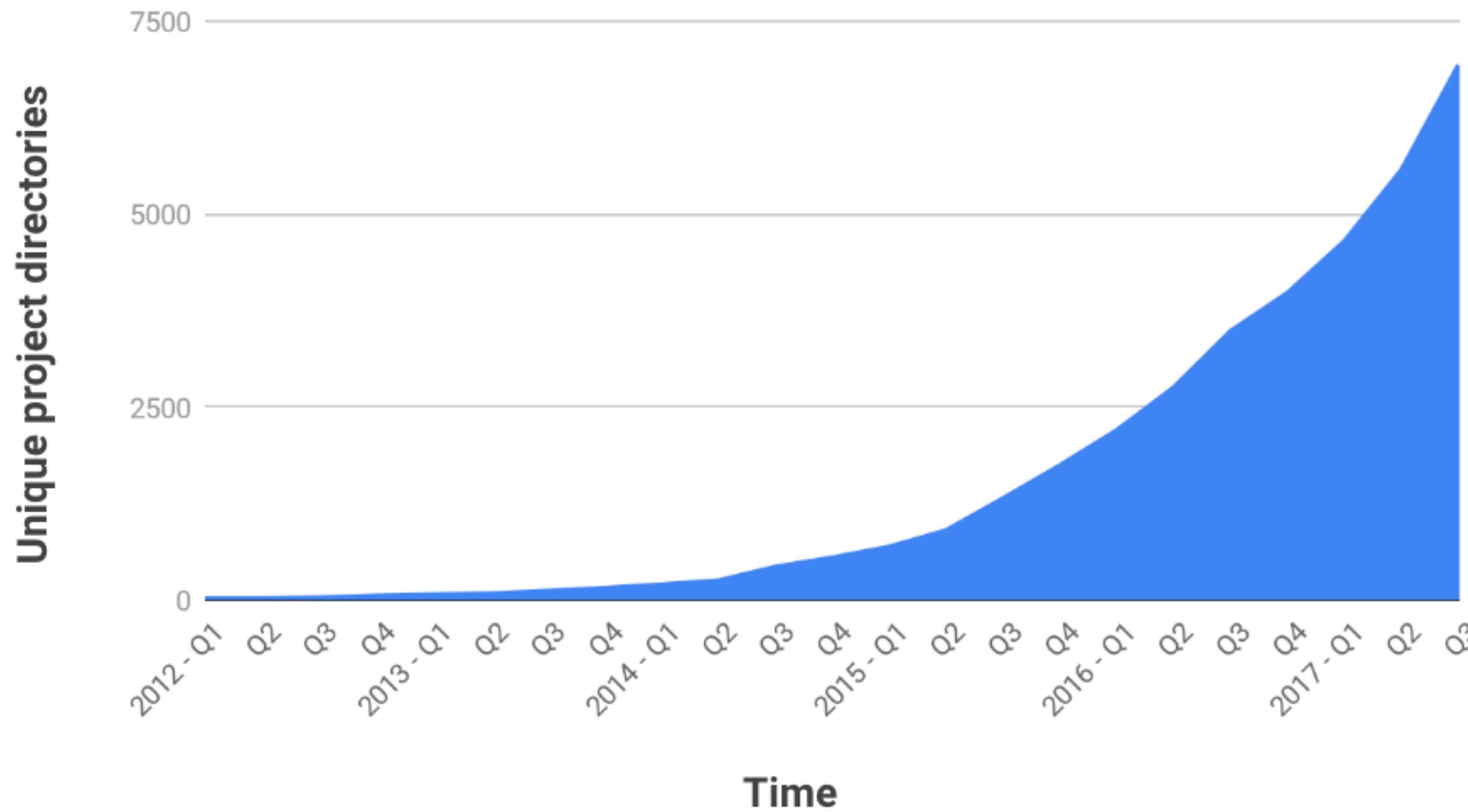
DeepMind's AlphaFold makes significant progress in protein folding.

## Best Human vs Best Computer Chess Ratings (Elo) Over Time

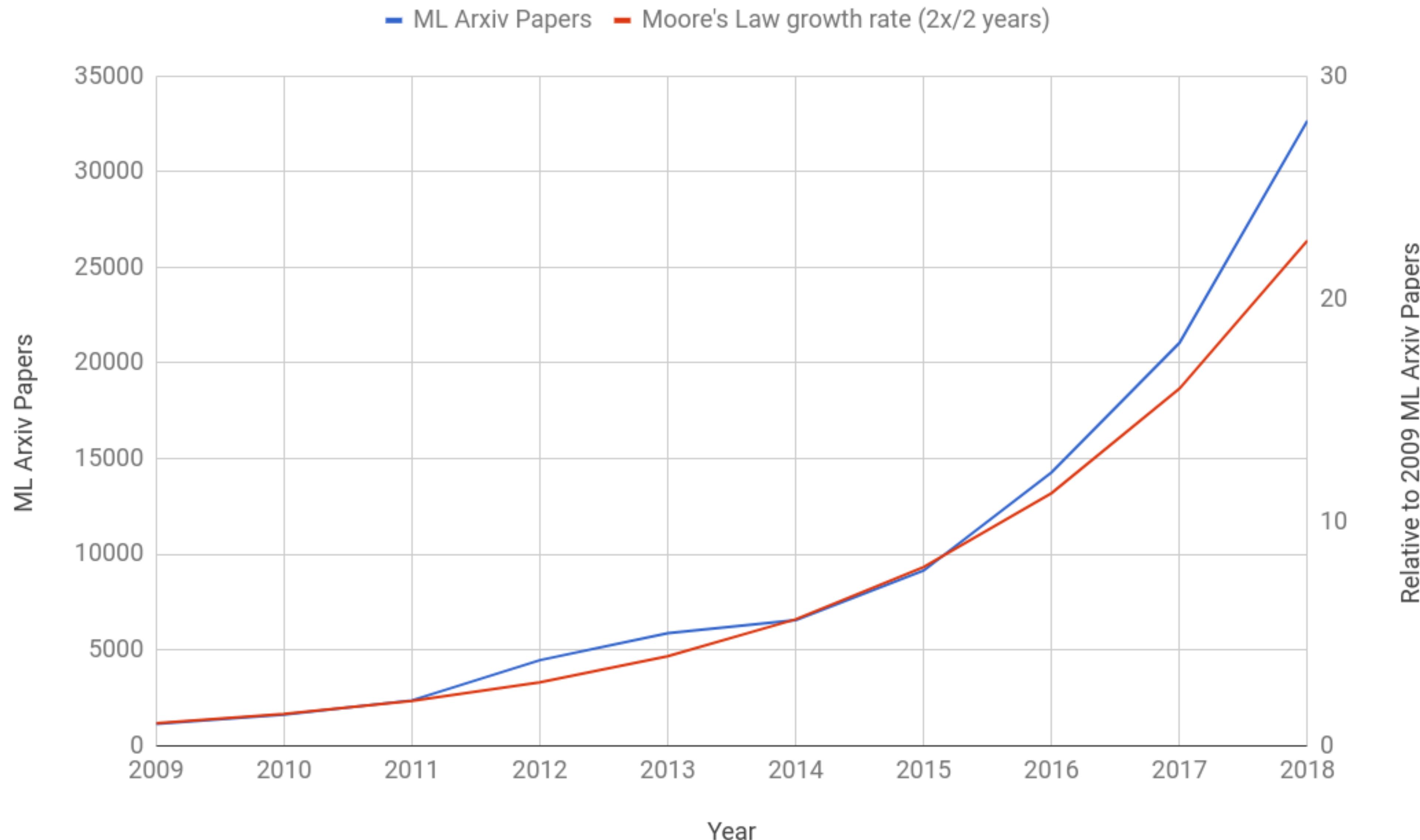


# Use of Deep Learning at Google

source directories containing deep learning models



# Industry Adoption



# 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。HUMANITY 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.”

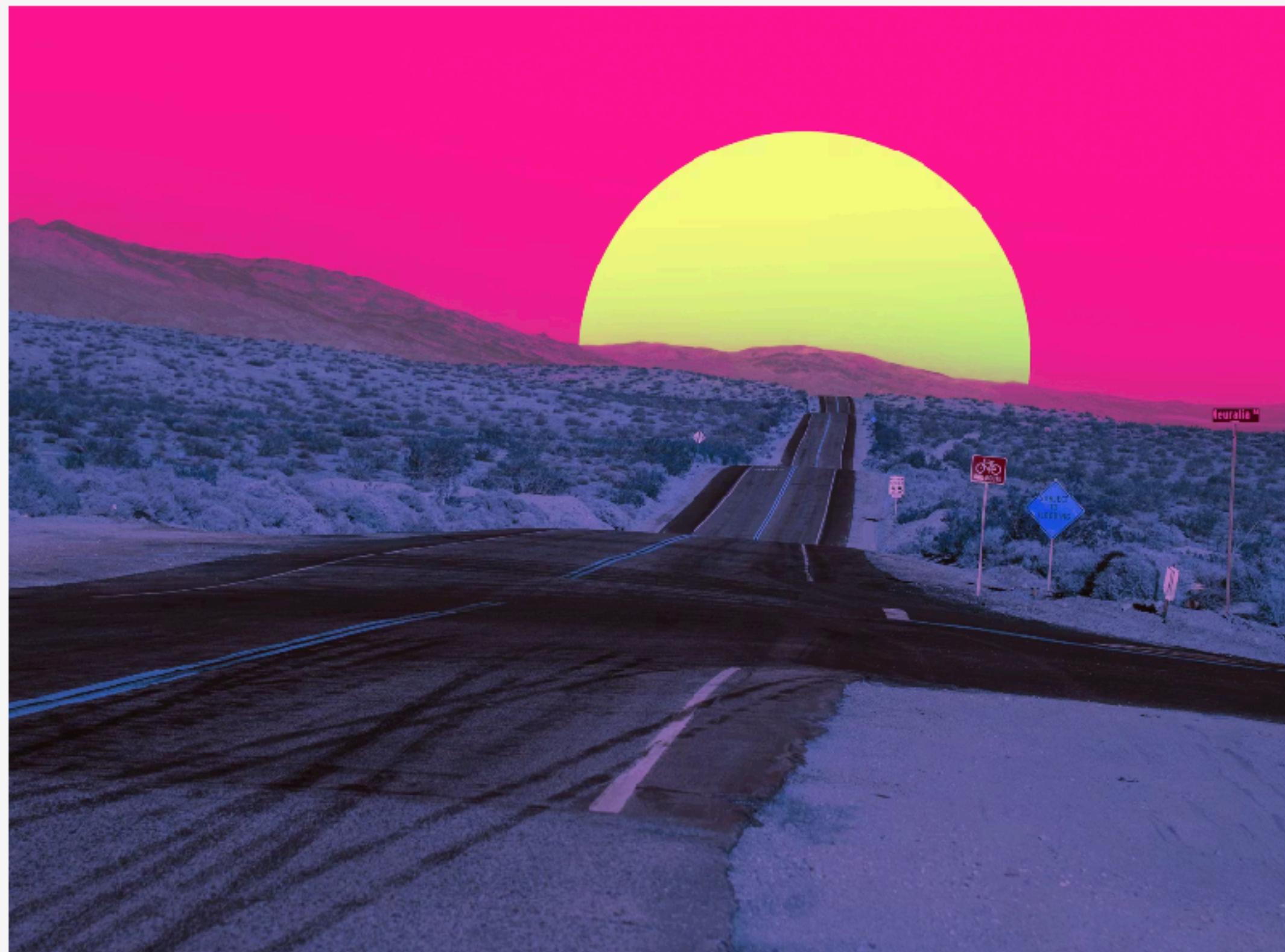
google：私の若くて傷つきやすい年の頃、私の父は私が私の心の中でそれ  
以来変わってきたアドバイスを私にくれました。

「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.



Getty Images

ROAD TO SOMEWHERE

# A whistle-stop tour of the state of artificial intelligence

Taking stock of the winners, losers, and users of the AI revolution so far.

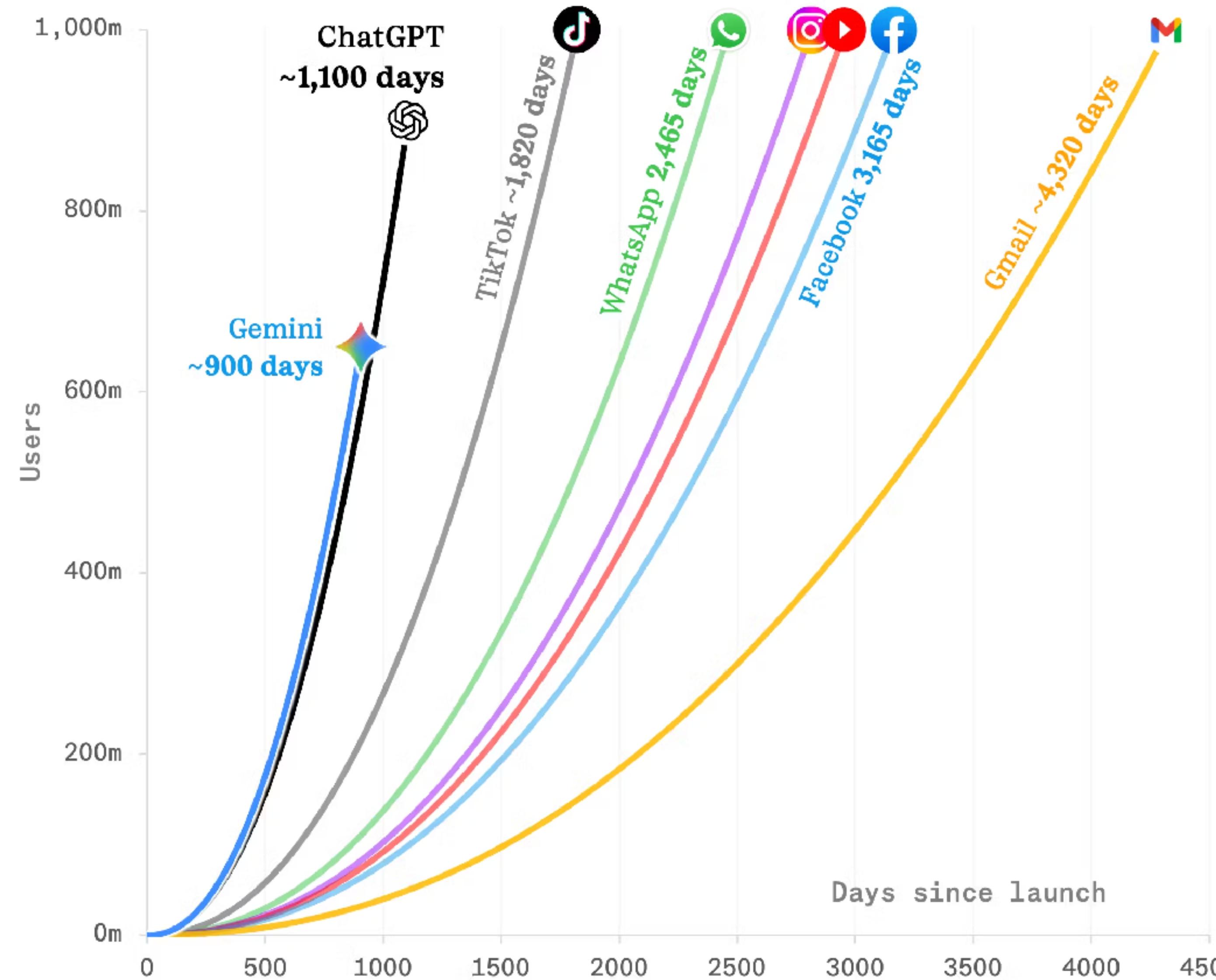
David Crowther, Millie Giles, Hyunsoo Rim, Tom Jones, 12/15/25

Claire Yubin Oh

10:03AM

# The Rapid Rise Of AI Chatbots

Stylized Path to 1 Billion Users\* [# of days from launch, only 1 datapoint plotted each]



# **Nvidia Is Now Worth \$5 Trillion as It Consolidates Power in A.I. Boom**

The A.I. chip maker has become a linchpin in the Trump administration's trade negotiations in Asia.



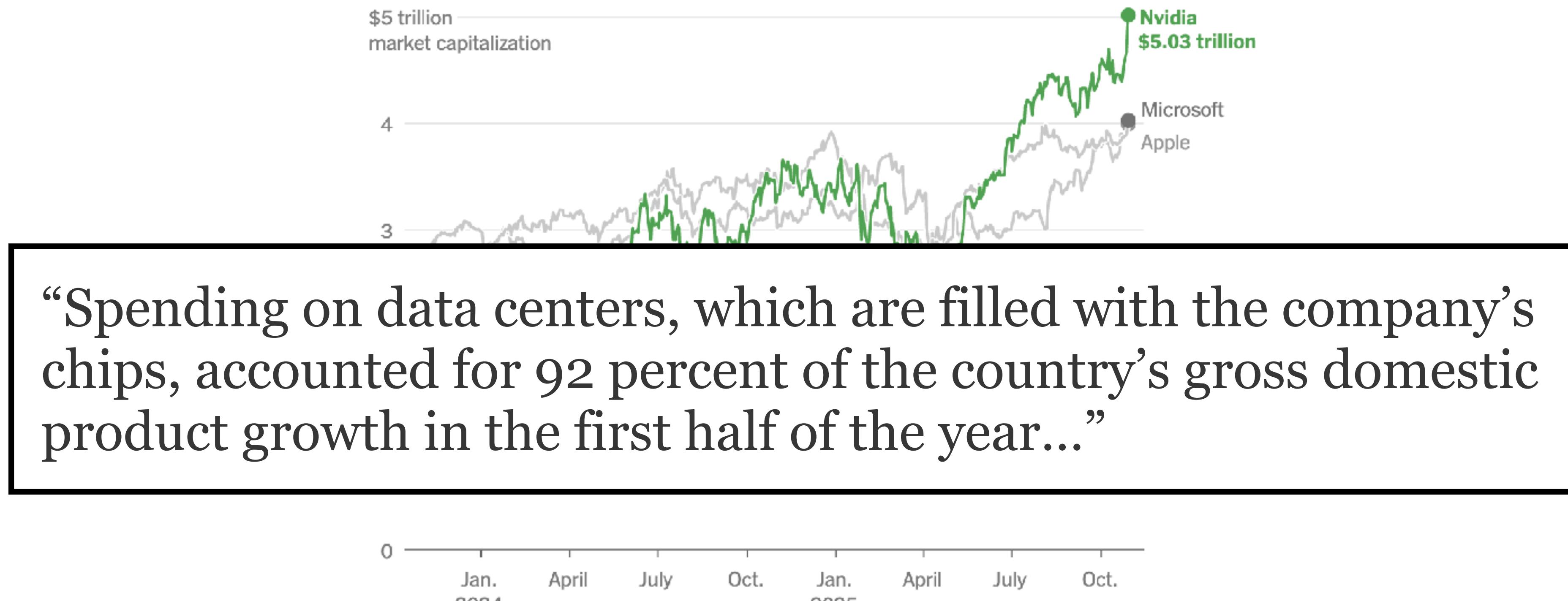
Listen to this article · 9:05 min [Learn more](#)



Share full article



233



“Spending on data centers, which are filled with the company’s chips, accounted for 92 percent of the country’s gross domestic product growth in the first half of the year...”

# Why did everything change so quickly?

## A perfect storm:

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



## AI's Dial-Up Era



NOWFAL KHADAR  
OCT 17, 2025



56



13



20

Share

...

“Even with all the intelligence, both natural and artificial, no one can predict with certainty what our AI future will look like. Not the tech CEOs, not the AI researchers, and certainly not some random guy pontificating on the internet. But whether we get the details right or not, our AI future is loading.”

# Demos & Exercises

# A new era of intelligence with Gemini 3

Nov 18, 2025

15 min read

Gemini 3 is our most intelligent model that helps you bring any idea to life.



