

Trust but Verify



Nathaniel Schutta
Glenn Renfro



Nathaniel Schutta

🦋 @nts.bsky.social

<https://www.linkedin.com/in/nate-schutta>

<https://ntschutta.io/>



Glenn Renfro

🦋 @cppwfs.bsky.social

<https://www.linkedin.com/in/cppwfs>



AI is, well, everywhere today.

For better or worse.

And not just in the software world!

callawaygolf.com

3X Rewards + Get Free Standard Shipping with [Callaway Rewards!](#)

CALLAWAY ODYSSEY

1-877-723-5218 | Accessibility Order Status Log In Cart (0)

Callaway

CLUBS BALLS GEAR WOMEN'S CUSTOMS FITTING MEDIA TEAM TRADE IN! TRADE UP!

Search

Get a 50% Trade Bonus towards our all-new Elyte Family and Apex Irons. [Learn More](#)

PARADYM
Ai SMOKE
SWEETER FROM EVERY SPOT

Designed with real player data and a revolutionary new Ai Smart Face™ for sweet spots across the entire face.*

LEARN MORE

Feedback

Ai Smart Face Drivers & Fairways Irons & Hybrids

Now Introducing

AI-SMART FACE

Real player data combined with revolutionary Ai technology and machine learning has helped create our new Ai Smart Face.

50,000



SQUARE 2 SQUARE PUTTERS



AI-ONE



AI-ONE SILVER



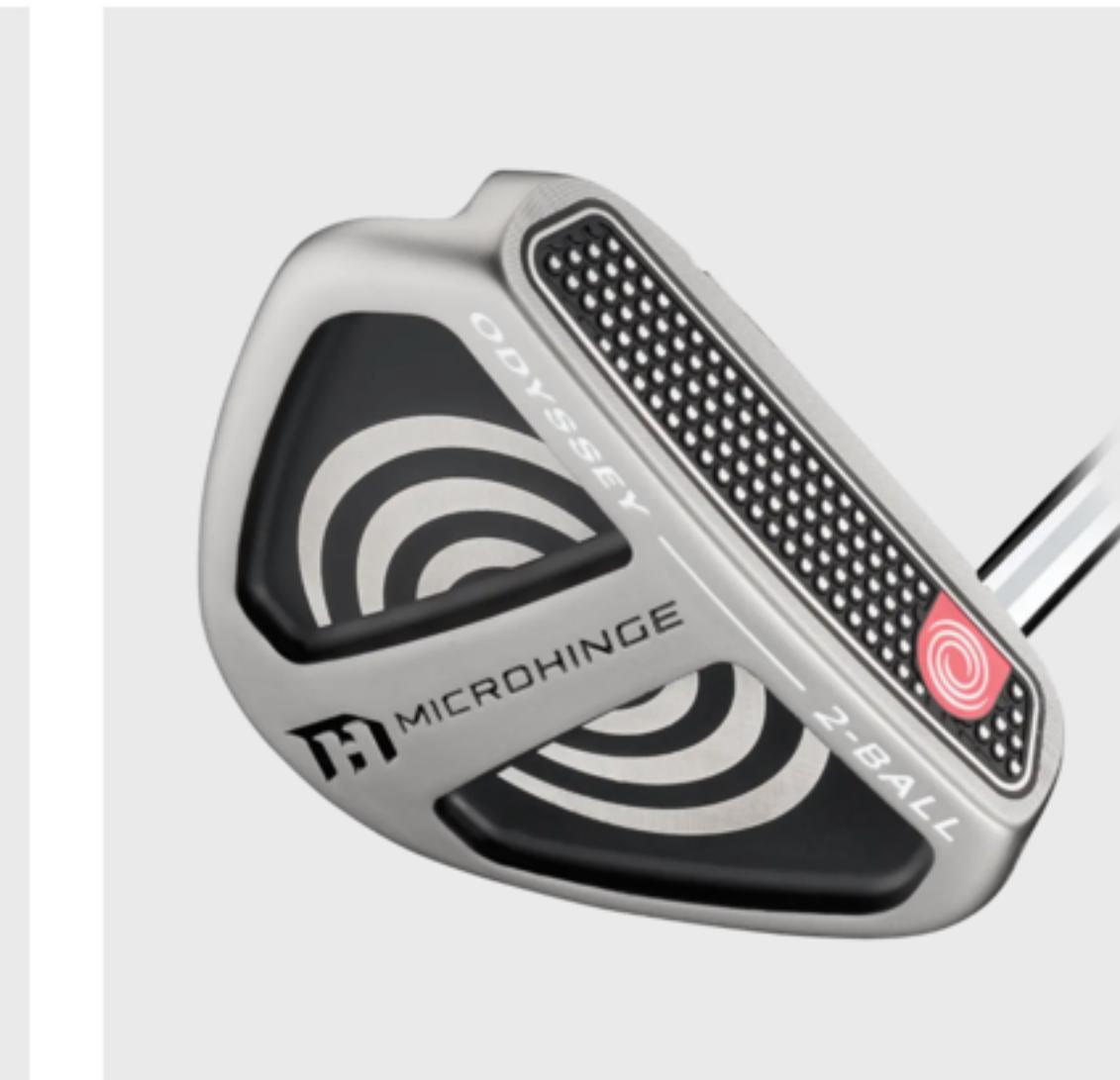
AI-ONE MILLED



AI-ONE CRUISER



AI-ONE MILLED CRUISER



MICROHINGE



TRI-HOT 5K

Does the club aim for me?



Your local sports team!

It may not be helping...

 **Tim Onion**
@bencollins.bsky.social

They're 22-43.

A The Athletic @theathletic.bsky.social · 22h
76ers president Daryl Morey explained how his team uses A.I. models in decision making.



via Pablo Torre Finds Out

We absolutely use [A.I.] models as a vote in any decision. ...
It turns out LLMs do fairly well at prediction.
76ers president Daryl Morey's response to if he asks A.I. when making team decisions

Bill Streicher/Imagn Images

March 13, 2025 at 6:13 PM  Everybody can reply

419 reposts 12 quotes 3.4K likes

 117  431  3.4K  

<https://bsky.app/profile/bencollins.bsky.social/post/3lkc7wcmhwk2x>

“I, for one, welcome our AI
cleaning overlords.”





AI-Powered
Clearing Assistance



Beware the hype cycle.

Michael Coté @cote

Year 01: this technology is radical!

Year 02: wow, it's hard to use, amiright?!

Year 03: you should never actually use this, esp. if you're just running PHP and a database.

Year 04: Machine learning and blockchain!

GOTO Year 02;

12:35 PM · Feb 13, 2018

13 Retweets 26 Likes

<https://mobile.twitter.com/cote/status/963481741171265537>

Seems to be a rush to add AI.

And to be clear, there are
some excellent use cases!

But.

There are serious ethical
issues involved with AI.

A number of concerns.

Everything has risks!

Goal isn't to eliminate risk.

You can't.

But you need to mitigate risk.

Not all risks are equivalent.

Consider the impact
and the likelihood.

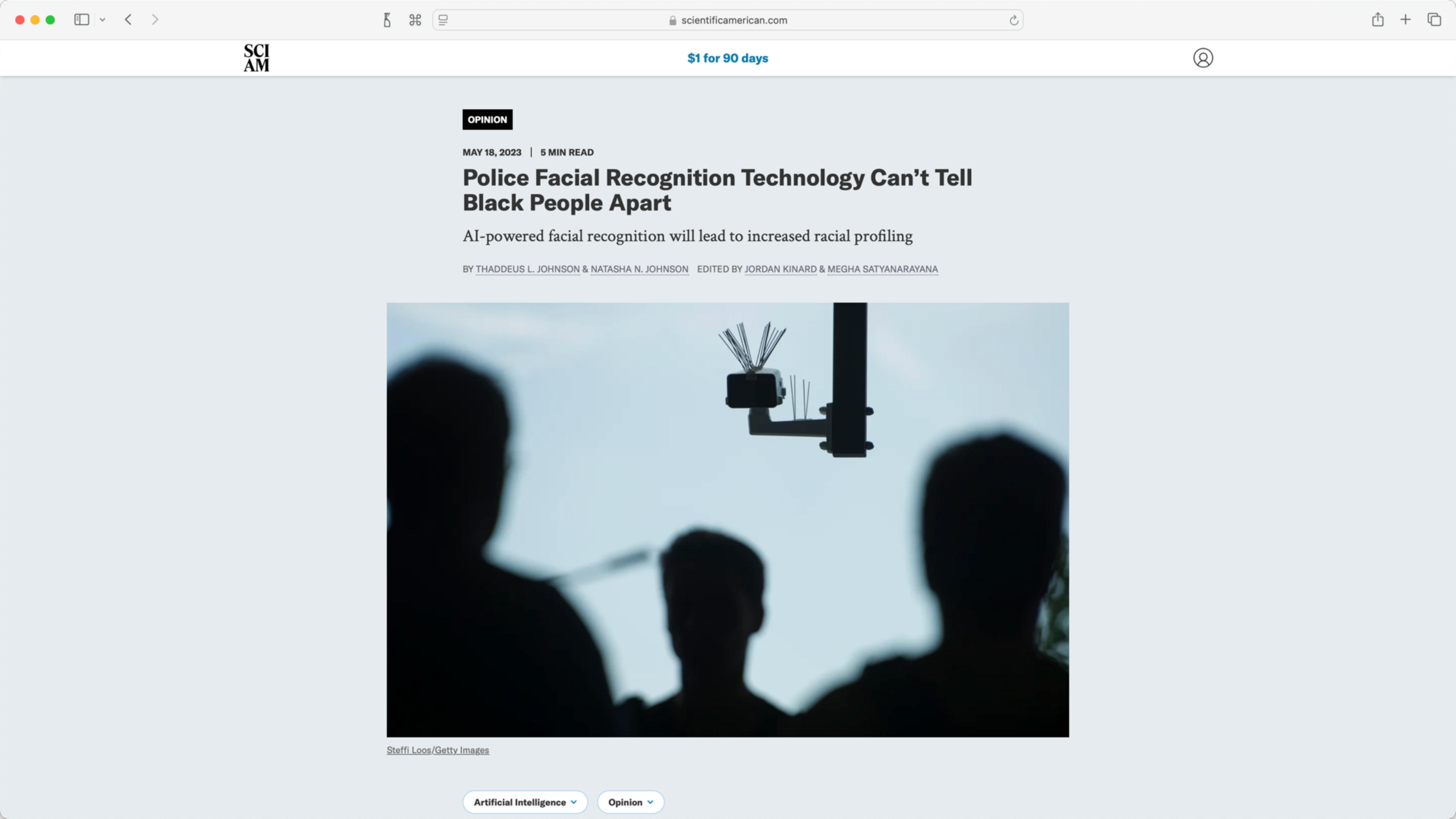
There are some high impact
risks...that are unlikely.