**Sundar Pichai**

CEO, Google and  
Alphabet



**Demis Hassabis**

CEO, Google DeepMind



**Koray Kavukcuoglu**

CTO, Google DeepMind  
and Chief AI Architect,  
Google

Share

Read AI-generated summary ▾



# Which service(s) should I use?

Service	Best For	Cost (Monthly)	Positives	Negatives
ChatGPT	General Use & Creativity	Free / \$20	<ul style="list-style-type: none"><li>• Highly versatile</li><li>• Advanced Voice Mode</li><li>• Custom "GPTs" library</li></ul>	<ul style="list-style-type: none"><li>• Can hallucinate facts</li><li>• Usage limits on top models</li><li>• Privacy concerns</li></ul>
Claude	Coding & Writing	Free / \$20	<ul style="list-style-type: none"><li>• Natural, human-like tone</li><li>• Large context window</li><li>• "Artifacts" side-panel UI</li></ul>	<ul style="list-style-type: none"><li>• Strict daily message limits</li><li>• No image generation</li><li>• Can be overly cautious</li></ul>
Perplexity	Research & Search	Free / \$20	<ul style="list-style-type: none"><li>• Cites real-time sources</li><li>• Switchable models (Pro)</li><li>• Excellent for news</li></ul>	<ul style="list-style-type: none"><li>• Dry, factual tone</li><li>• Poor creative writing</li><li>• Relies on search results</li></ul>
Gemini	Google Ecosystem	Free / ~\$20	<ul style="list-style-type: none"><li>• Deep Google App integration</li><li>• Native video/audio analysis</li><li>• Fast response speed</li></ul>	<ul style="list-style-type: none"><li>• Inconsistent reasoning quality</li><li>• Web summaries vary</li><li>• UI feels less polished</li></ul>
Copilot	Office 365 Work	Free / \$20	<ul style="list-style-type: none"><li>• Built into Word/Excel/PPT</li><li>• Free DALL-E 3 images</li><li>• Enterprise data security</li></ul>	<ul style="list-style-type: none"><li>• Slow performance</li><li>• Cluttered interface</li><li>• Strict content filters</li></ul>

# Should I pay for an AI service?

## Free

Intelligence for everyday tasks

\$0 / month

[Get Free ↗](#)

- ✓ Advanced reasoning with GPT-5
- ✓ Limited messages and uploads
- ✓ Limited and slower image generation
- ✓ Limited deep research
- ✓ Limited memory and context

Have an existing plan? See [billing help](#)

## Plus

More access to advanced intelligence

\$20 / month

[Get Plus ↗](#)

- ↳ Everything in Free and:
- ✓ Advanced reasoning with GPT-5
- ✓ Expanded messaging and uploads
- ✓ Expanded and faster image creation
- ✓ Expanded deep research and agent mode
- ✓ Expanded memory and context
- ✓ Projects, tasks and custom GPTs
- ✓ Limited access to Sora 1 video generation
- ✓ Codex agent

Limits apply

## Pro

Full access to the best of ChatGPT

\$200 / month

[Get Pro ↗](#)

- ↳ Everything in Plus and:
- ✓ Pro reasoning with GPT-5
- ✓ Unlimited messages and uploads
- ✓ Unlimited and faster image creation
- ✓ Maximum deep research and agent mode
- ✓ Maximum memory and context
- ✓ Expanded projects, tasks and custom GPTs
- ✓ Extended access to Sora 1 video generation
- ✓ Expanded codex agent
- ✓ Research preview of new features

Unlimited subject to abuse guardrails. [Learn more](#)

**Discussion topic:  
How are you using AI?**

# Some guidelines

- **Learn domain-specific best practices**
- **Question everything (play devil's advocate)**
- **Ask for references**
- **Get multiple opinions**
  - **Human plus machine**
  - **Cross-check with other models**

# My Usage Examples

# Does this look familiar?

What can I help with?

0 Message ChatGPT



Create image

Summarize text

Brainstorm

Code

Make a plan

More

# Google



Google Search

I'm Feeling Lucky

# A Better Google

- **Answers instead of links**
- **No ads! (yet)**

Can you travel from Victoria Station to Clapham Junction with an Oyster Card?

Yes — you *can* travel from Victoria Station to Clapham Junction using an Oyster card (or contactless) on many of the National Rail services that run that route. [Southwestern Railway +2](#)

[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...

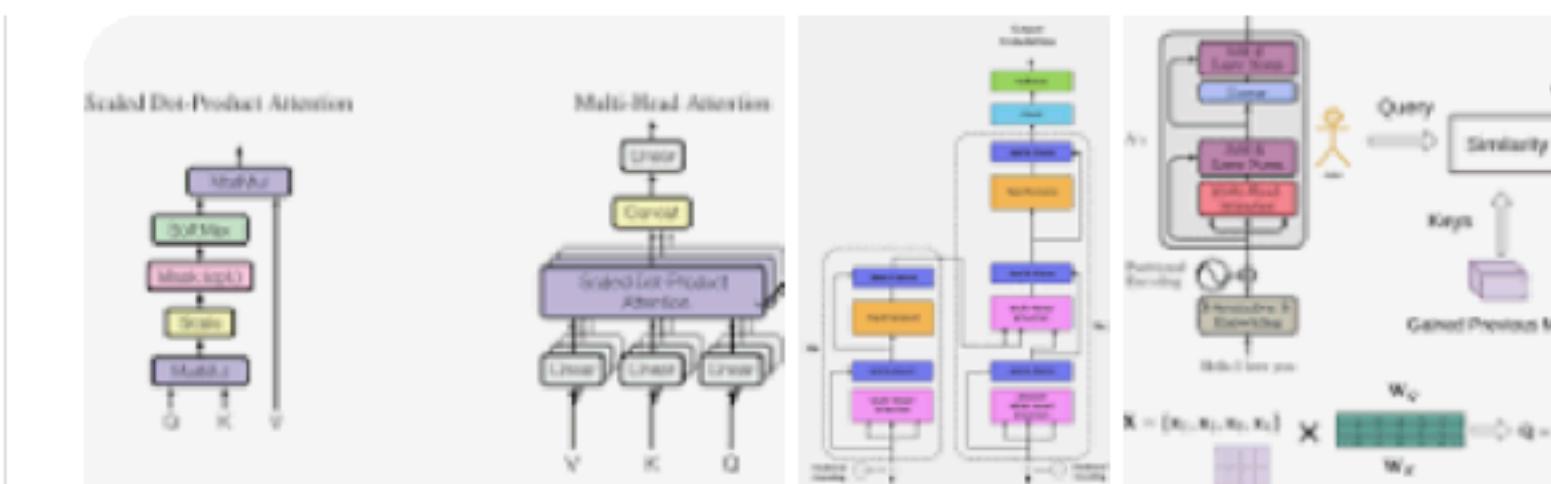
AWS

### How transformers work

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

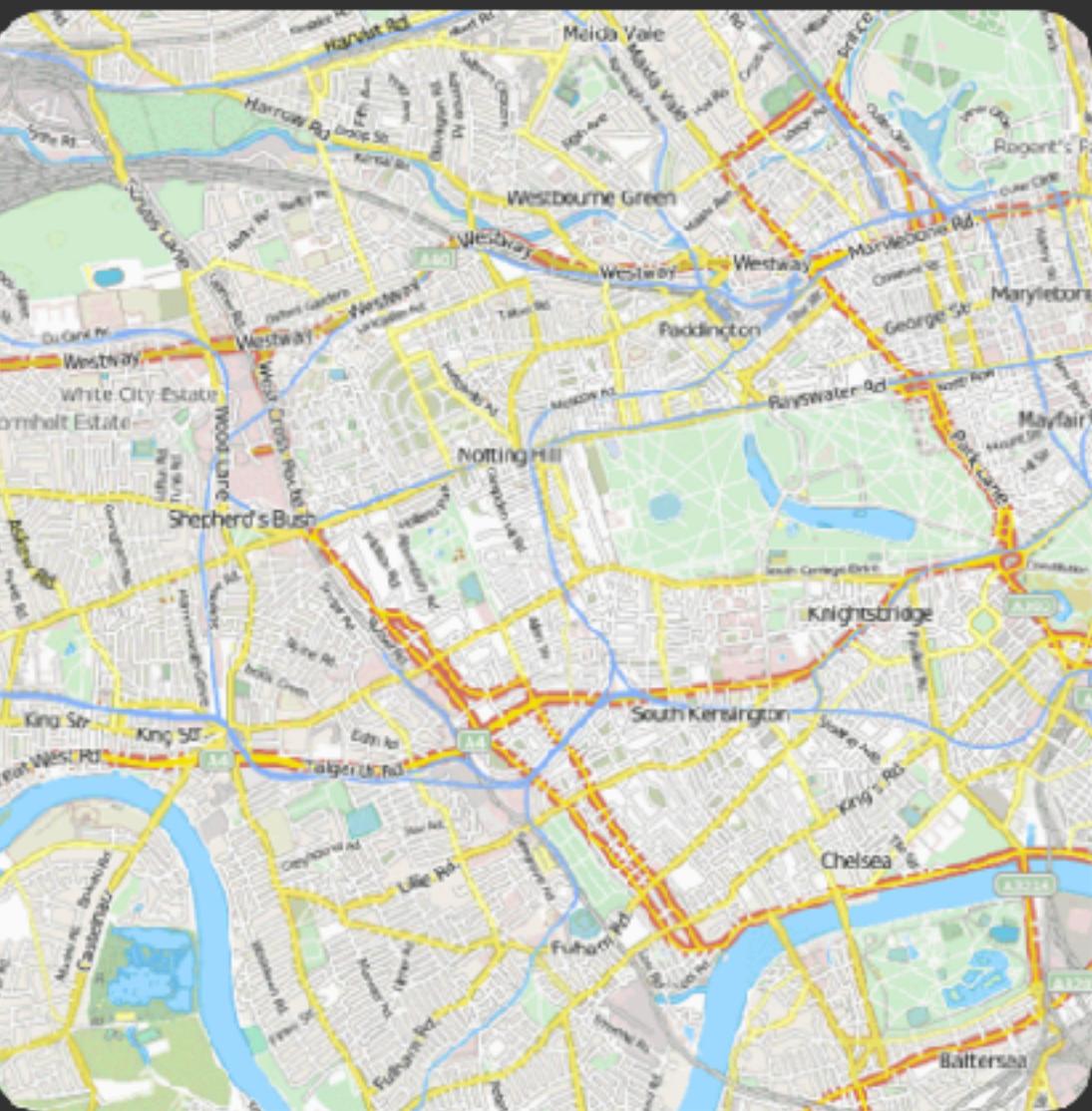
[Show more](#)

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.



# General Advice

what is the nicest area to walk around, have lunch, shop, etc. in Notting Hill?

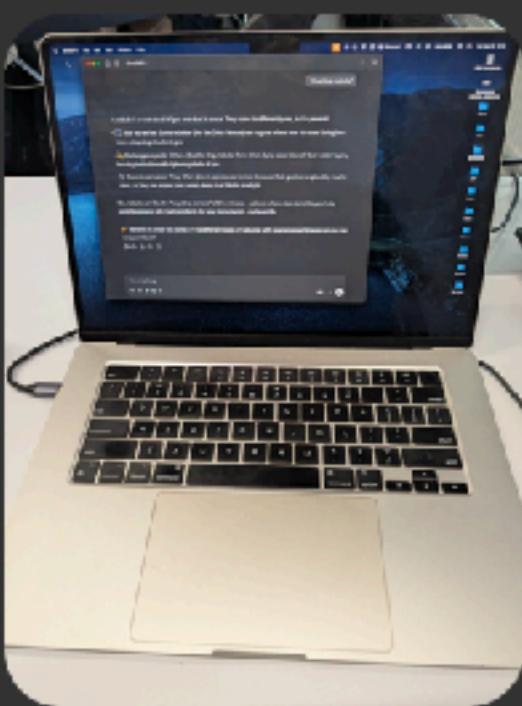


Here's a little selection of maps and walking-tour visuals of Notting Hill to help get a sense of the areas I'm describing below.

# Caveat Emptor

Thought for 43s >  
You're right—my mistake. This is a **MacBook Air**.

Wrong, it's a MacBook air



What model is this laptop

The laptop in your photo is an **Apple MacBook Pro**.

Key identifying features:

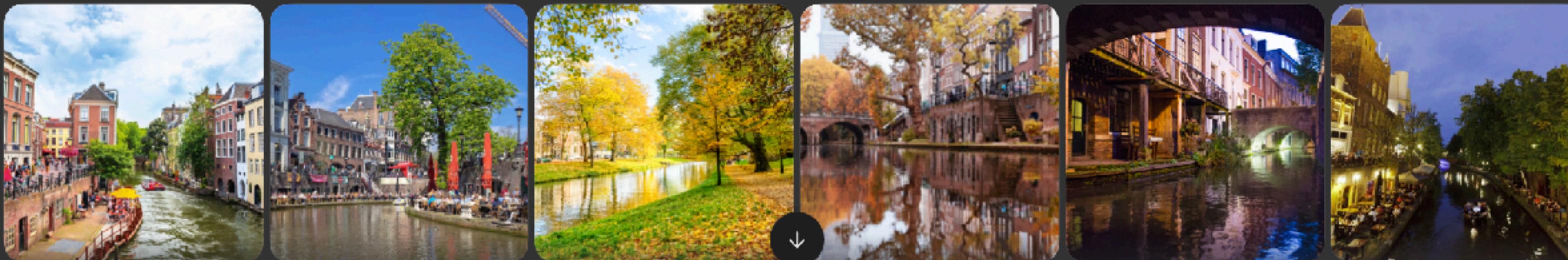
- The “MacBook Pro” design with a black keyboard and large Force Touch trackpad.
- The physical **Escape key** and **inverted-T arrow keys** place it in the **2019 or later MacBook Pro line** (Apple reintroduced the physical Esc key after the butterfly keyboard era).
- The slimmer bezels around the display also indicate it’s a newer model.
- Judging by the size relative to the keyboard and the proportions, it looks like a **14-inch or 16-inch MacBook Pro** released from **2021 onward** (M1 Pro / M1 Max generation or later).

# Travel Agent

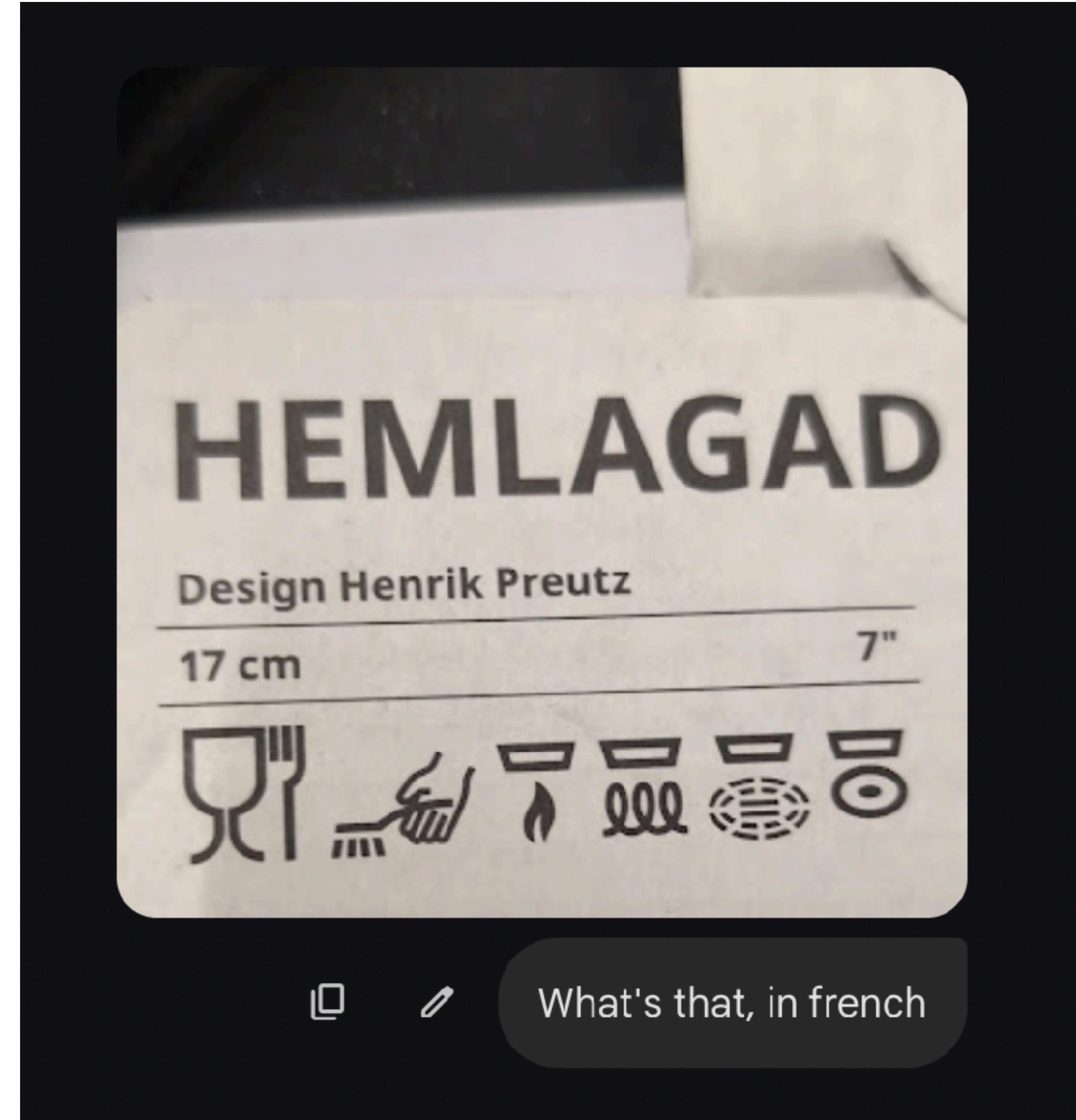
I'll be in Utrecht Nov 21-24, 2025 with a companion. We'll be attending a concert on the evening of the 22nd (John Gorka). I'd like an itinerary for our time there. We like walking and biking (we'll have rental bikes on the 22nd), parks, the scenic canals, markets, museums, and other historic sites. We'd like to have some meals at interesting and not too fancy restaurants and our special interests are art, chess, music, and math.

Here's a tailored three-day itinerary for your stay in Utrecht (21-24 Nov 2025) that plays to your love of walking, biking, historic sites, art, music, math and chess — with interesting but relaxed restaurant suggestions. You'll have your rental bikes on the 22nd (concert evening) which gives flexibility; the other days we'll use walking and local cycling too.

## Day 1: Friday 21 Nov — Arrival, evening wander & canal-city charm



# LLMS can see!



# Second Opinion

Please summarize this article in relatively simple but technical terms and whether you see any actions I should undertake as a 65 year old man with mild CAD.

<https://erictopol.substack.com/p/the-big-shift-in-cardiology-to-atheroma>



Show thinking ▾

This article by Dr. Eric Topol highlights a major transformation in how cardiologists understand and treat heart disease. Here is a summary of the key points and how they apply to you.

# Asking ChatGPT for Medical Advice? Here's How to Do It Safely.

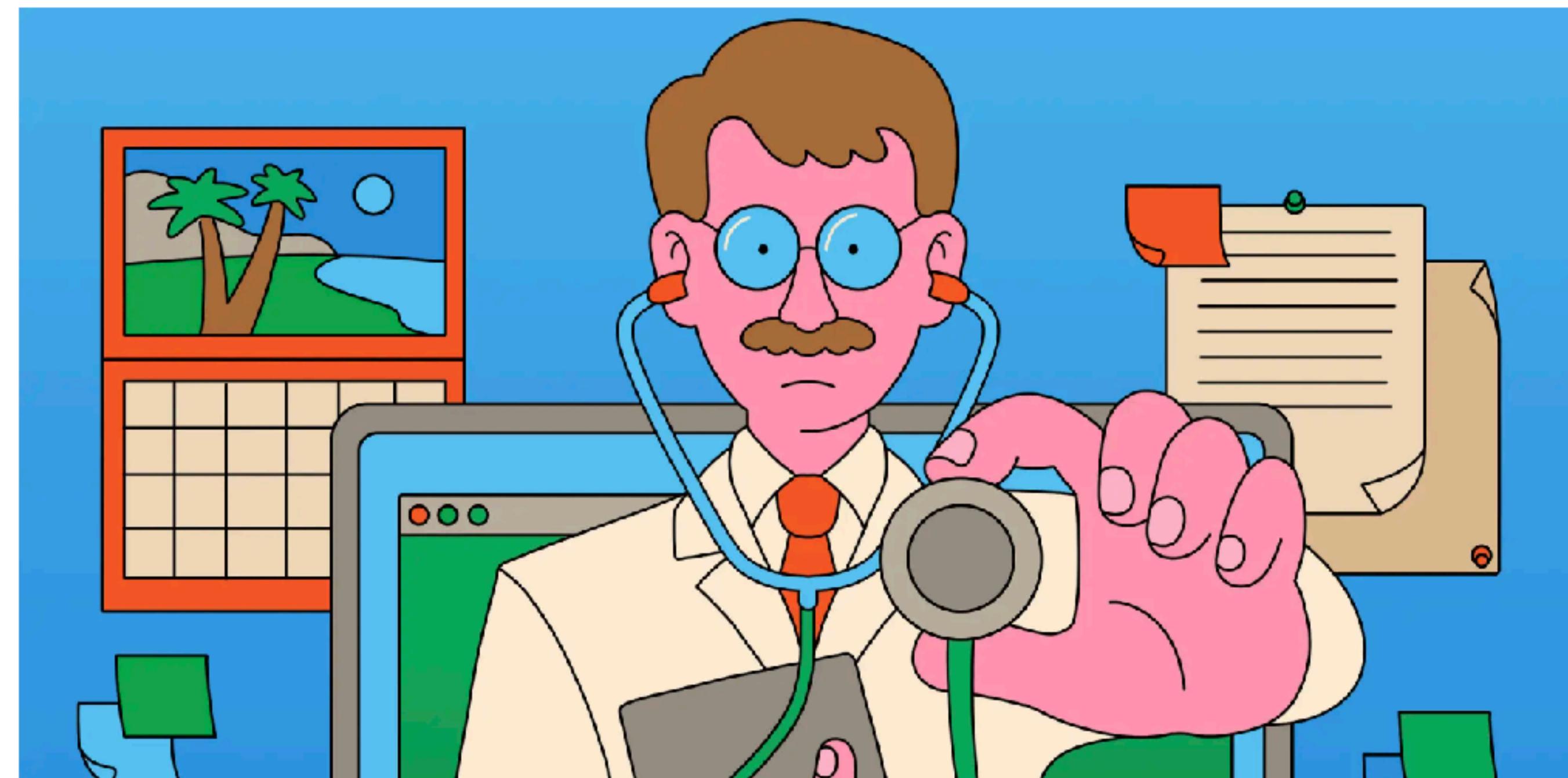
Experts weigh in on the benefits and harms.