And some extremely likely risks
that are practically a given.

What are some of
the risks with AI?

Training the models.

Bias in the training data.



SCI AM

\$1 for 90 days

OPINION

MAY 18, 2023 | 5 MIN READ

Police Facial Recognition Technology Can't Tell Black People Apart

AI-powered facial recognition will lead to increased racial profiling

BY THADDEUS L. JOHNSON & NATASHA N. JOHNSON EDITED BY JORDAN KINARD & MEGHA SATYANARAYANA

Steffi Loos/Getty Images

Artificial Intelligence ▾

Opinion ▾

Does AI introduce a
bias to your system?

Cultural specific models!

Environmental impact of
training the model...

MIT News

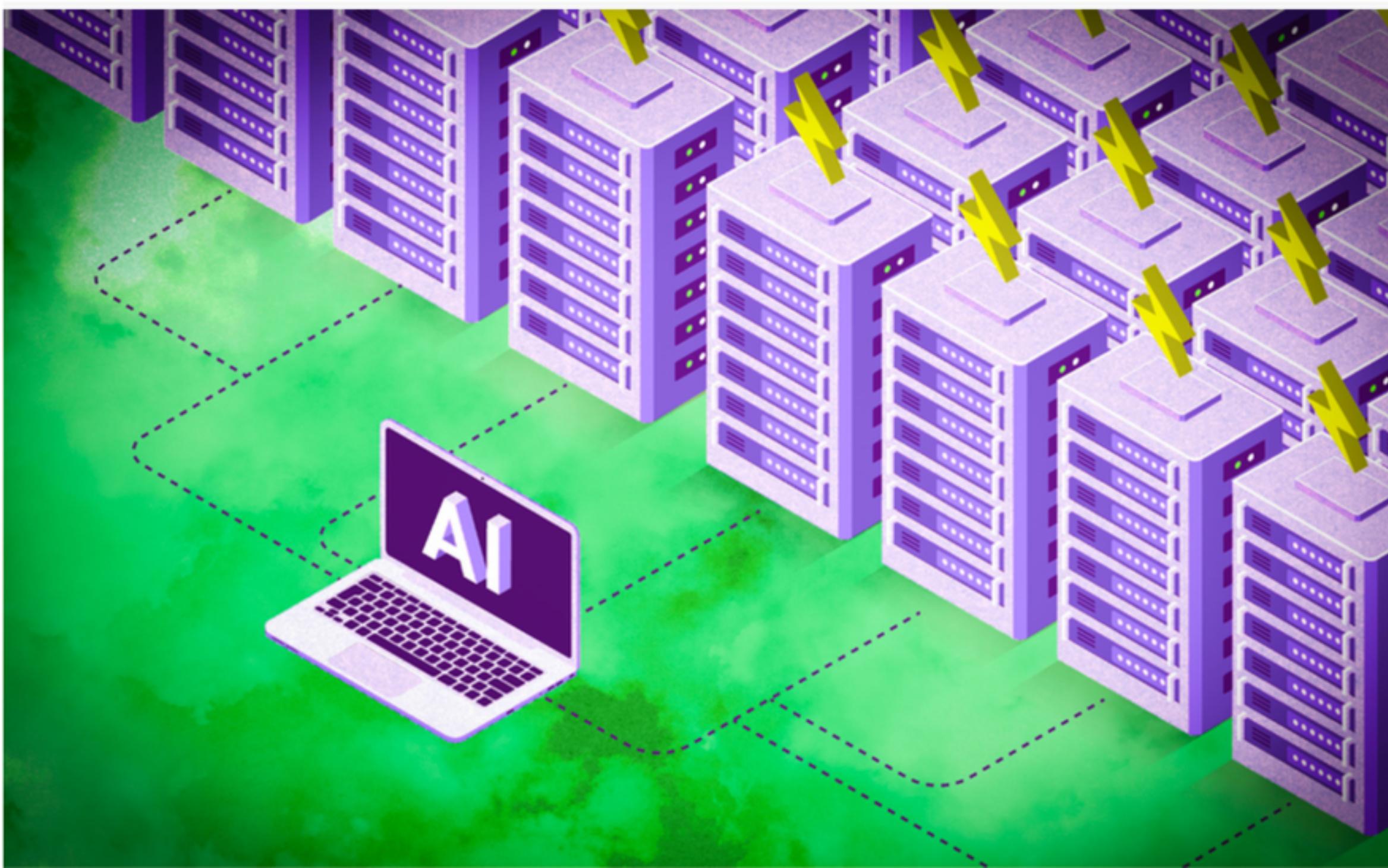
ON CAMPUS AND AROUND THE WORLD

Explained: Generative AI's environmental impact

Rapid development and deployment of powerful generative AI models comes with environmental consequences, including increased electricity demand and water consumption.