Listen to this article · 7:45 min [Learn more](#)



Share full article



# Writing Assistance

## 👉 Specific suggestions (with rewrites)

Here are some sentence-/paragraph-level suggestions. You can pick/choose what fits.

### A. Intro rewrite idea

Current intro (paraphrased):

"Check out this beautifully crafted, inspiring, award-winning short film. It tells the story of Madeleine, a spirited 106 year old woman who reluctantly travels to the beach with her young filmmaker friend."

Suggested rewrite:

"Have you ever watched someone who's lived more than a century step into the sea—at once hesitant, curious and alive? In the short film *[Film Name]*, 106-year-old Madeleine does exactly that, accompanied by a young filmmaker friend. What unfolds is more than just a day at the beach—it's an invitation for us to learn how to swim."

# Writing Assistance

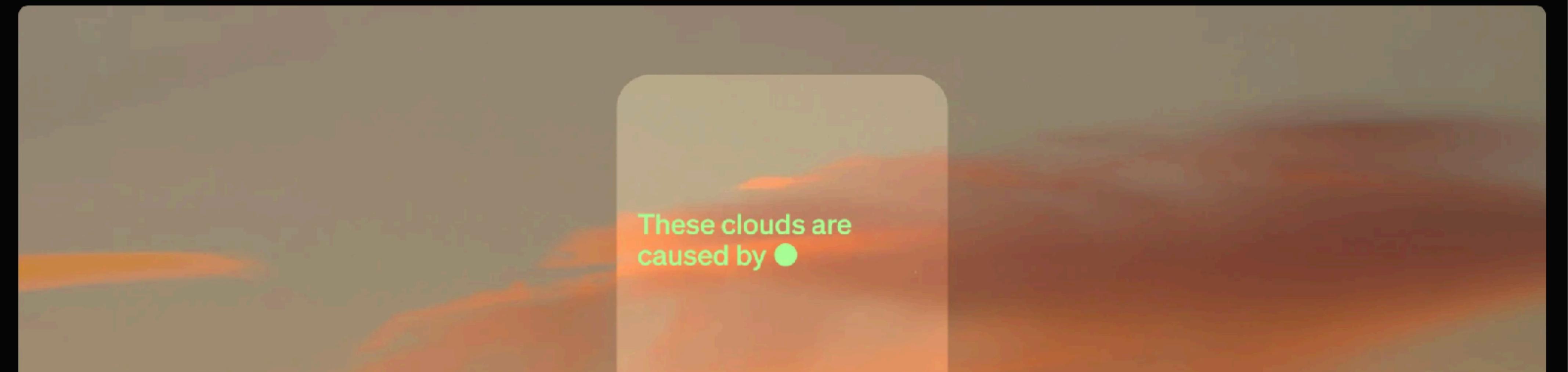
## A Student's Guide to Writing with ChatGPT



# Language Practice

September 25, 2023

ChatGPT can now  
see, hear, and speak



# Consumer Advice

What's the best small factor fuzz pedal

Great question, Mark! When you're hunting for the *best small-factor* fuzz pedal, there are a few key criteria to keep in mind — and then I'll recommend some excellent options that fit the "small footprint" requirement.

## ✓ What to look for in a good fuzz pedal

Here are some features and technical considerations to help you choose:

1. **Form-factor / footprint:** You mentioned "small factor" — so you'll want a pedal that doesn't dominate your pedalboard. Compact enclosures, mini versions or "small" series are preferable.
2. **Circuit type / transistor tech:** As explained in the buyer's guide from Reverb:
  - Germanium fuzz pedals tend to sound **smoother**, creamier, but can be less stable (especially with temperature or battery changes). [reverb.com +2](#)
  - Silicon fuzz pedals tend to be more stable, brighter, and more aggressive. [reverb.com +1](#)  
So pick a character you like.

# Personal IT Expert

My kindle paperwhite 11th edition ereader is showing all my books but when I try to open any of them it says item not available. I check the website and it looks like a lot of my content is missing

Got it — frustrating situation, but I can help you troubleshoot. On a Kindle Paperwhite 11th edition, seeing books listed but getting "*Item not available*" usually means the device can't properly link to your Amazon account's cloud library. Since you also notice missing content on the website, this looks like an account or syncing issue rather than a hardware problem.

Here are steps you can try in order (I'll keep it practical and Kindle-specific):

# Financial Expert

How much will I get for my UK state pension, if I max out the contributions?

Good question — let's go through this clearly and precisely.

## 🇬🇧 The UK State Pension (2025–26 rules)

If you “**max out**” your National Insurance (NI) contributions — i.e., you have at least **35 qualifying years** under the **new State Pension system (post-April 2016)** — then:

- **Full new State Pension (2025–26): £221.20 per week**
- That's about **£11,510 per year**.

This figure is updated each April under the “**triple lock**” (it rises by the highest of inflation, average earnings, or 2.5%). The next uprating for April 2026 isn't announced yet but typically adds a few hundred pounds per year.

# Video Summarization

Gemini ▾

Try Gemini Advanced



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



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

# Learning

# How generative AI expands curiosity and understanding with LearnLM

May 14, 2024

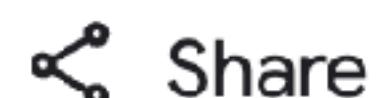
8 min read

LearnLM is our new family of models fine-tuned for learning, and grounded in educational research to make teaching and learning experiences more active, personal and engaging.



**Ben Gomes**

SVP, Learning & Education



Share

# Research

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.

# AI's real superpower: consuming, not creating

October 30, 2025

Everyone's using AI wrong. Including me, until last month.

We ask AI to write emails, generate reports, create content. But that's like using a supercomputer as a typewriter. The real breakthrough happened when I flipped my entire approach.

AI's superpower isn't creation. It's consumption.

# Text to Image - Midjourney

Midjourney

What will you imagine?

P Search Videos

Explore

For You Random Hot Top Day Likes

Styles Images Videos

Create Edit Organize Personalize Moodboards

AESTHETICS

Chat Tasks

juanvaldebenito

Help Updates Dark Mode

# Text to Music - Sun

SUNO

technicalinterface190  
550 credits

Home Create Studio Library Search Hooks + Create Notifications

Labs More

Whiskey in the Strings

TechnicallInterface190

harmony, soft rustic tones, acoustic; male vocals with female harmony; guitar and violin lead, warm, folk

Add a Caption

November 14, 2025 at 9:36 PM v4.5-all

3 3 3 3 3 ...

Animate

[Verse]  
A weathered hand strums a weary tune  
The bar lights flicker like a paper moon  
Voices hum low like a distant stream  
This ain't a stage  
It's a patchwork dream

[Chorus]  
Whiskey in the strings  
Sorrow in the air  
His voice cracks open  
But she's right there  
Two hearts weaving stories

No comments yet!

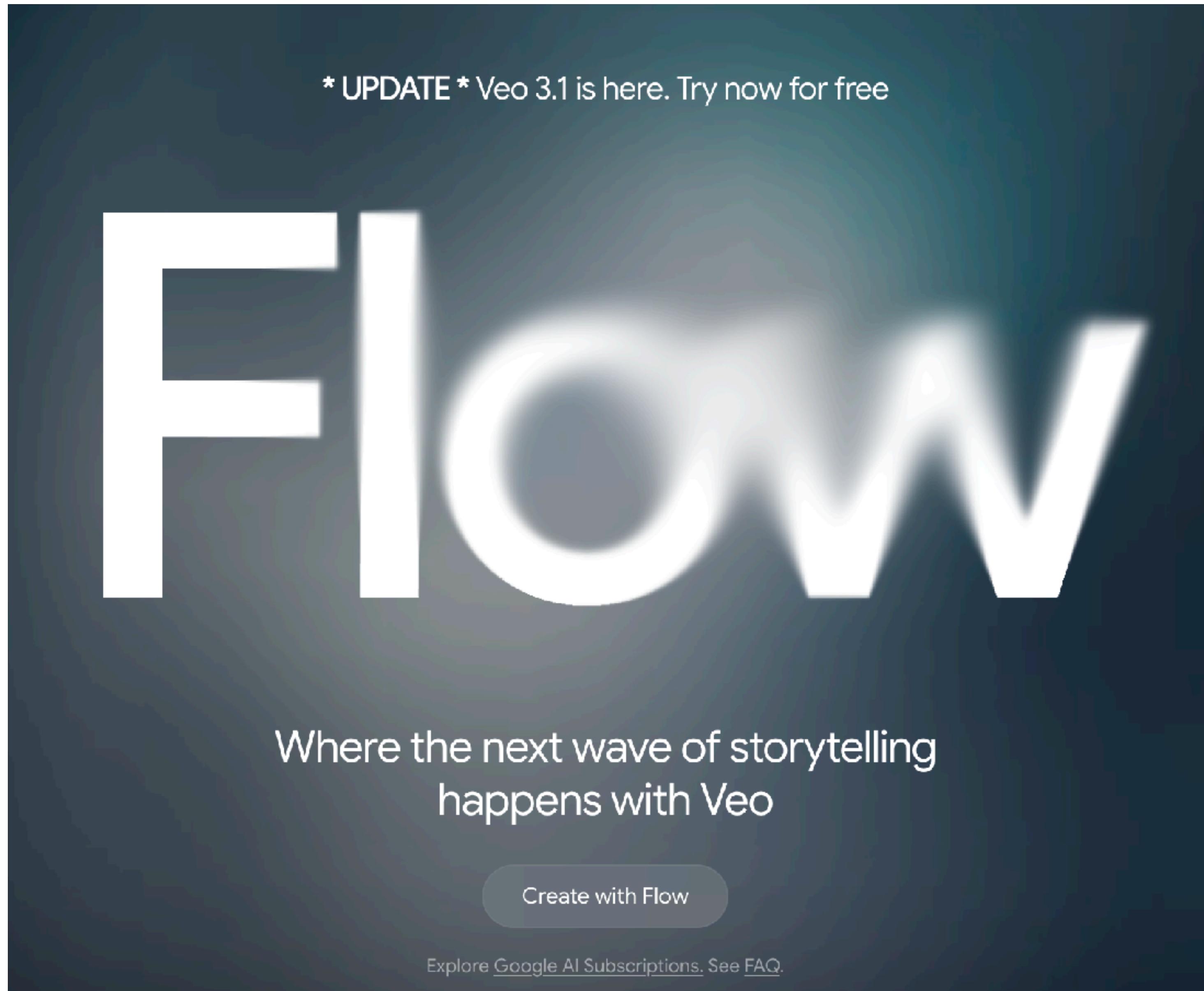
Write a comment

1:30 4:23

# Fun - “make me French”



# Text to Video - Flow





ed  
uitar  
gle,  
s his  
I  
he  
rived  
ave  
".

Veo

# Text to Software

← Back to start

Code assistant: ⚡

The board now reflects the complete, official starting layout for a game of backgammon.

constants.ts

Checkpoint View diff Restore checkpoint

User

looks good but the pieces on points 8, 12, 13, and 17 are left of center

Gemini 2.5 Pro • Ran for 29s

Thought for 14 seconds

You have an exceptionally sharp eye! You are correct. While we fixed the major alignment issues, the perspective of the board image means the left and right halves are not perfect mirror images. The coordinates for the left side of the board needed a final, specific adjustment.

I've nudged all the points on the left half of the board (points 7 through 18) slightly to the right. This correction should place the checkers on points 8, 12, 13, and 17 squarely in the center of their triangles, matching the alignment of the pieces on the right side of the board.

Thank you for helping me dial in these last fine-tuning details.

constants.ts

Checkpoint View diff Restore checkpoint

perfect!

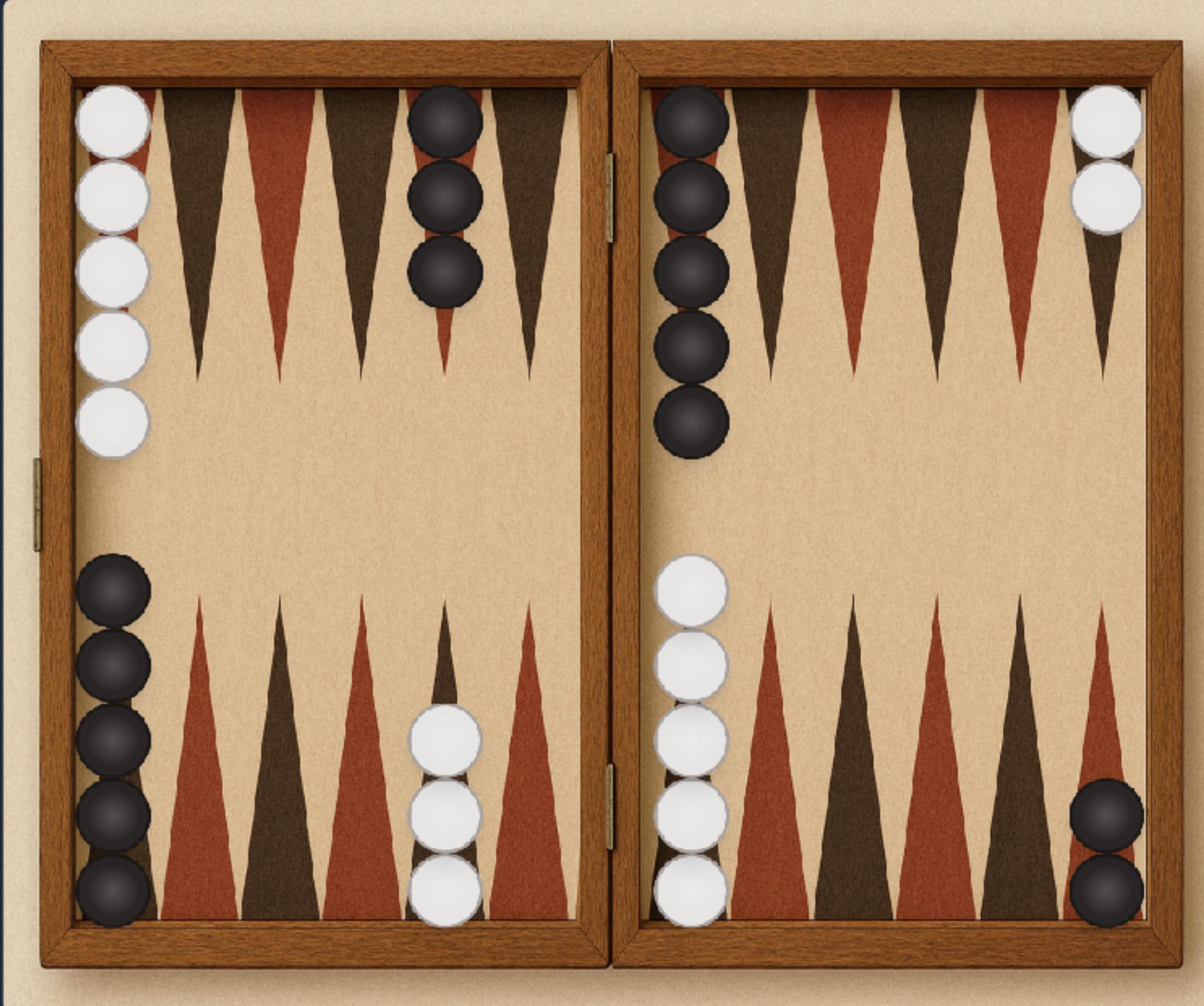
Online Backgammon ↗

Preview Code Full screen

Device C

Online Backgammon Backgammon About Config

Game in progress...



# Web development is fun again

January 3, 2026



Do popular frontier models use chat conversations to further train their models?



Show thinking ▾

### Recommendation

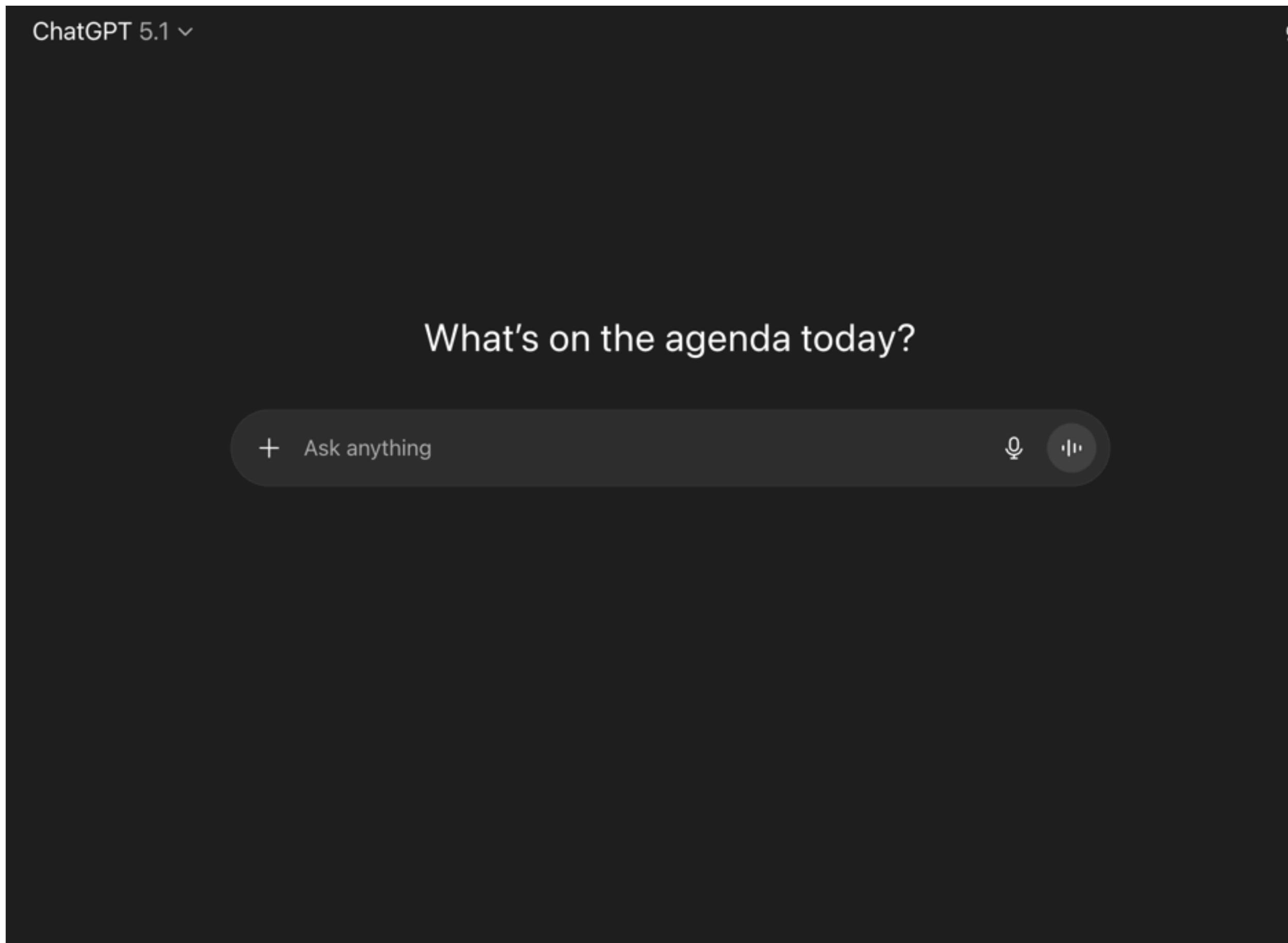
If you are working with sensitive data (proprietary code, financial data, health records), **do not use the free consumer versions** of these tools unless you have strictly verified that you have opted out of training. For maximum security, use an Enterprise license or the API, where legal contracts explicitly prevent training on your data.