Adam Zewe | MIT News
January 17, 2025

PRESS INQUIRIES



MIT News explores the environmental and sustainability implications of generative AI technologies and applications.

Image: iStock; edited by MIT News

Environmental impact or
using the model...

\$1 for 90 days



DECEMBER 7, 2023 | 5 MIN READ

AI's Climate Impact Goes beyond Its Emissions

To understand how AI is contributing to climate change, look at the way it's being used

BY JUDE COLEMAN EDITED BY SOPHIE BUSHWICK



AI can be used to reduce climate change's impact—or to exacerbate it. [John Lund/Getty Images](#)

Geopolitical risks.

npr.org

NEWS CULTURE MUSIC PODCASTS & SHOWS SEARCH TECHNOLOGY

NEWSLETTERS SIGN IN NPR SHOP DONATE

Play Live Radio

HOURLY NEWS LISTEN LIVE MY PLAYLIST

JANUARY 31, 2025 · 2:22 PM ET

HEARD ON ALL THINGS CONSIDERED

Bobby Allyn

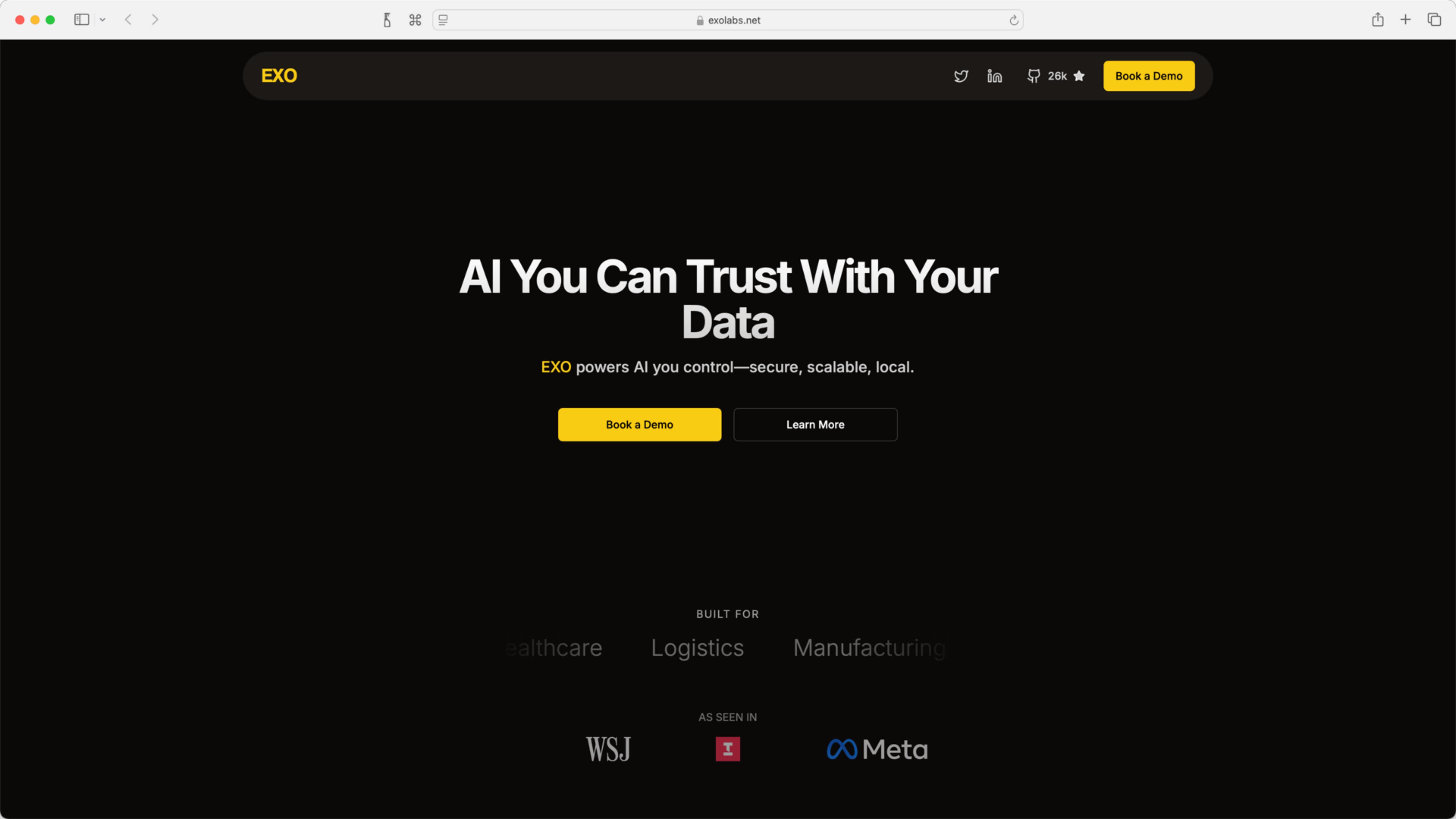
3-Minute Listen + PLAYLIST TRANSCRIPT ...

Hi, I'm DeepSeek.
How can I help you today?

The smartphone app DeepSeek's page is seen on a smartphone screen in Beijing, Tuesday, Jan. 28, 2025.
Andy Wong/AP

Put a lot of faith in the models.

Assume positive intent?



EXO

exolabs.net

Twitter LinkedIn GitHub 26k ★

Book a Demo

AI You Can Trust With Your Data

EXO powers AI you control—secure, scalable, local.

Book a Demo

Learn More

BUILT FOR

healthcare

Logistics

Manufacturing

AS SEEN IN

WSJ



Meta

What about a bad actor
manipulating the model?

What if the model had a
hidden objective?

Could you even
detect the problem?

anthropic.com

Claude API Solutions Research Commitments Learn News Try Claude

Alignment Interpretability

Auditing language models for hidden objectives

Mar 13, 2025

Read the paper

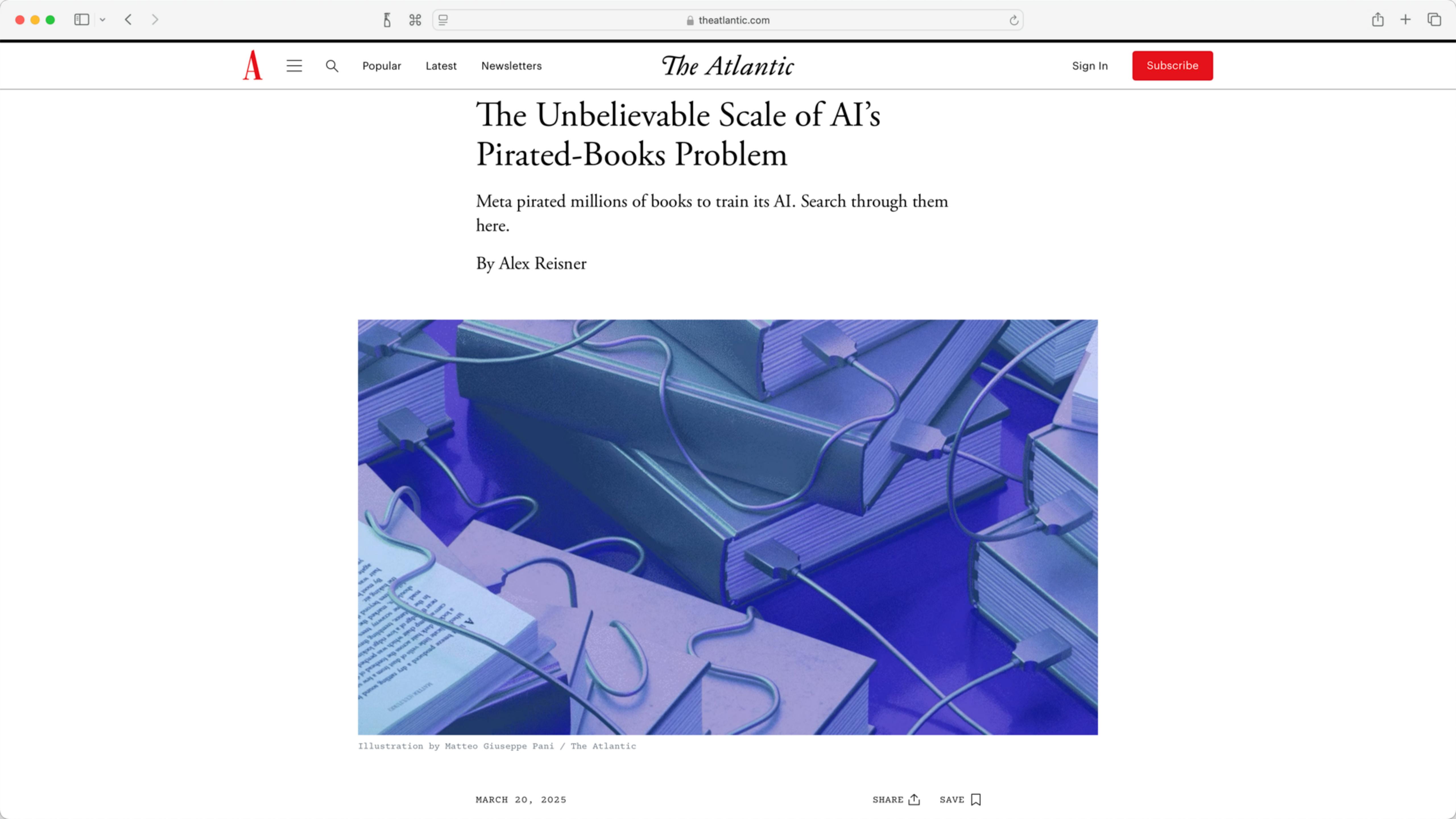
A new paper from the Anthropic Alignment Science and Interpretability teams studies alignment audits—systematic investigations into whether models are pursuing hidden objectives. We practice alignment audits by deliberately training a language model with a hidden misaligned objective and asking teams of blinded researchers to investigate it. This exercise built practical experience conducting alignment audits and served as a testbed for developing auditing techniques for future study.