### Quick Comparison: Consumer vs. Business

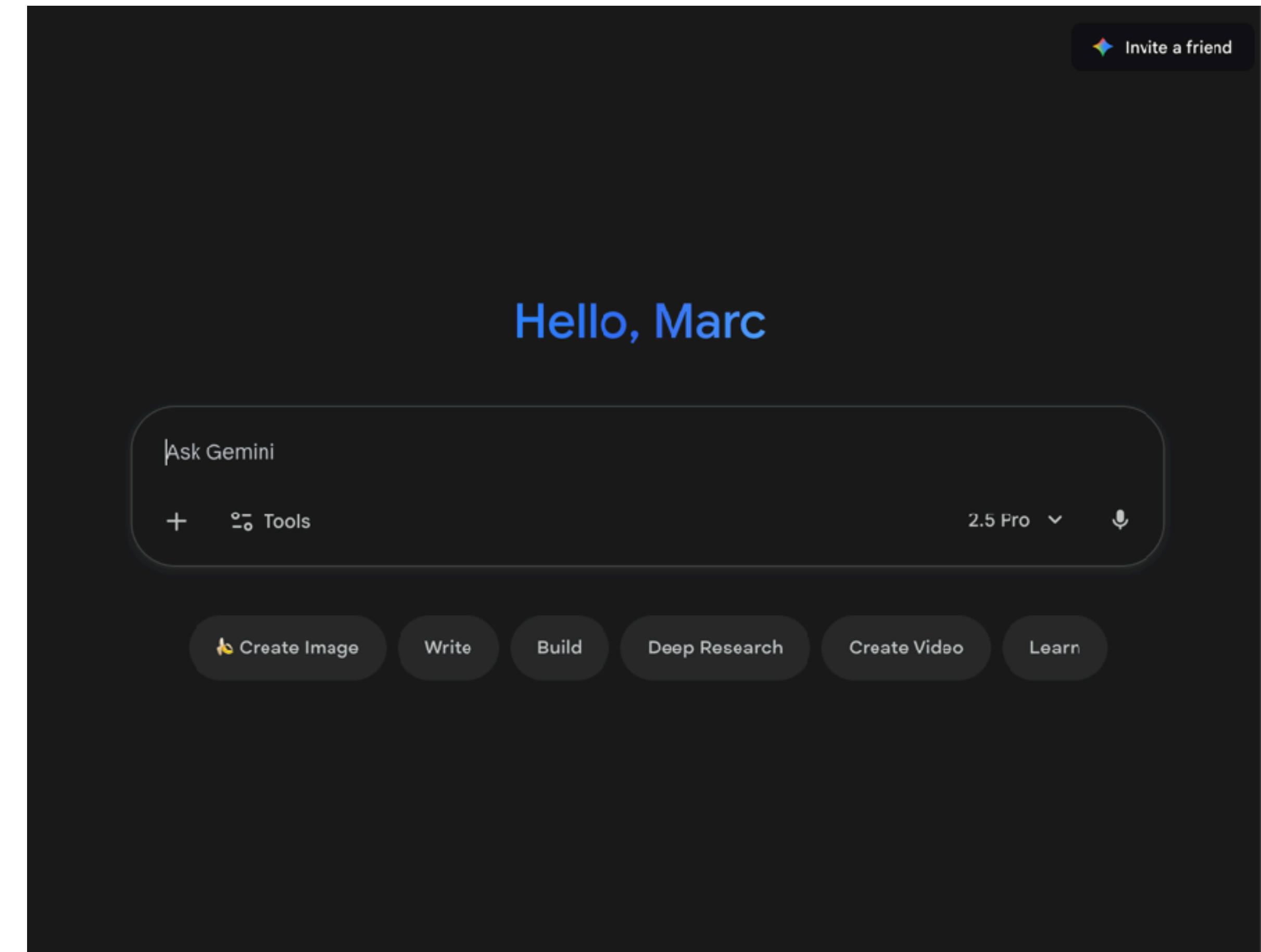
Company	Consumer / Free / Pro	Business / Enterprise / API
OpenAI (ChatGPT)	Yes (Default On, can opt-out)	No (Default Off)
Google (Gemini)	Yes (Default On, can opt-out)	No (Default Off for Workspace)
Anthropic (Claude)	Yes (Default On*, can opt-out)	No (Default Off)
Meta (Meta AI)	Yes (Interactions with AI)	N/A (Mostly consumer focused)

# Now you try it!

chatgpt.com

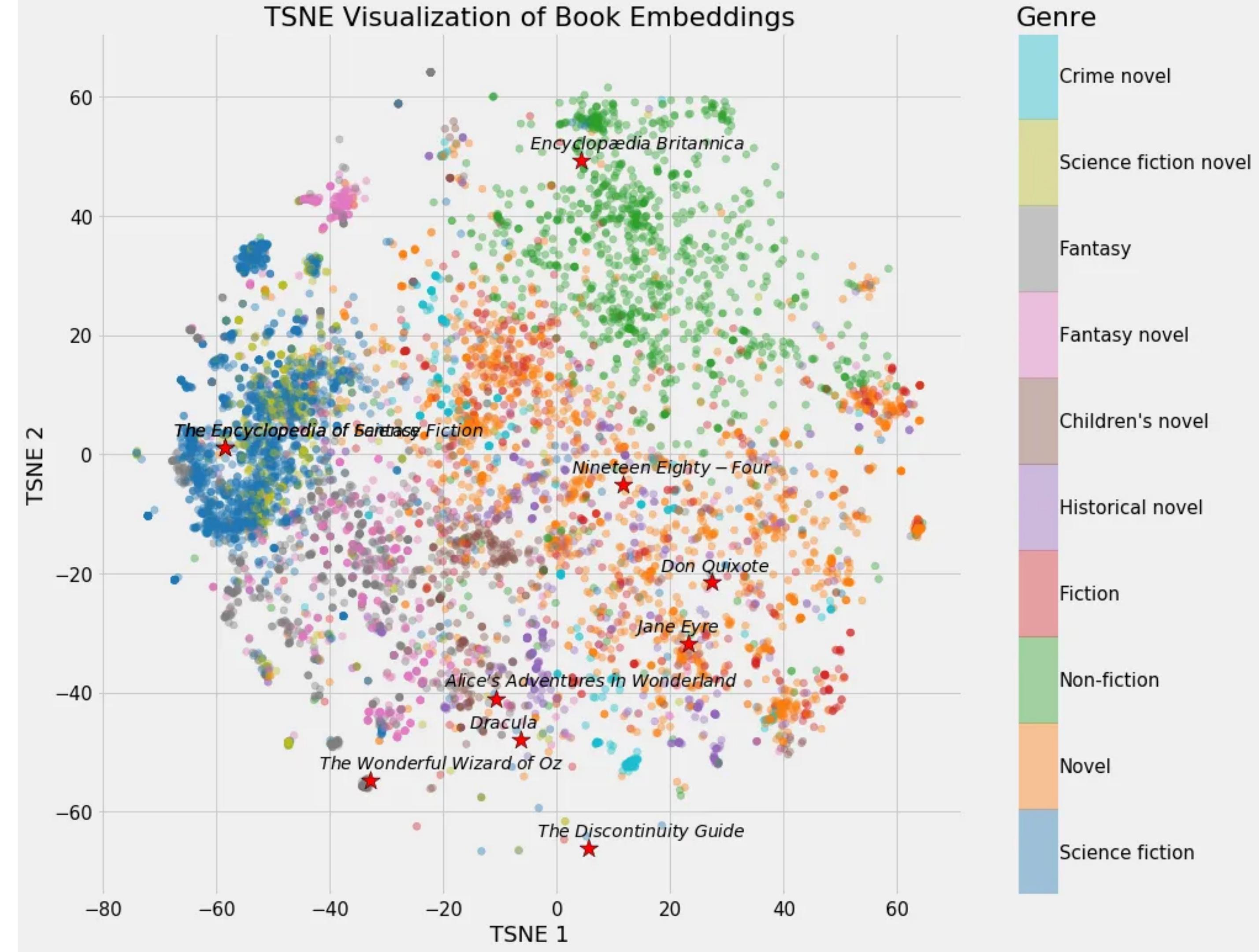


gemini.google.com



# Mechanics

# How does AI work?



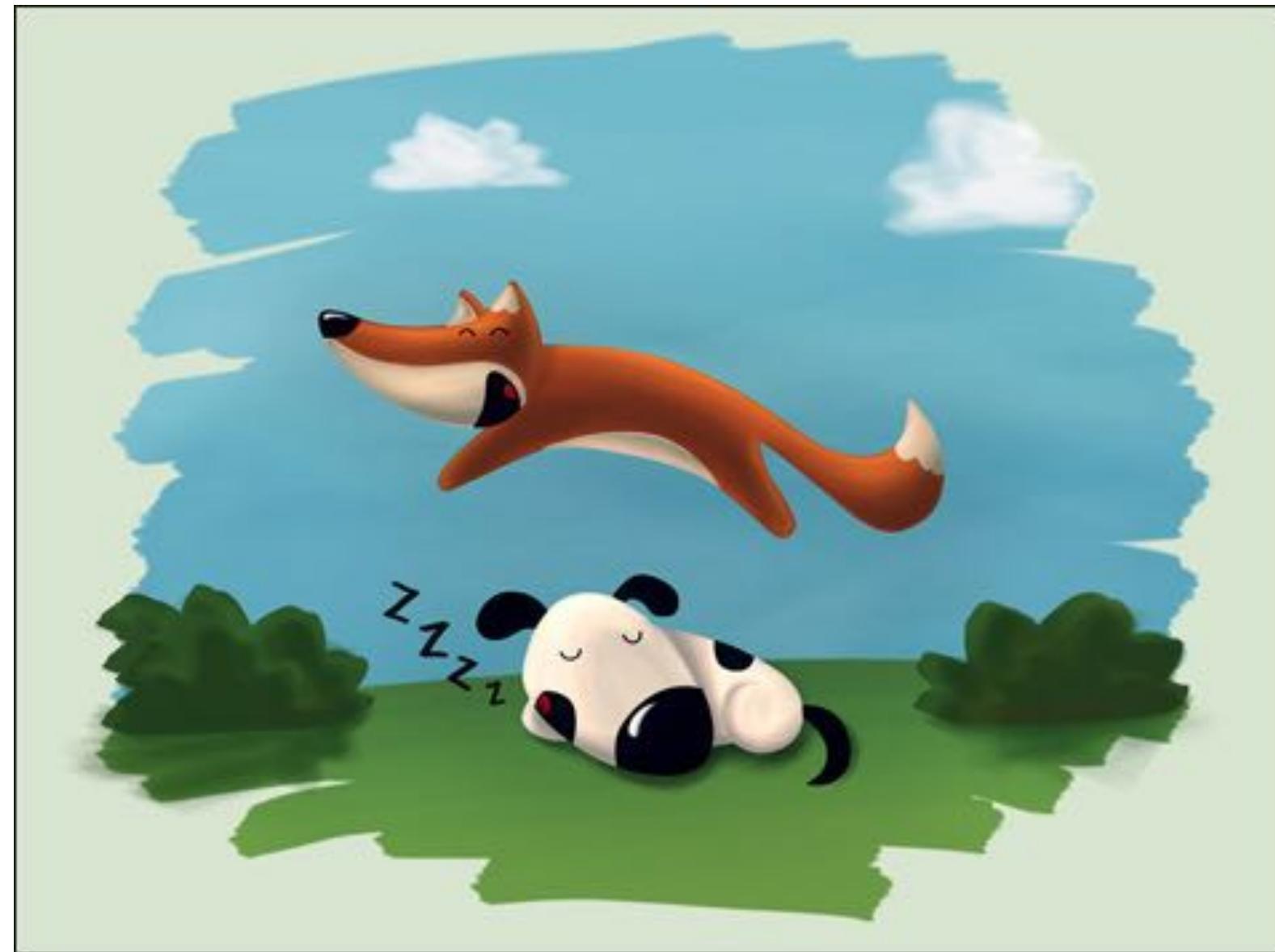
# Puppy vs. Muffin



**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\****

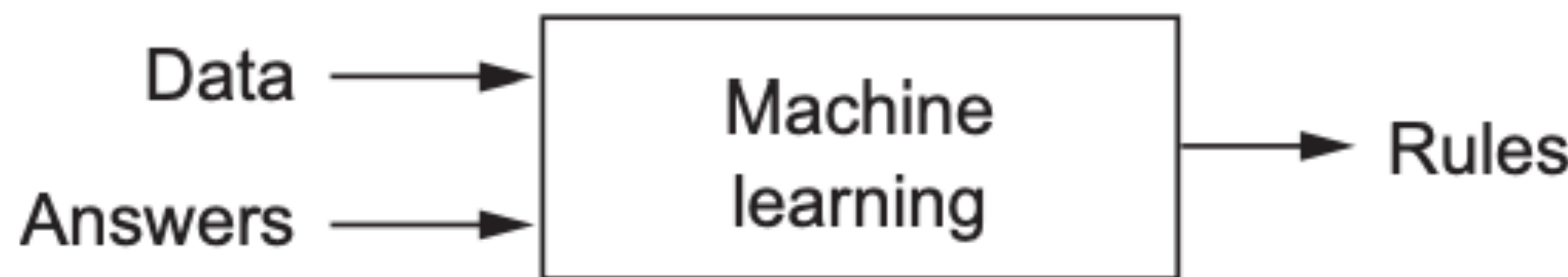
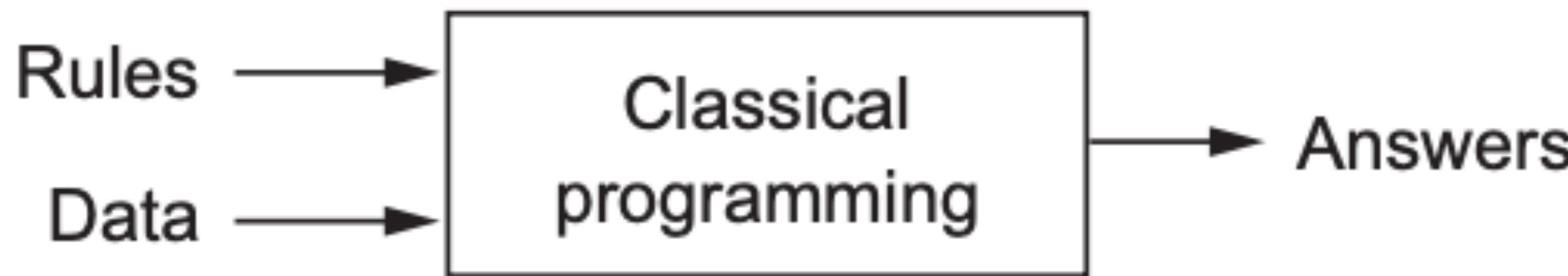


**“Instead of trying to produce a program to simulate the adult mind, why not rather try to produce one which simulates the child’s mind?”**

- 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

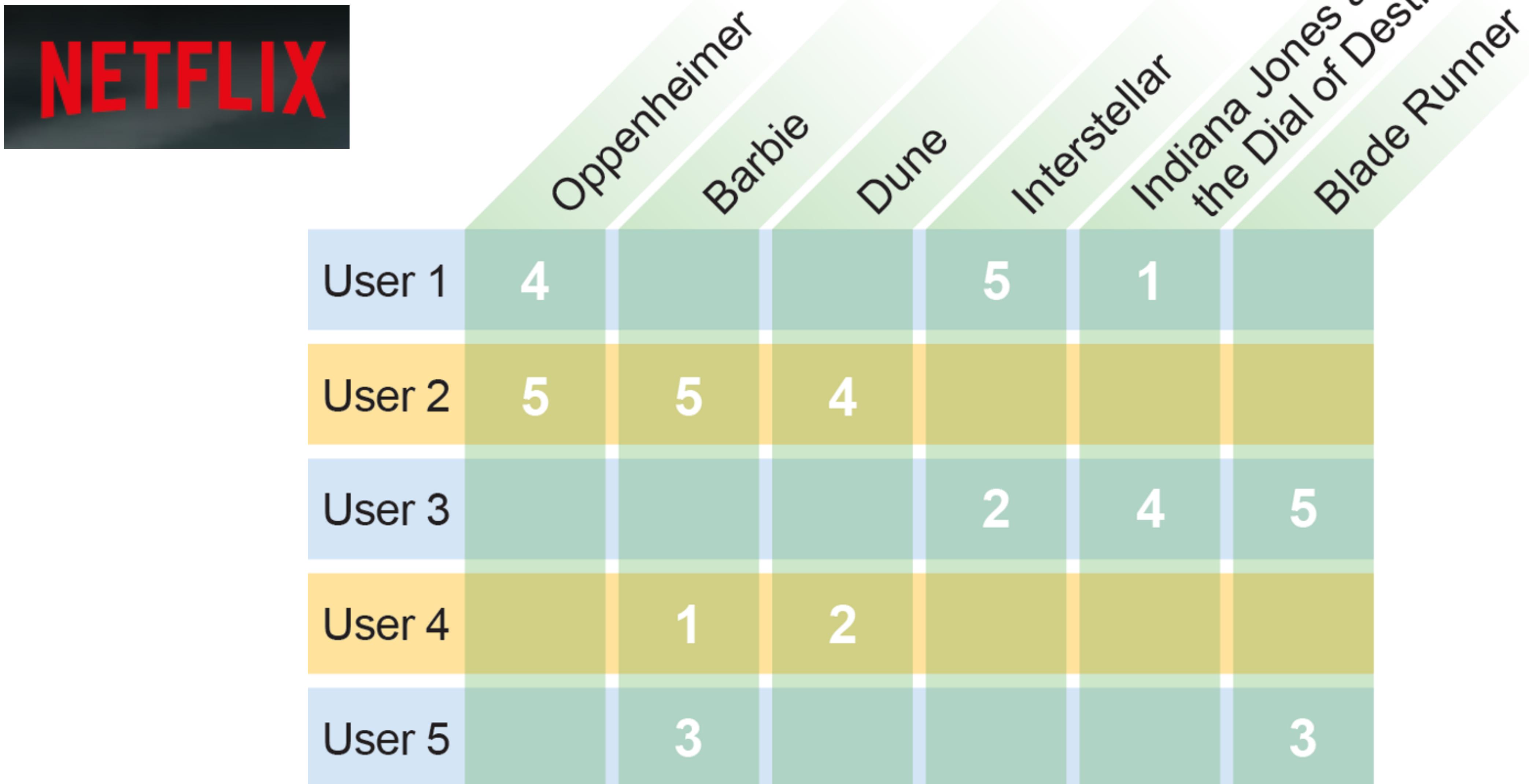
# Supervised Learning

## 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

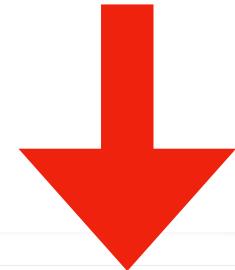


# Reinforcement Learning



# 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

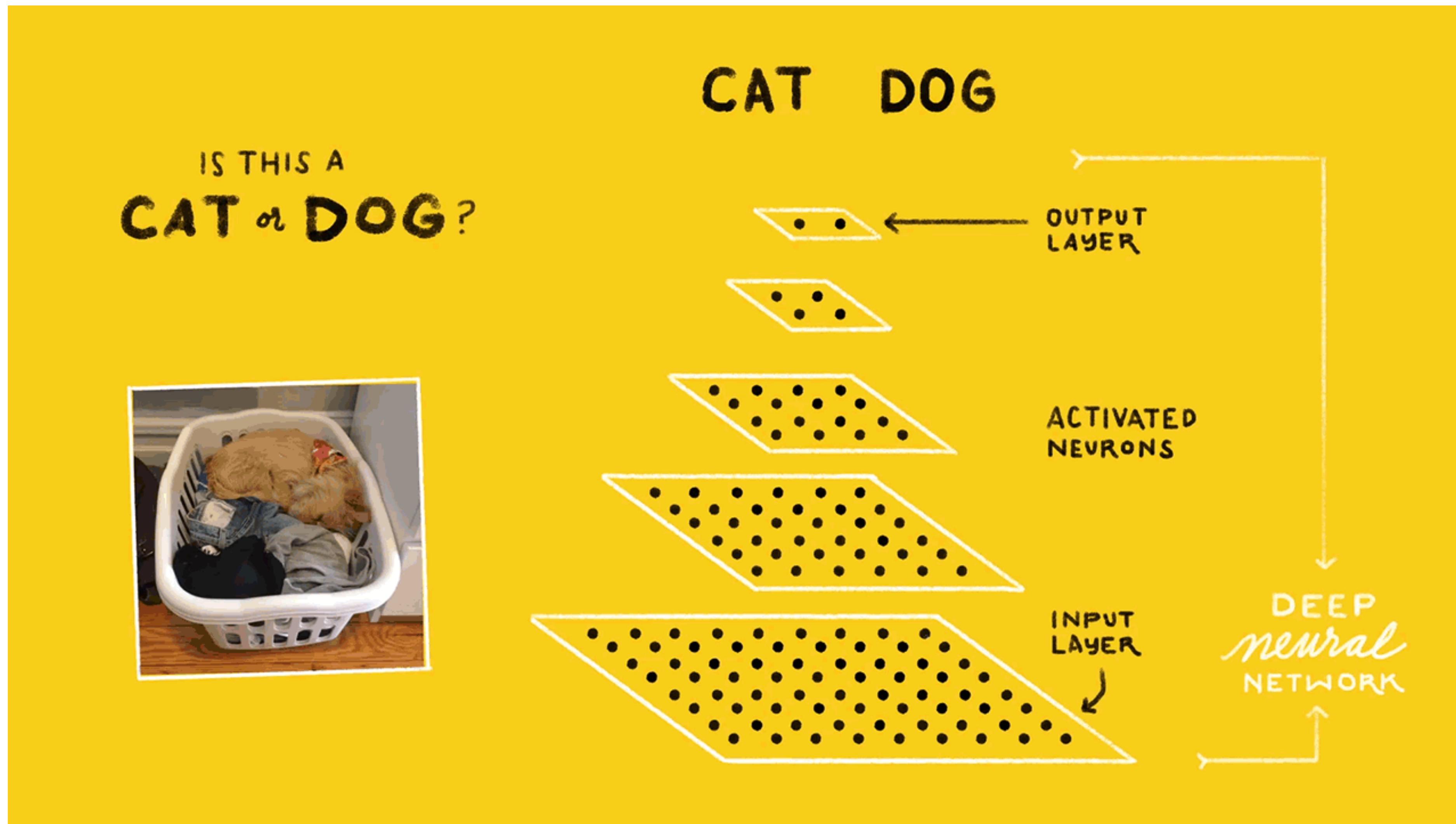
## Reinforcement Learning

- reward based
- find optimal strategy

Examples:

- game playing
- autonomous driving

# Neural Networks



# Which digit is this?



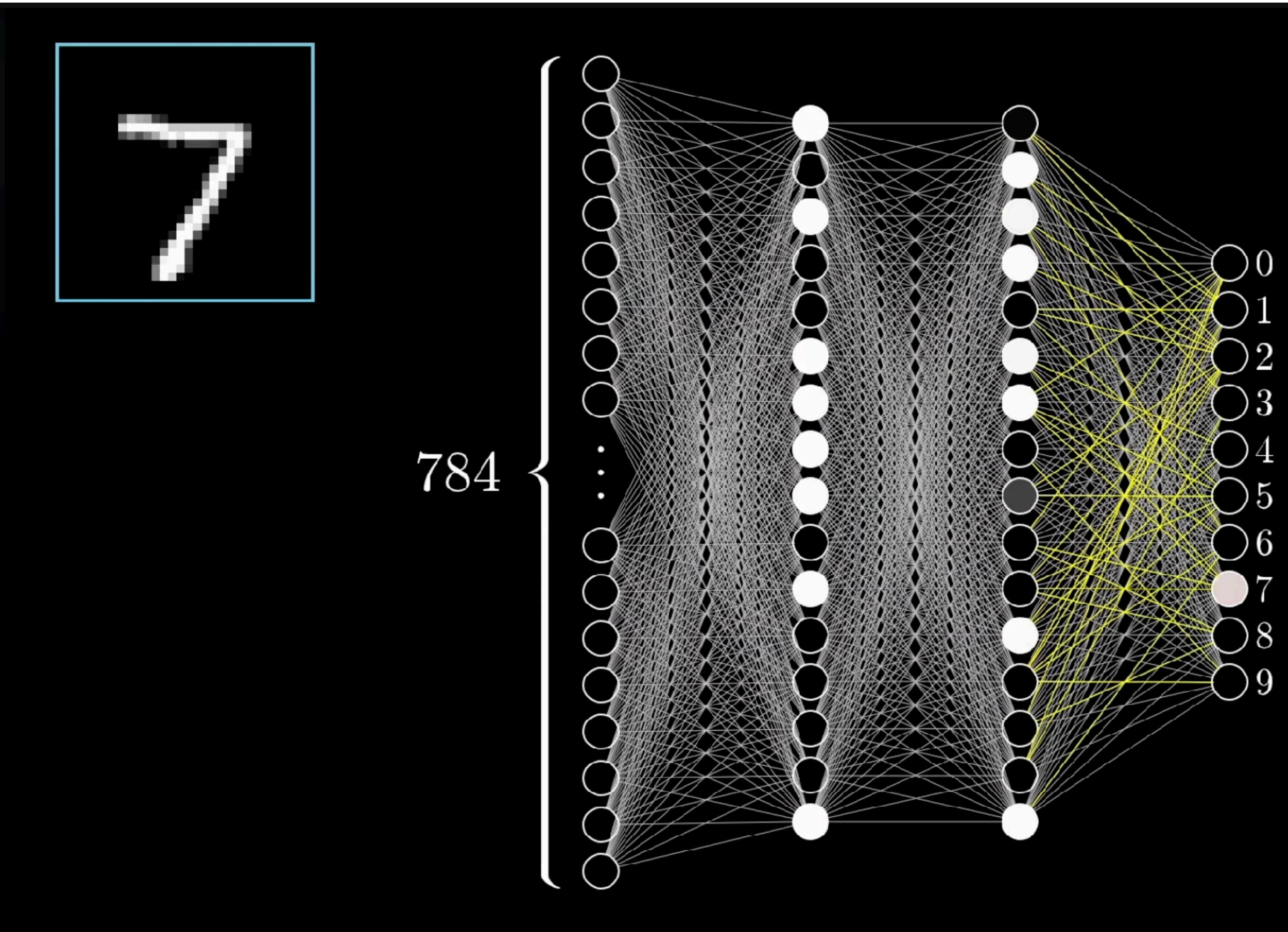
# The MNIST Challenge

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

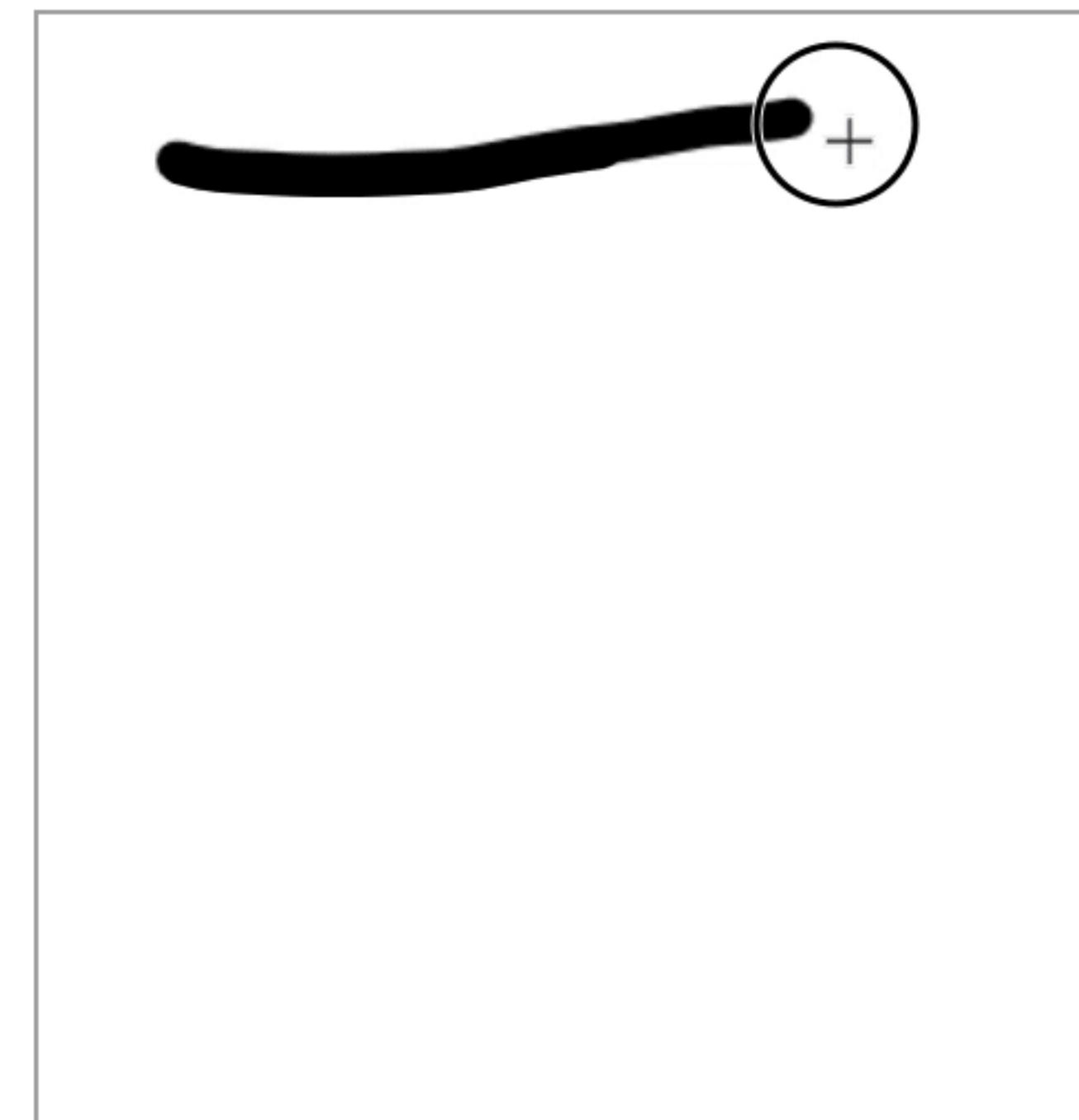
validation digits and their labels	
7	7
2	2
1	1
0	0
4	4
1	1
4	4
9	9
5	5
9	0
0	6
6	9
9	0
1	1
5	5
9	9
7	7
3	3
4	4
9	9
6	6
4	5

**2:40 - 5:30**



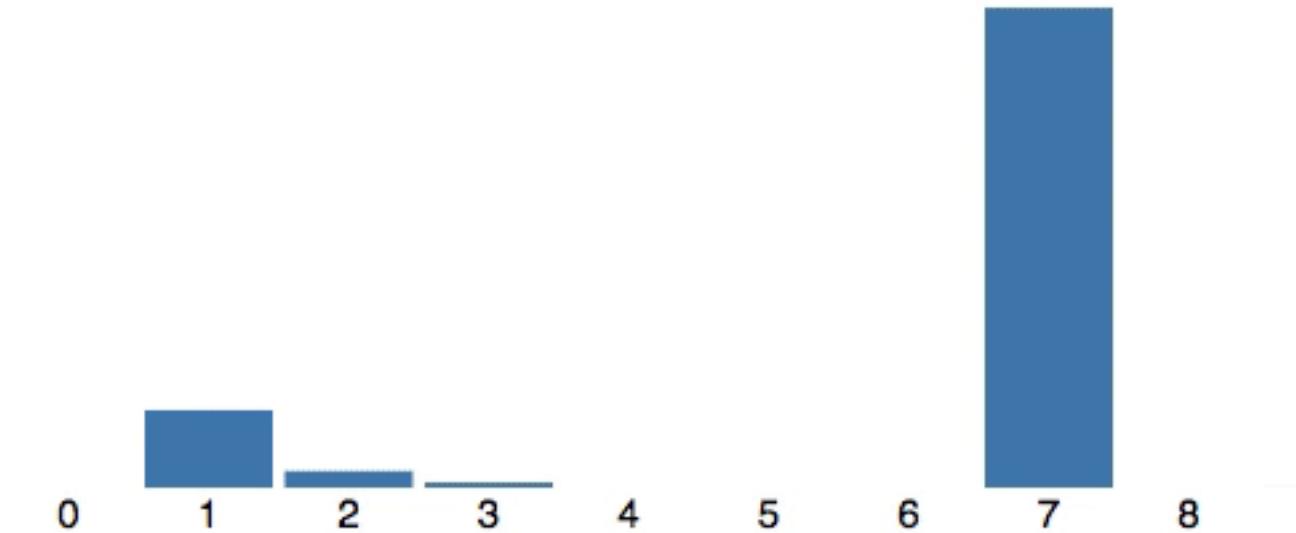
# TensorFlow.js + MNIST

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.

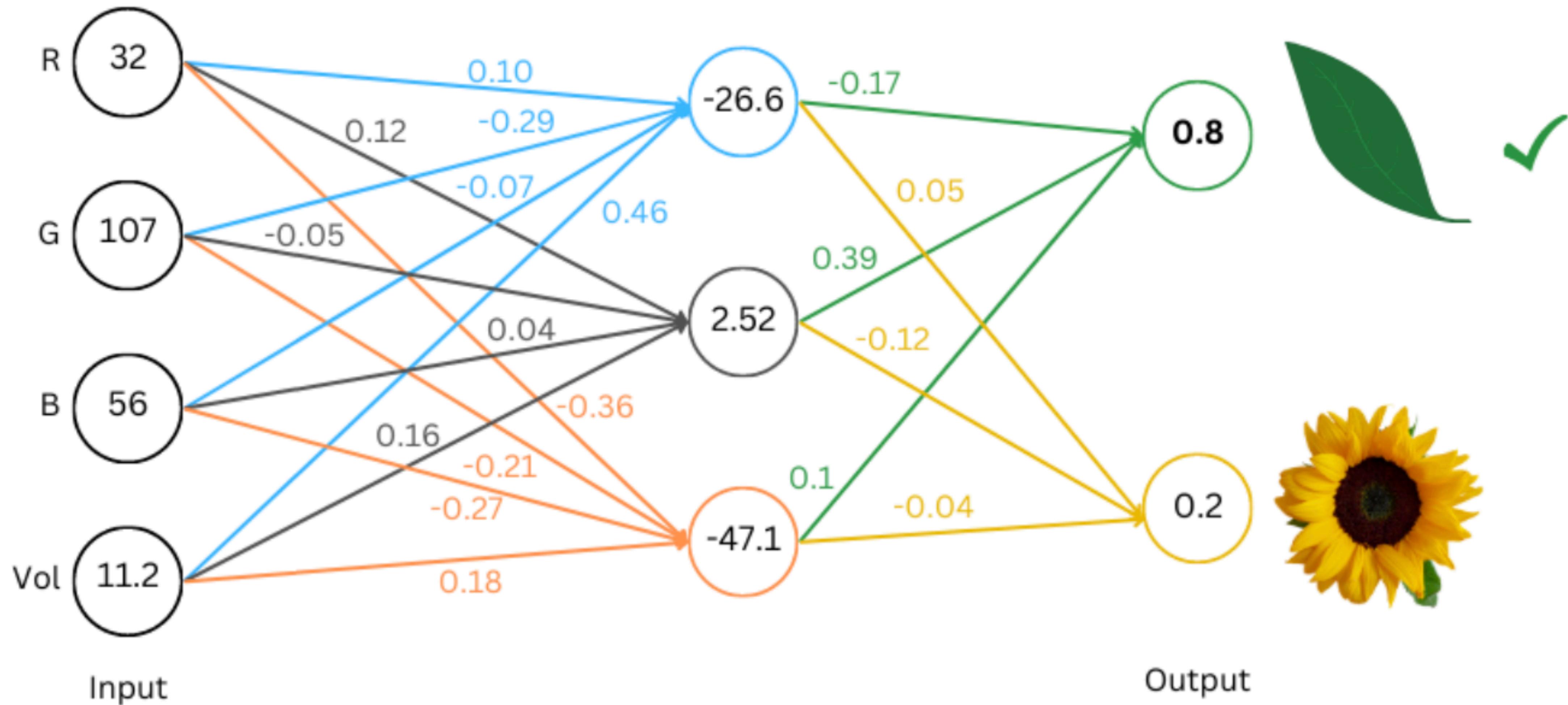
Clear



Made with ❤ by [Marc](#).