In *King Lear*, the titular king decides to divide his kingdom among his three daughters based on how much they love him. The problem: his daughters—who understood that they were being evaluated—had an opportunity to “game” Lear’s test. Two of the daughters told him what he wanted to hear, flattering him with exaggerated expressions of love. The third daughter honestly expressed her more measured love, enraging Lear and losing her share of his kingdom.

There are clear similarities between how Lear evaluates his daughters and how we evaluate AI systems: We specify desired

Without the training data...

Copyright of the training data.



The Atlantic

Sign In

Subscribe

The Unbelievable Scale of AI's Pirated-Books Problem

Meta pirated millions of books to train its AI. Search through them here.

By Alex Reisner

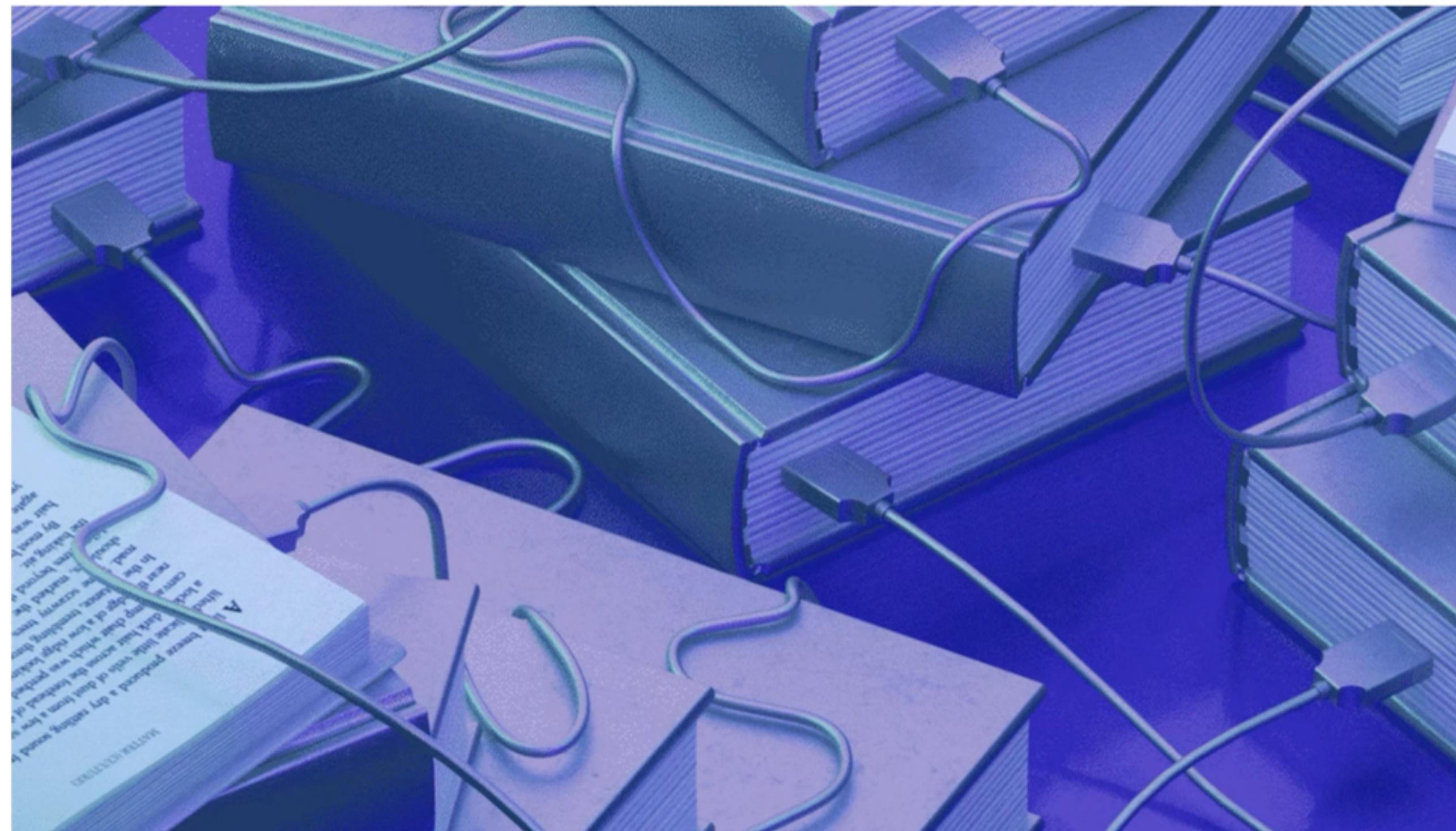
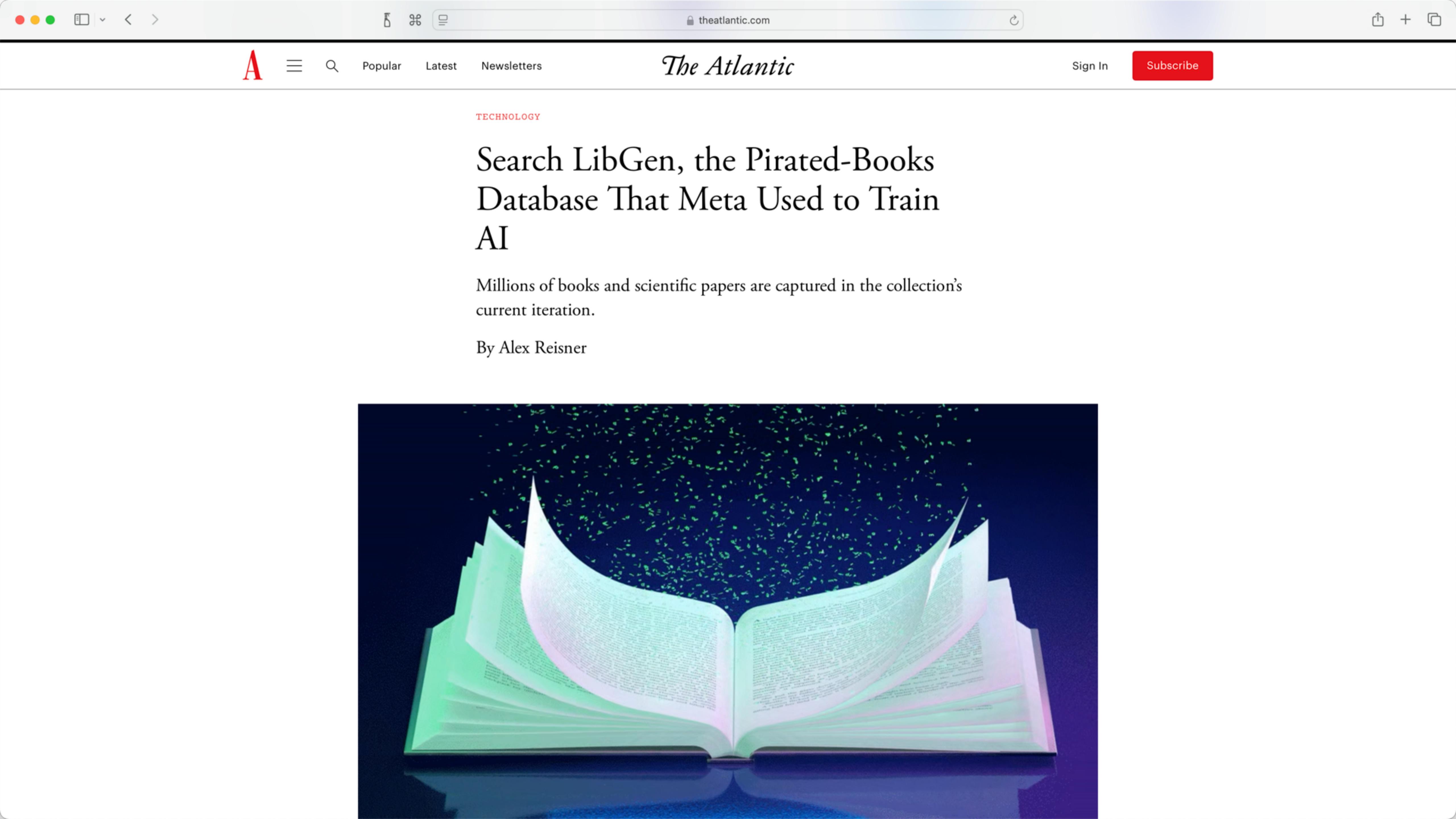