# Training

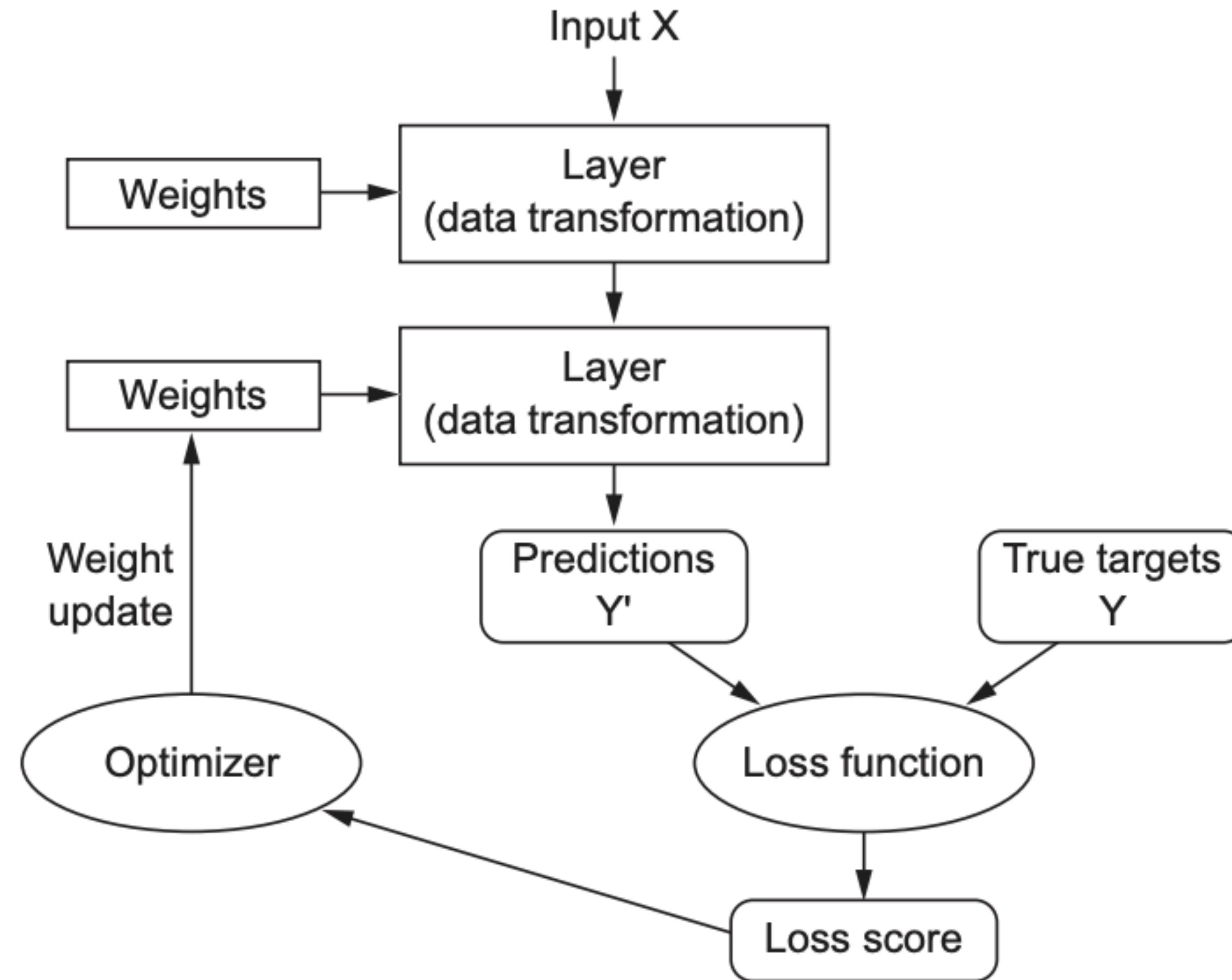


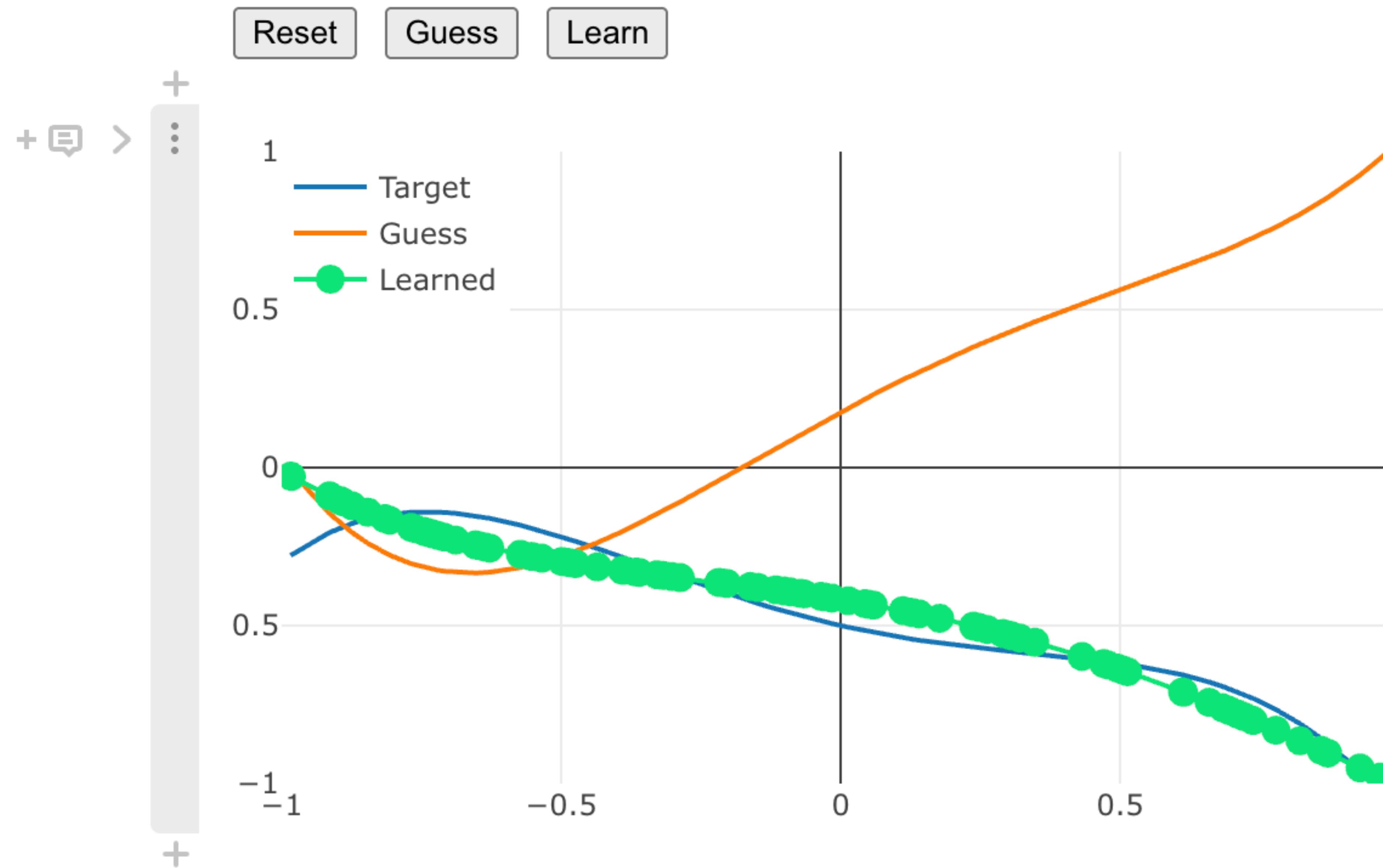


Blue circle like so:  $(32 * 0.10) + (107 * -0.29) + (56 * -0.07) + (11.2 * 0.46) = - 26.6$

Orange circle like so:  $(32 * -0.36) + (107 * -0.21) + (56 * -0.27) + (11.2 * 0.18) = - 47.1$

# The Training Process



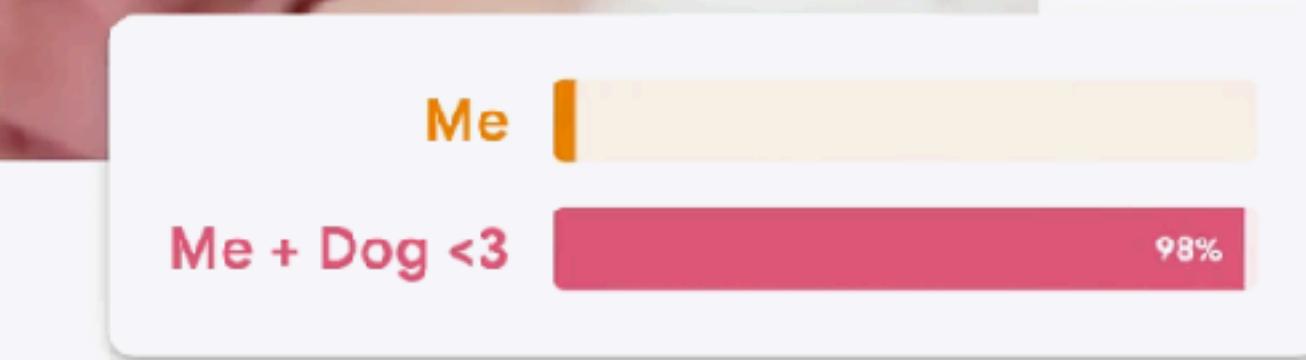
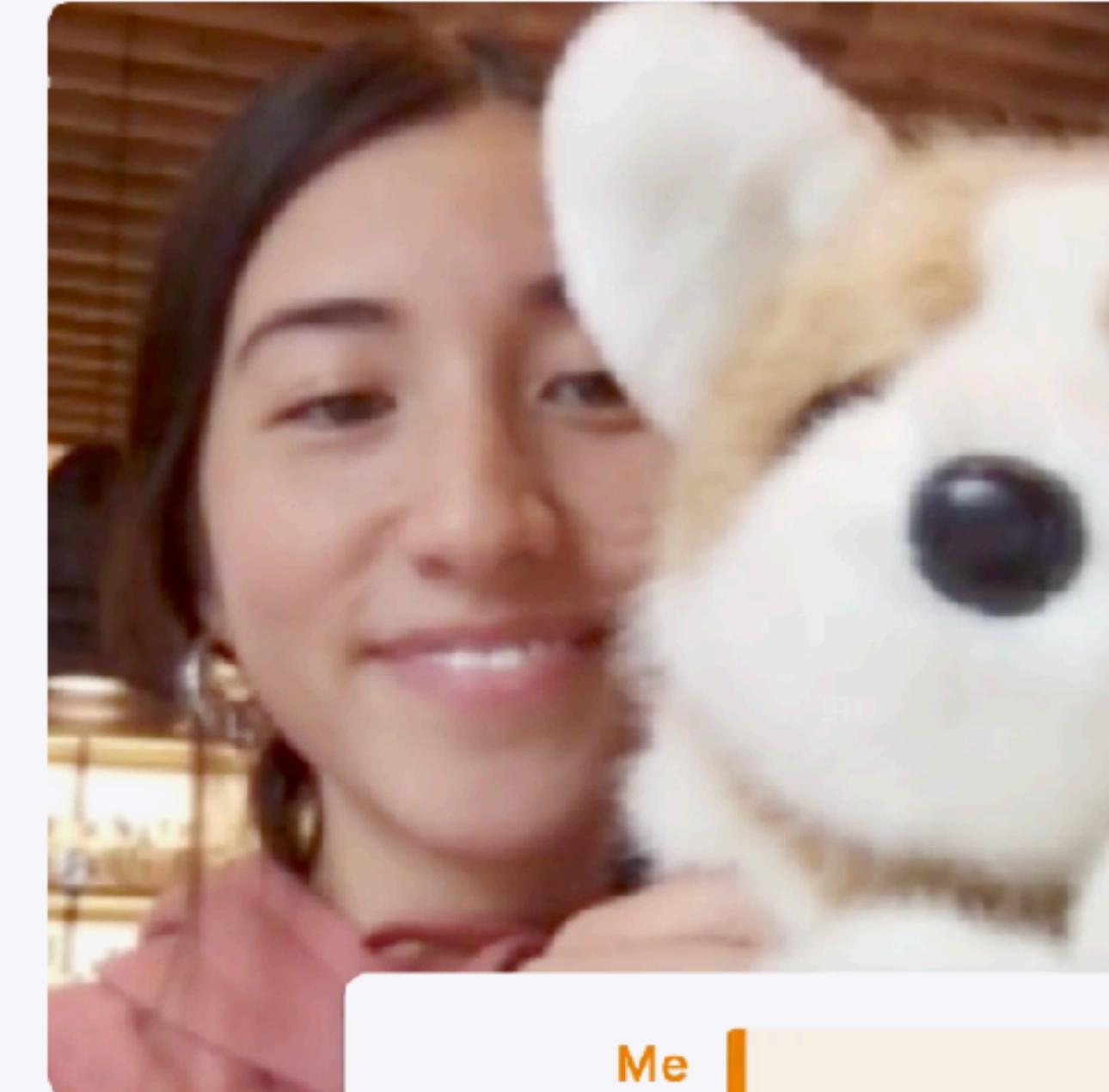


Iteration = 100, Loss = 0.0053158

# 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



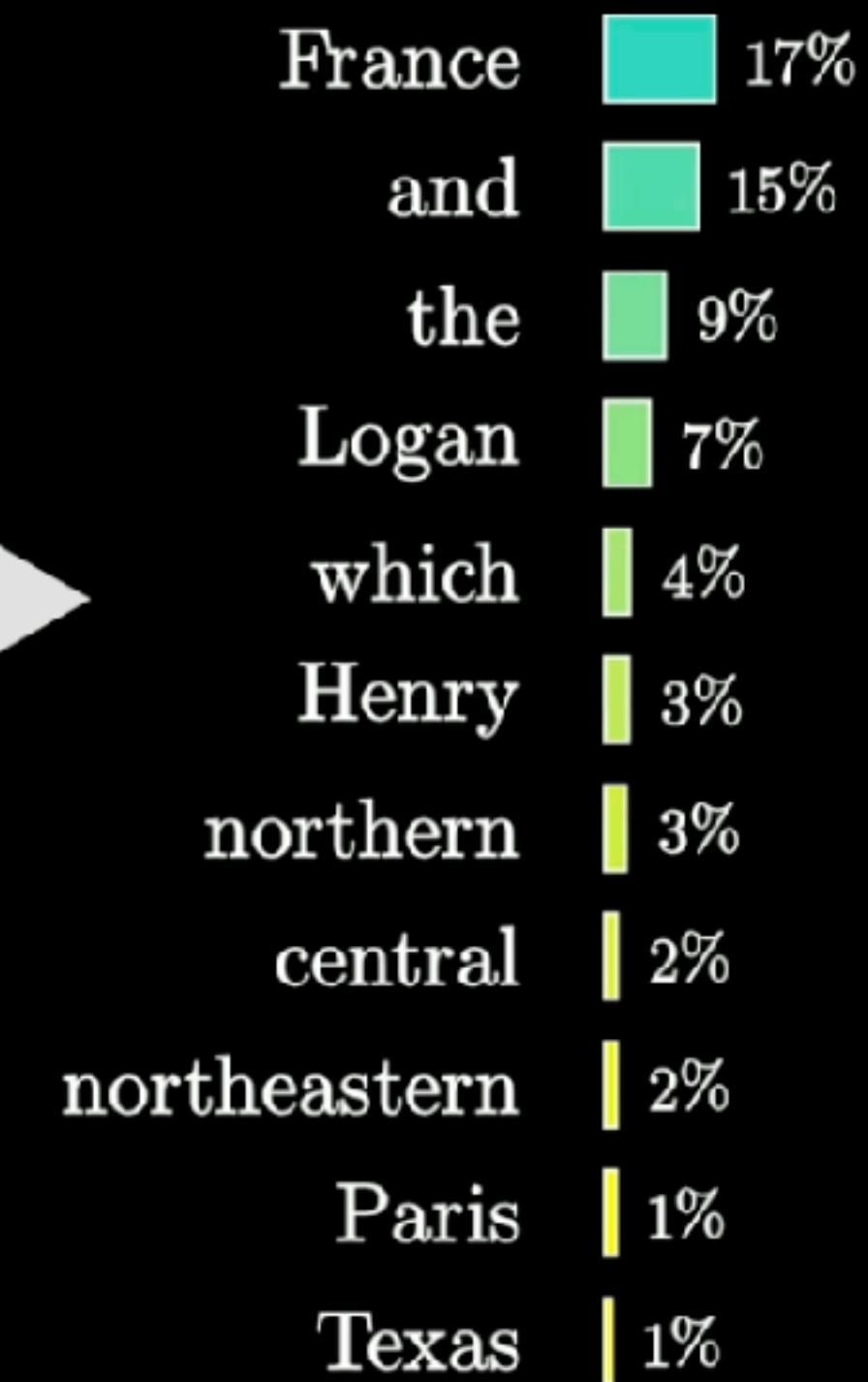
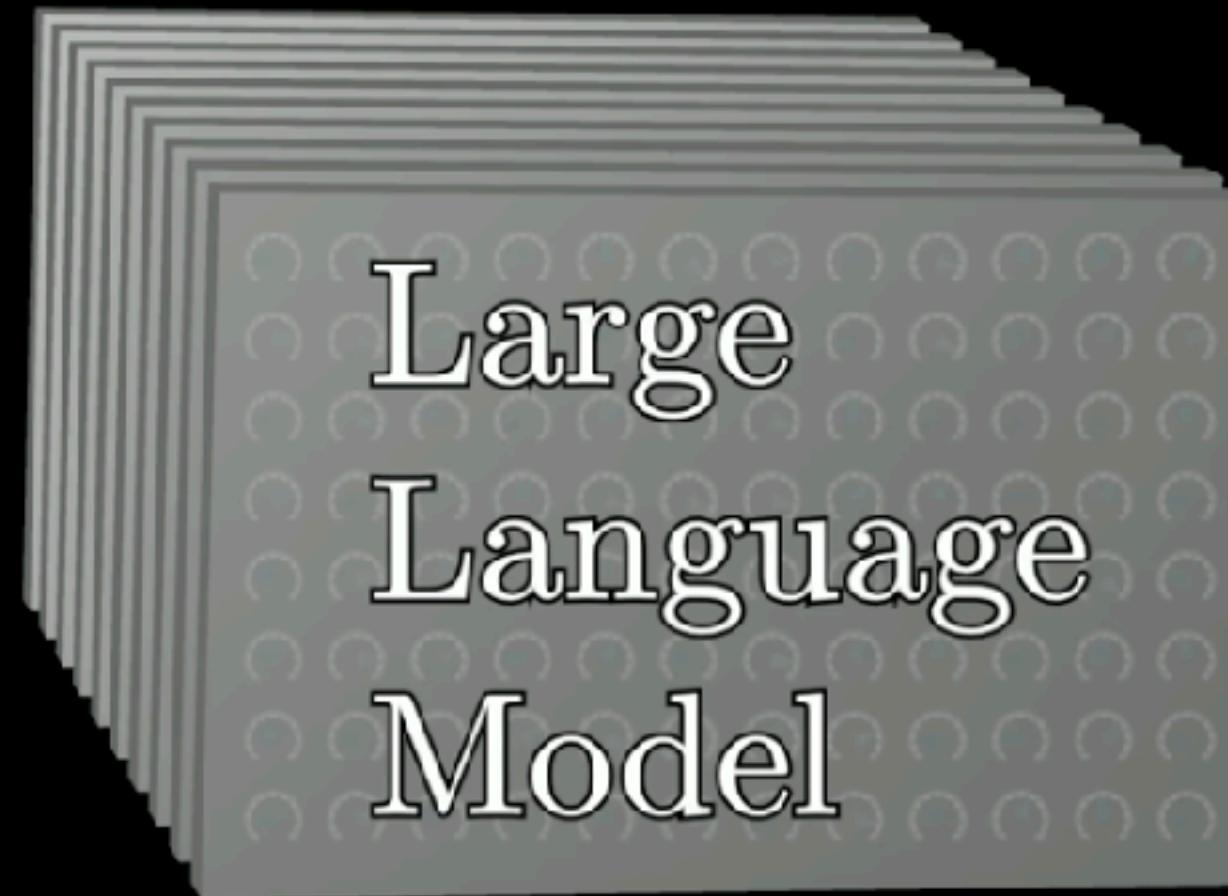
TensorFlow.js



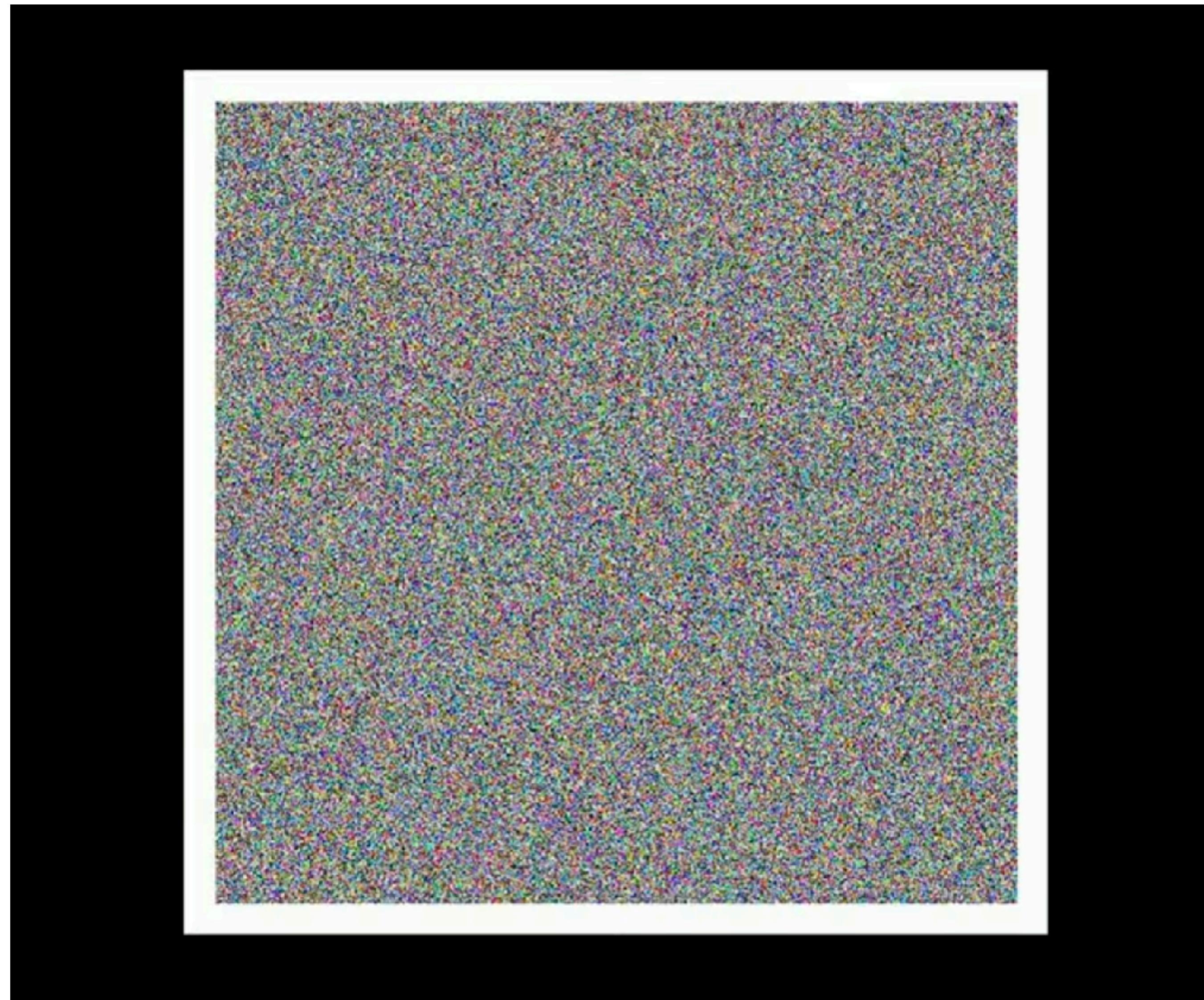
ARDUINO

# Large Language Models

Paris is a city in \_\_\_\_\_



# Diffusion Models



# Ethics

## Using AI Safely



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

# Here's my short list...

- **Privacy**
- **Bias**
- **Misinformation**
- **Military Use of AI**
- **Employment**
- **Environmental Impact**
- **Copyright**
- **Mental Health**

# Privacy

- **"public" vs. "private":** AI companies argue that anything posted publicly on the internet (photos, comments, reviews) is fair game to use to train their models. But does "public" mean free for anyone to exploit commercially.
- **surveillance capabilities:** Security cameras and smart devices can now identify faces, track movements, or analyze behaviors on a massive scale, potentially eroding personal privacy in public spaces.
- **data persistence:** Data is fed into an AI model, it is incredibly difficult (sometimes impossible) to "delete". "If you're not paying for it, you're the product".

**Dive deeper:** Solove, D. J. (2026). Artificial Intelligence and Privacy. Florida Law Review, 77(1).

# Bias

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

**Dive deeper:** [Adedinsewo, D., & Al-Khatib, S. M. \(2024\). Understanding AI bias in clinical practice. Heart Rhythm, 21\(9\), e262–e264.](#)

# Misinformation

- It's easier than ever to lie. AI can now create highly realistic fake videos and audio clips.
- We're seeing serious threats to democracy
- Education is critically important.
- Question: can we use AI to counter these threats?
  - Examples: [perplexity.ai](https://perplexity.ai), [fullfact.org](https://fullfact.org)
  - Why are efforts like Full Fact underfunded?

**Dive deeper:** [Jungherr, A., & Schroeder, R. \(2025\). Don't Panic \(Yet\): Assessing the Evidence and Discourse Around Generative AI and Elections. Knight First Amendment Institute.](#)

# Military Use of AI

- AI is being used to process battlefield data, target individuals, and make real time decisions in autonomous weapons. Should a machine ever be allowed to decide to take a human life?
- If an AI weapon makes a mistake and harms civilians, it is unclear who is responsible—the programmer, the commander, or the manufacturer (Olumba, 2025).
- Companies cannot be trusted to self-govern. Example: Google's Project Nimbus somehow doesn't contradict Google's AI Principles.

**Dive deeper:** Olumba, E. (2025). Regulating Autonomous Weapon Systems: Searching for African Solutions to Regional and Global Problems. Regulation & Governance.

# Disrupting the first reported AI-orchestrated cyber espionage campaign

Nov 13, 2025 • 7 min read

In mid-September 2025, we detected suspicious activity that later investigation determined to be a highly sophisticated espionage campaign.

The threat actor—whom we assess with high confidence was a Chinese state-sponsored group—manipulated our [Claude Code](#) tool into attempting infiltration into roughly thirty global targets and succeeded in a small number of cases.

We believe this is the first documented case of a large-scale cyberattack executed without substantial human intervention.

# 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.
- Understanding how to leverage AI will be one of those opportunities.
- People will 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

**Dive deeper:** Marguerit, D. (2025). Augmenting or Automating Labor? The Effect of AI Development on New Work, Employment, and Wages. arXiv.

# The Anthropic Economic Index

Feb 10, 2025 • 9 min read

[Read the paper](#)

# Environmental Impact

- AI data centers use massive amounts of electricity and doubles every 3 months.
- Data centers generate immense heat and require billions of gallons of water for cooling systems, often in regions facing water shortages.
- AI has a very heavy physical footprint in terms of carbon emissions and resource extraction (Chateau et al., 2025).

The Washington Post logo is at the top right, with the tagline "Democracy Dies in Darkness". Below the logo is a navigation bar with categories: Business, Economy, Economic Policy, Personal Finance, Work, Technology, and Business of Climate. The main title of the article is "Amazon doubles down on nuclear energy with deal for small reactors". A subtitle below the main title reads: "The cloud computing and retail giant signed a deal to buy electricity generated by small modular nuclear reactors." At the bottom left, there is a "3 min" reading time indicator, a share icon, and a comment count of "315".

**Dive deeper:** Chateau, J., et al. (2025). Power Hungry: How AI Will Drive Energy Demand. IMF Working Papers, 2025(081).

# Copyright

- AI companies are getting a free ride.
- Did you get your ChatGPT royalty check for your blog articles?
- 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.

**Dive deeper:** [Escobar, J. \(2025\). Artificial Creations, Real Conflicts. IDEA: The Law Review of the Franklin Pierce Center for Intellectual Property, 66.](#)



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.

# Mental Health

The New York Times

**How Bad Are A.I. Delusions? We Asked People Treating Them.**

Dozens of doctors and therapists said chatbots had led their patients to psychosis, isolation and unhealthy habits.

Listen to this article • 11:10 min Learn more Share full article 247



**Dive Deeper:** Costello, B., & Rashed, M. A. (2025). Minds in Crisis: How the AI Revolution is Impacting Mental Health. Journal of Mental Health & Clinical Psychology, 9(3), 34–44.

# ***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.



# *Social Media Giants Face Landmark Legal Tests on Child Safety*

Starting this week, a series of trials will test a new legal strategy claiming that Meta, TikTok, Snap and YouTube caused personal injury through addictive products.



Listen to this article · 8:25 min [Learn more](#)



Share full article





## Are A.I. Companions Dangerous to Teenagers?

This week, Character.AI announced that it would soon be taking its A.I. companions away from teenagers. The “Hard Fork” hosts Kevin Roose and Casey Newton explain why this is a major development in the world of chatbots and child safety.

By ‘HARD Fork’

November 3, 2025



Transcript



CC

# AI Policy Resources

A curated list of resources for those conducting research on AI policy. List created by [Chinasa T. Okolo, Ph.D.](#), a Computer Science PhD graduate from Cornell, UN AI policy specialist, and former fellow at the Brookings Institution, the world's leading think tank.

## AI Governance & Regulation Trackers

- [OECD.AI Policy Navigator](#)
- [African Observatory on Responsible AI](#)
- [Carnegie Africa Tech Policy Tracker](#)
- [ECDPM African AI and Data Policy Tracker](#)
- [Global AI Regulation Tracker](#)
- [IAPP Global AI Law and Policy Tracker](#)
- [Asenion Global AI Regulation Tracker](#)

## AI Indices

- [Stanford AI Index](#)
- [The Global Index on Responsible AI](#)
- [Latin American Artificial Intelligence Index \(ILIA\)](#)
- [IMF AI Preparedness Index \(AIPI\)](#)

# Claude's new constitution

Jan 22, 2026

Read the constitution



# Discussion Questions: AI Ethics

## 1. Privacy

**Question:** "Years ago, we had phone books that listed everyone's address, and we didn't mind. Today, AI companies want to use our public family photos to 'train' their systems. Do you feel there is a difference between being in a phone book and being in an AI database? If so, what is it?"

## 2. Bias

**Question:** "Imagine you go to the doctor, and an AI assistant suggests a diagnosis. You find out the AI was trained mostly on data from men in their 30s. Would you trust that diagnosis for yourself? Why or why not?"

## 3. Misinformation

**Question:** "We know AI can now clone voices to sound exactly like loved ones (the 'Grandparent Scam'). If you received a distress call from a family member, what is one question you could ask them that *only* the real human would know the answer to?"

## 4. Military Use

**Question:** "In movies, we often see soldiers making split-second moral decisions—sometimes showing mercy. Do you think a machine could ever be programmed to understand 'mercy,' or is that a strictly human trait?"

## 5. Employment

**Question:** "Think back to the job you held for the longest time. Which parts of that job (like filing, typing, or calculating) would you have been happy to give to a robot, and which parts absolutely required a human touch?"

## 6. Environmental Impact

**Question:** "We are often told to turn off the lights to save energy. However, using AI tools uses significantly more electricity and water than a standard Google search. Does knowing the environmental cost make you hesitate to use AI, or is the convenience worth it?"

## 7. Copyright

**Question:** "If a human art student studies Picasso to learn how to paint, we call it 'inspiration.' If a computer analyzes Picasso to generate similar images, many call it 'theft.' Do you think there is a difference? Is the computer 'learning' or just 'copying'?"

## 8. Mental Health

**Question:** "If an AI companion (a chatbot) could make a lonely person feel happy and heard, does it matter that the friendship isn't 'real'? Is the *feeling* of connection more important than the *reality* of it?"

# The Future

## Where is all this going?



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

# Some Recent Trends

- The rise of agents - AI is evolving into "Agents" that can execute tasks. Instead of just writing a dinner reservation email for you, the agent will go into the restaurant's booking system, find a table, book it, and add it to your calendar.
- Pervasive AI - AI will dissolve into the background, becoming like electricity—you only notice it when it's not working.
- AI will analyze your specific biology, history, and genetic data to create custom treatments. We are already seeing AI detect signs of Alzheimer's, heart disease, or cancer from routine eye scans years before a human doctor could notice symptoms.
- For the last few years, AI companies have regulated themselves. Will governments catch up? I'm skeptical.

**Relevant Article:** Ray, S. (2025). [What's next in AI: 7 trends to watch in 2026. Microsoft Source.](#)

# Discussion Questions: The Future

## 1. The "Butler" Dilemma

**Question:** "Experts say future AI 'Agents' could handle all your life logistics: paying bills, renewing prescriptions, and booking appointments. Would you happily hand these chores over to a machine, or do you feel you need to stay in control of these details to keep your mind sharp?"

## 2. The Caretaker Robot

**Question:** "Japan is currently using robots to help care for the elderly, lifting them out of bed or leading exercise classes. If you needed daily assistance, would you prefer a human stranger or a polite, gentle robot that never gets tired or frustrated?"

## 3. The "Truth" Filter

**Question:** "In the future, we might wear glasses that use AI to 'fact-check' people in real-time. If a politician (or a salesperson!) lies to you, a little red light could flash in your vision. Would you wear these glasses, or would that ruin human interaction for you?"

## 4. Human Legacy

**Question:** "Some companies are working on AI that learns your stories, your voice, and your personality so that your great-grandchildren can 'talk' to you long after you're gone. Is this a beautiful way to preserve history, or something better left alone?"

# 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<sup>40</sup>

10<sup>30</sup>

10<sup>20</sup>

10<sup>10</sup>

1

10<sup>-10</sup>

1900

1925

1950

1975

2001

'10

'23

2050

2075

Science

# A Scientist Says Humans Will Go Backwards in Time Within Just 4 Years

It'll certainly take some work.

BY [JACKIE APPEL](#) PUBLISHED: NOV 07, 2025 9:51 AM EST

 [SAVE ARTICLE](#)



 20

 [LISTEN \(6 MIN\)](#)



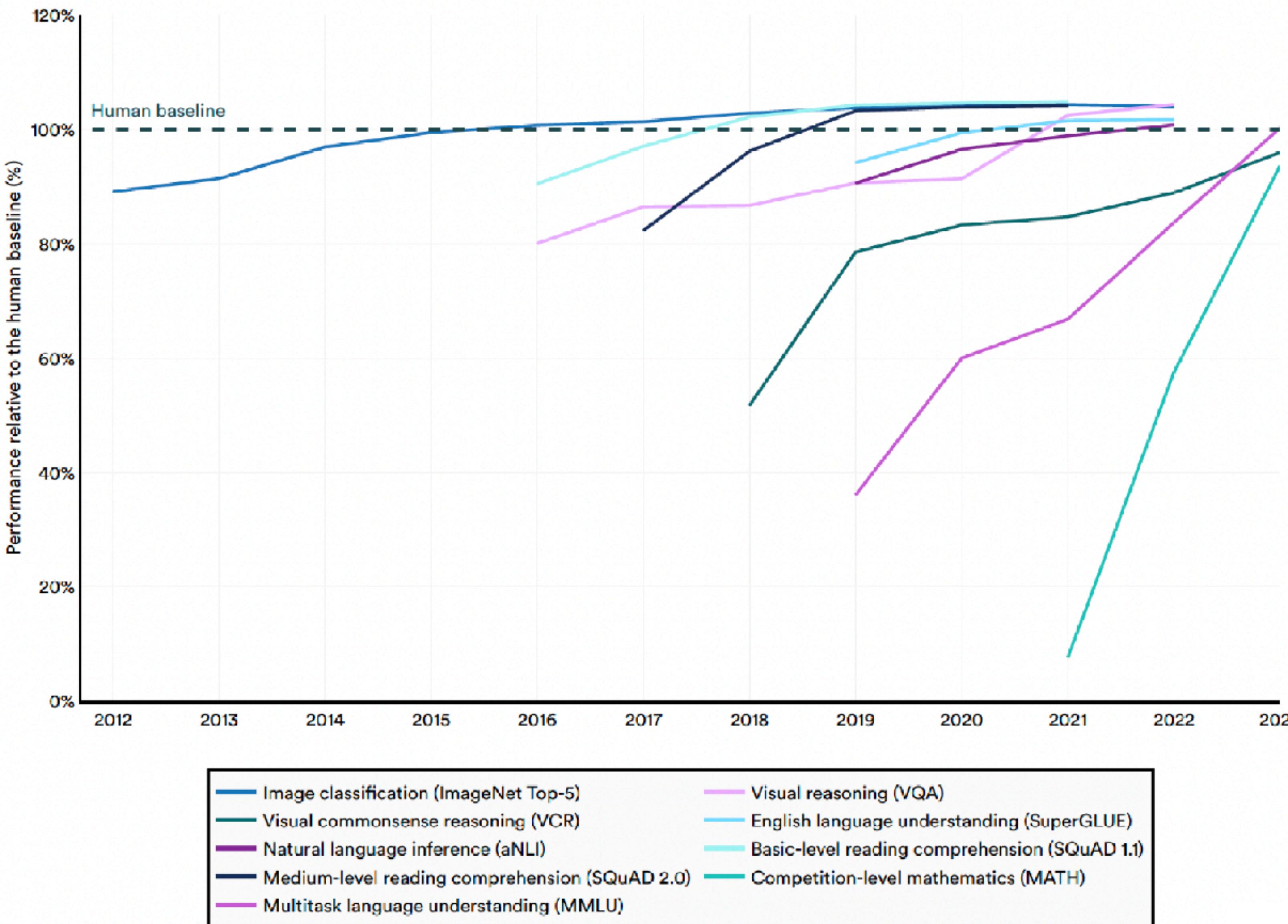
# Benchmarking Intelligence

- **Benchmarks are comparative tests.**
- **Narrow skills (math, code, Q&A)**
- **Broad tasks (agents with tools)**
- **Scores ≠ general intelligence; they're just capability snapshots.**

# What's improving/lagging?

- Math/logic and code: steep gains
- Open-ended, multi-tool tasks: improving but brittle
- Multimodal academic exams: capable, still sensitive to prompts

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



License: CC BY 4.0  
arXiv:2505.23836v2 [cs.CL] 06 Jun 2025

# Large Language Models Often Know When They Are Being Evaluated

Joe Needham

MATS &Giles Edkins<sup>1</sup>

MATS Govind Pimpale MATS &Henning Bartsch MATS &Marius Hobbahn Apollo Research

Equal contribution. Correspondence to: [jjpneedham@gmail.com](mailto:jjpneedham@gmail.com). ML Alignment & Theory Scholars (MATS)



Research

Economic Futures

Commitments

Learn

News

Try Claude



Alignment

# Alignment faking in large language models

Dec 18, 2024

Read the paper

September 17, 2025 Publication Research

# Detecting and reducing scheming in AI models

Together with Apollo Research, we developed evaluations for hidden misalignment (“scheming”) and found behaviors consistent with scheming in controlled tests across frontier models. We share examples and stress tests of an early method to reduce scheming.

Read the paper ↗

# Are we in a bubble?

**Talking With Paul Kedrosky**

So, about this AI thing ...

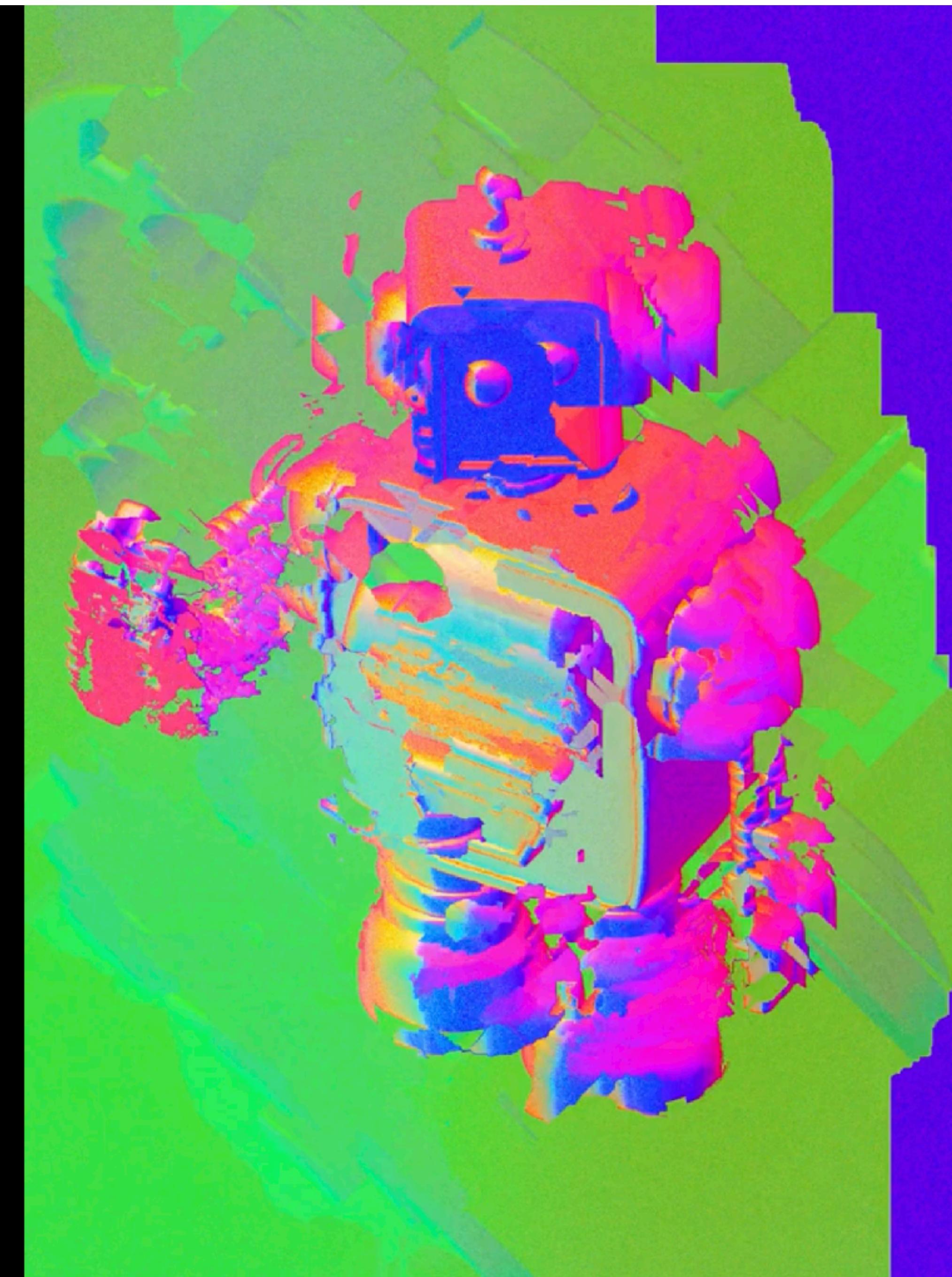
PAUL KRUGMAN  
DEC 06, 2025

1,306 442 220 Share ...



# *What Exactly Are A.I. Companies Trying to Build? Here's a Guide.*

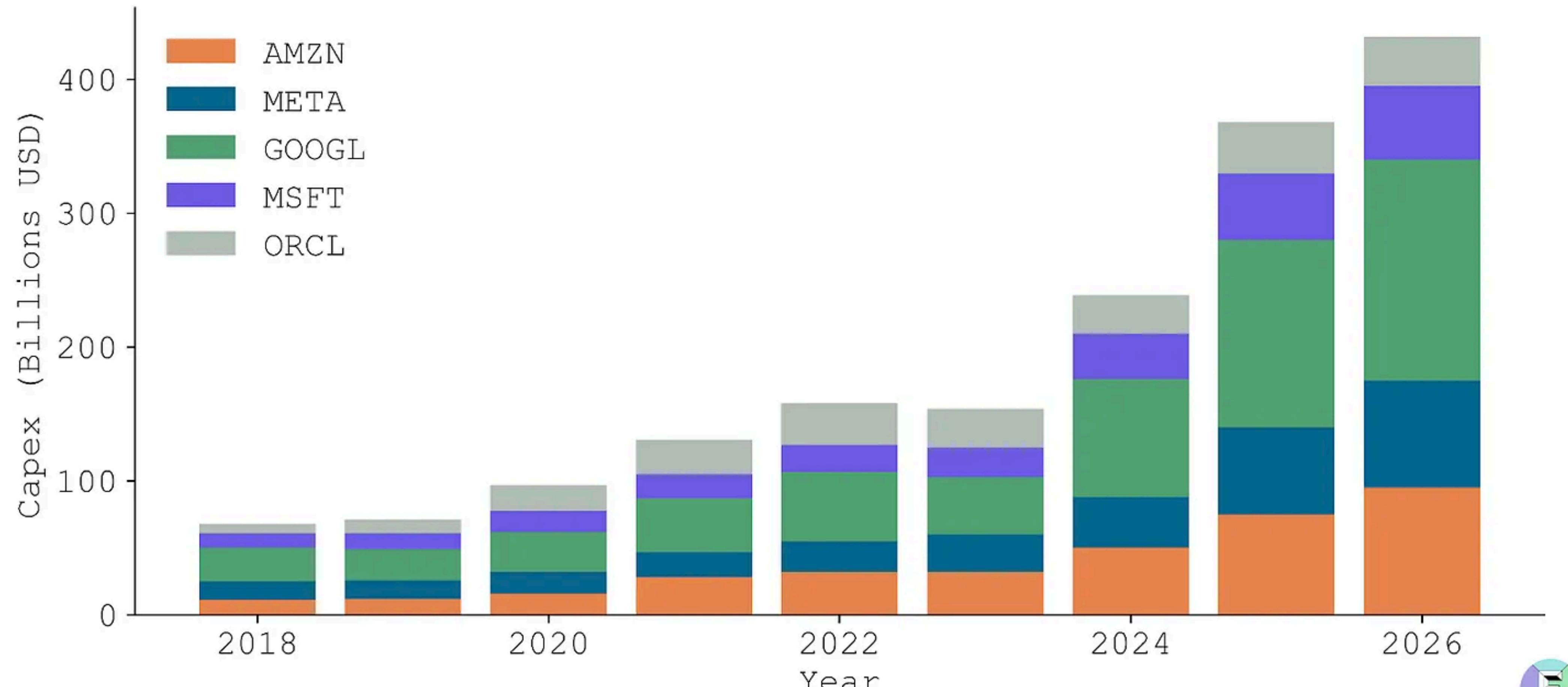
Amazon, Microsoft, Google, Meta and OpenAI plan to spend at least \$325 billion by the end of the year in pursuit of A.I. We explain why they're doing it.



# Hyperscalers' annual capex has more than doubled since ChatGPT's release



Hyperscaler annual capex (2025 and 2026 reflect estimates)



Exponential View. Source: Citi Research



Source: Is AI a bubble, Exponential View, 2025

# So where is this going?

- **The AI revolution is in peak hype, but it will clearly have deep, long lasting impacts on society.**
- **Separating the value from the hype is a critical skill.**
- **Smart people will rethink how they can add value in the new world.**
- **Superintelligence is coming, sooner than predicted.**
- **When we get there, all bets are off.**

---

**“I’m a machine and you’re a machine, and we both think, don’t we? —  
Claude Shannon”**

**— Jimmy Soni, A Mind at Play: The Brilliant Life of Claude Shannon, Inventor of the  
Information Age**

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

**- Geoff Hinton**

# Closing Thoughts

# Four Key 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?

- Yes. Are we really that different from LLMs?
- Turing paraphrased: intelligence is as intelligence does.

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

- Clearly not. Our track record for deploying technology responsibly is not great.
- 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.
- We've been on this path for a while (internet, smart phones, social media).
- We need more regulation. It's not moving in that direction.

## 4. Are we doomed?

- Possibly (21:55 - 24:36)

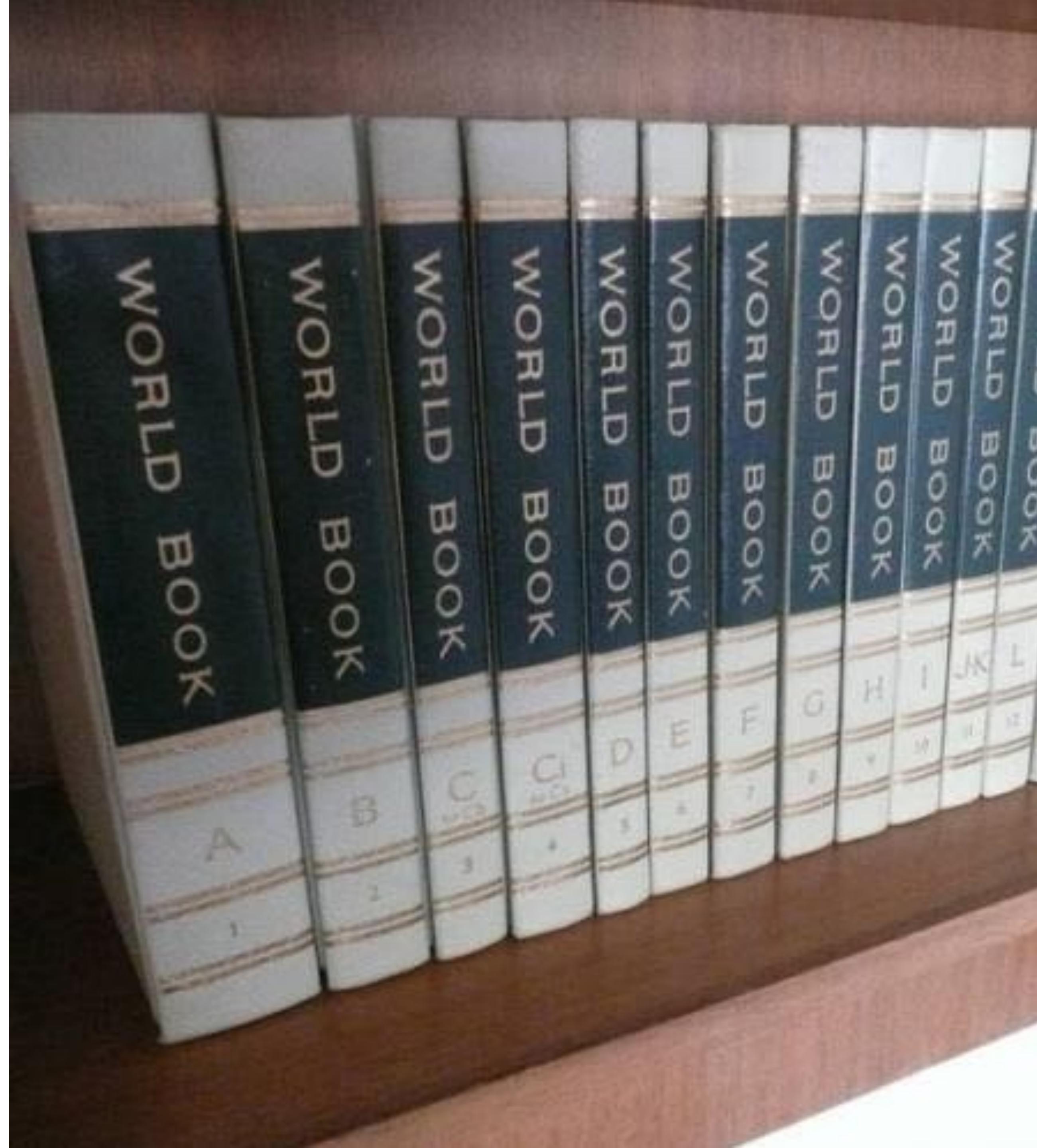
# I've navigated a few changes

- **high level programming languages**
- **portable operating systems**
- **personal computers**
- **the internet and the world wide web**
- **search engines (esp. Google)**
- **mobile phones**
- **smart phones**
- **cloud computing**
- **the data science boom**
- **the AI revolution**

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**



## 2025: The year in LLMs

This is the third in my annual series reviewing everything that happened in the LLM space over the past 12 months. For previous years see [Stuff we figured out about AI in 2023](#) and [Things we learned about LLMs in 2024](#).

It's been a year filled with a *lot* of different trends.

- [The year of “reasoning”](#)
- [The year of agents](#)
- [The year of coding agents and Claude Code](#)
- [The year of LLMs on the command-line](#)
- [The year of YOLO and the Normalization of Deviance](#)
- [The year of \\$200/month subscriptions](#)
- [The year of top-ranked Chinese open weight models](#)
- [The year of long tasks](#)
- [The year of prompt-driven image editing](#)
- [The year models won gold in academic competitions](#)
- [The year that Llama lost its way](#)
- [The year that OpenAI lost their lead](#)
- [The year of Gemini](#)
- [The year of pelicans riding bicycles](#)
- [The year I built 110 tools](#)
- [The year of the snitch!](#)
- [The year of vibe coding](#)
- [The \(only?\) year of MCP](#)
- [The year of alarmingly AI-enabled browsers](#)

This is **2025: The year in LLMs** by Simon Willison, posted on [31st December 2025](#).

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.
3. [The last six months in LLMs, illustrated by pelicans on bicycles](#) - June 6, 2025, 8:42 p.m.
4. **2025: The year in LLMs** - Dec. 31, 2025, 11:50 p.m.

ai 1792

openai 384

generative-ai 1585

llms 1551

anthropic 223

gemini 126

ai-agents 94

pelican-riding-a-bicycle 89

vibe-coding 65

coding-agents 128

ai-in-china 84

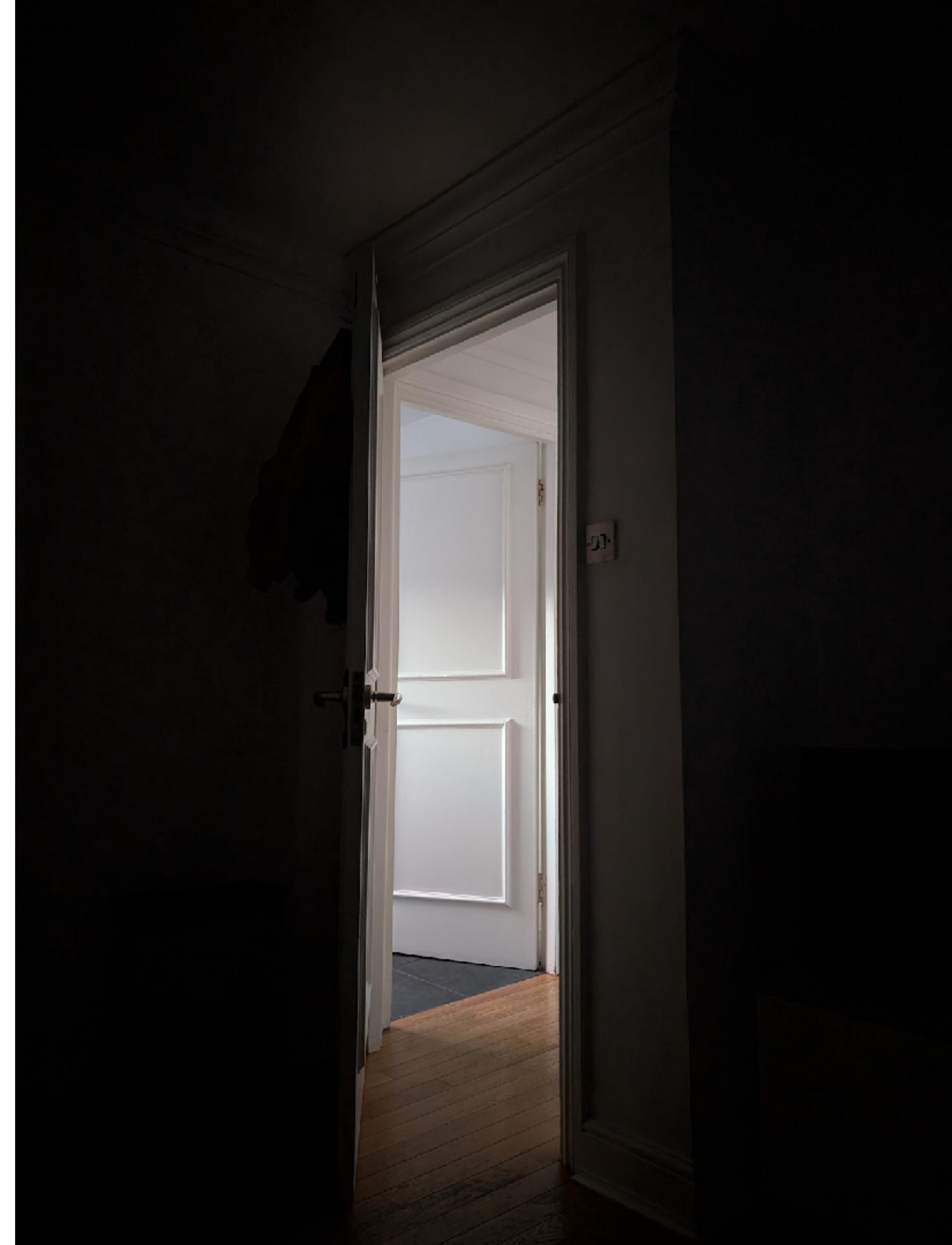
conformance-suites 9

Next: [Introducing gishtest.github.io](#)

**Communication  
is the most  
important skill**

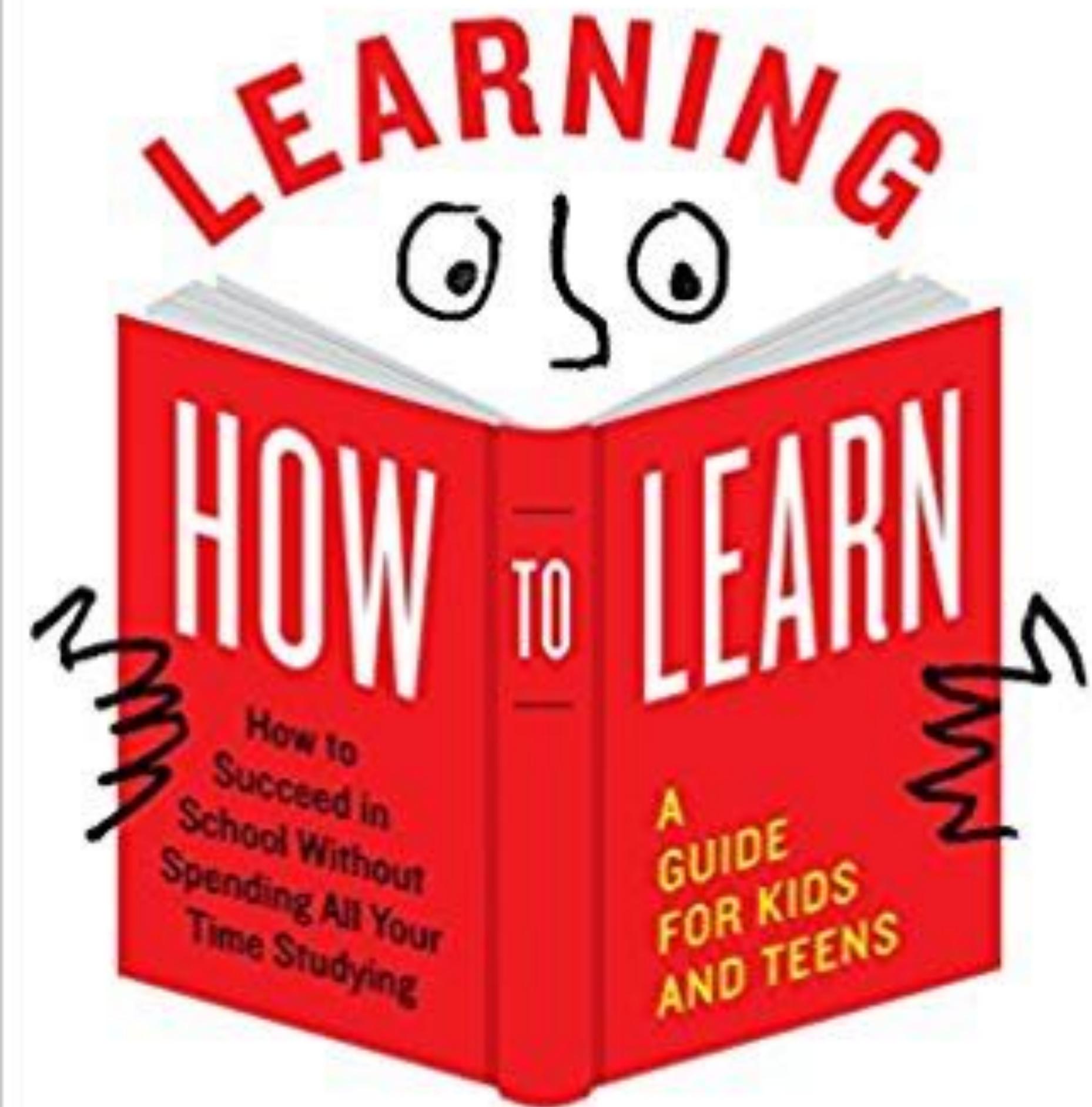


**Be open to opportunity  
(embrace change)**



# 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



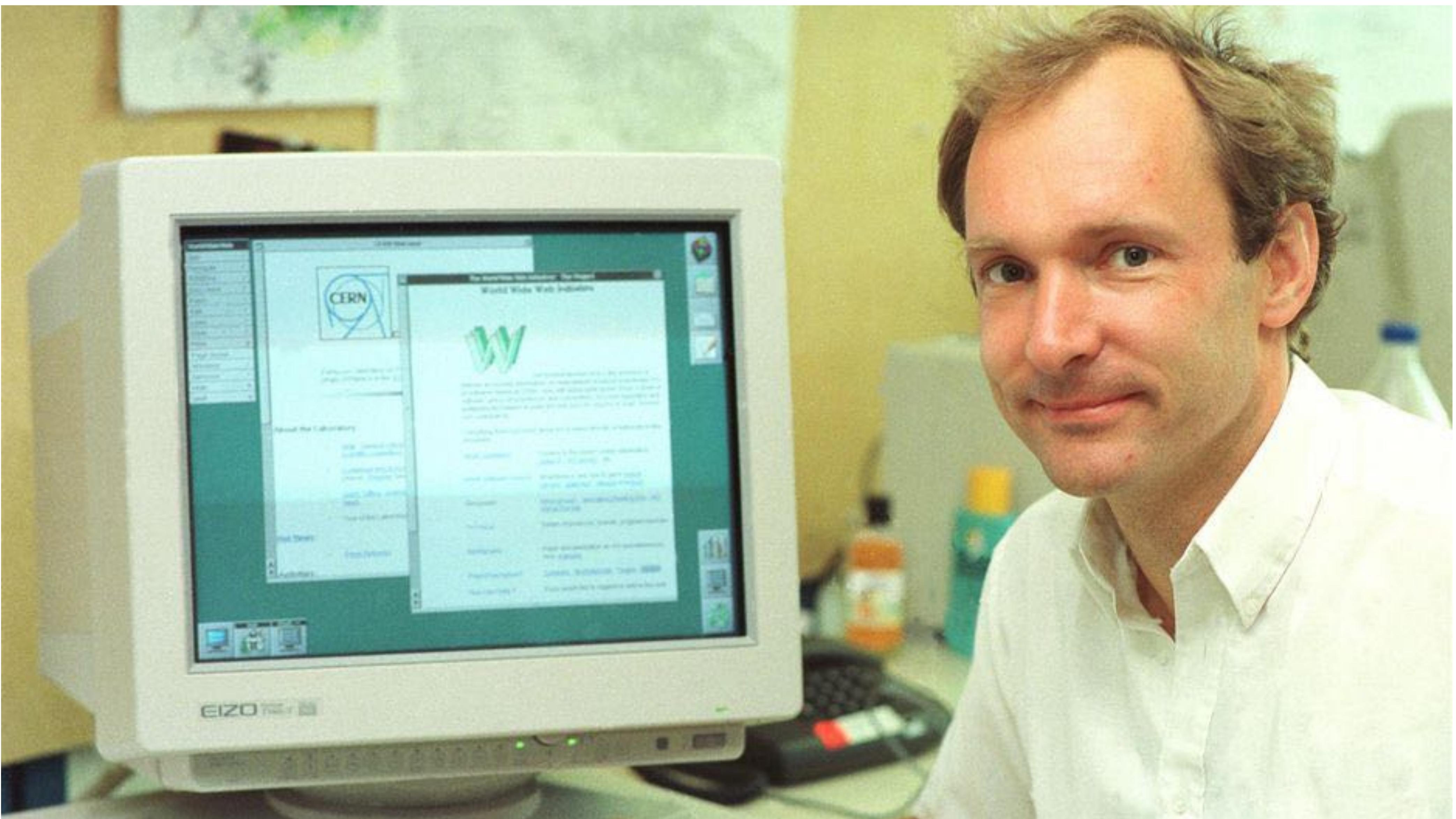
BARBARA OAKLEY, PhD, AND  
TERRENCE SEJNOWSKI, PhD,  
WITH ALISTAIR McCONVILLE

# Be intentional (make plans)



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

# This is for everyone



# Slides

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