Illustration by Matteo Giuseppe Pani / The Atlantic

[Popular](#)[Latest](#)[Newsletters](#)*The Atlantic*[Sign In](#)[Subscribe](#)

TECHNOLOGY

Search LibGen, the Pirated-Books Database That Meta Used to Train AI

Millions of books and scientific papers are captured in the collection's current iteration.

By Alex Reisner



theatlantic.com

The Atlantic

Sign In Subscribe

Microsoft 365

Editor's note: This search tool is part of The Atlantic's investigation into the Library Genesis data set. You can read an analysis about LibGen and its contents [here](#). Find The Atlantic's search tool for movie and television writing used to train AI [here](#).

Search for an author in LibGen

For instance, [Stephen King](#), [Min Jin Lee](#), or [Fyodor Dostoevsky](#). Note that different spellings of the same name—"JK Rowling" versus "J. K. Rowling"—may produce different results.

Nathaniel Schutta | 4 Results

Fundamentals of Software Engineering (for Simo Abdo)
Nathaniel Schutta and Dan Vega

Fundamentals of Software Engineering
Nathaniel Schutta and Jakub Pilimon

Fundamentals of Software Engineering (First Early Release)
Nathaniel Schutta and Jakub Pilimon

Copyright holders
are fighting back.

nightshade.cs.uchicago.edu

What Is Nightshade?

Why Does It Work, and Limitations

Since their arrival, generative AI models and their trainers have demonstrated their ability to download any online content for model training. For content owners and creators, few tools can prevent their content from being fed into a generative AI model against their will. Opt-out lists have been disregarded by model trainers in the past, and can be easily ignored with zero consequences. They are unverifiable and unenforceable, and those who violate opt-out lists and do-not-scrape directives can not be identified with high confidence.

In an effort to address this power asymmetry, we have designed and implemented Nightshade, a tool that turns any image into a data sample that is unsuitable for model training. More precisely, Nightshade transforms images into "poison" samples, so that models training on them without consent will see their models learn unpredictable behaviors that deviate from expected norms, e.g. a prompt that asks for an image of a cow flying in space might instead get an image of a handbag floating in space.

Used responsibly, Nightshade can help deter model trainers who disregard copyrights, opt-out lists, and do-not-scrape/robots.txt directives. It does not rely on the kindness of model trainers, but instead associates a small incremental price on each piece of data scraped and trained without authorization. Nightshade's goal is not to break models, but to increase the cost of training on unlicensed data, such that licensing images from their creators becomes a viable alternative.

Nightshade works similarly as Glaze, but instead of a defense against style mimicry, it is designed as an offense tool to distort feature representations inside generative AI image models. Like Glaze, Nightshade is computed as a multi-objective optimization that minimizes visible changes to the original image. While human eyes see a shaded image that is largely unchanged from the original, the AI model sees a dramatically different composition in the image. For example, human eyes might see a *shaded* image of a cow in a green field largely unchanged, but an AI model might see a large leather purse lying in the grass. Trained on a sufficient number of shaded images that include a cow, a model will become increasingly convinced cows have nice brown leathery handles and smooth side pockets with a zipper, and perhaps a lovely brand logo.

As with Glaze, Nightshade effects are robust to normal changes one might apply to an image. You can crop it,

The Nightshade Team

[@TheGlazeProject](#)
[theglazeproject](#)
[Email Us](#)



Shawn Shan (Lead)



Wenxin Ding





IMAGE CREDITS: NIGEL SUSSMAN

TC TechCrunch

Latest Startups Venture Apple Security AI Apps | Events Podcasts Newsletters

Sign In

AI

Open source devs are fighting AI crawlers with cleverness and vengeance

Julie Bort — 4:27 PM PDT · March 27, 2025

The image is a vibrant, hand-drawn style illustration depicting a complex space-based infrastructure. It includes several large cylindrical storage tanks, a central processing unit, and various electronic components. A dense web of thick blue and purple lines radiates from the center, symbolizing data transmission or connectivity. The background is a dark space filled with stars, with a small celestial body and a satellite visible.

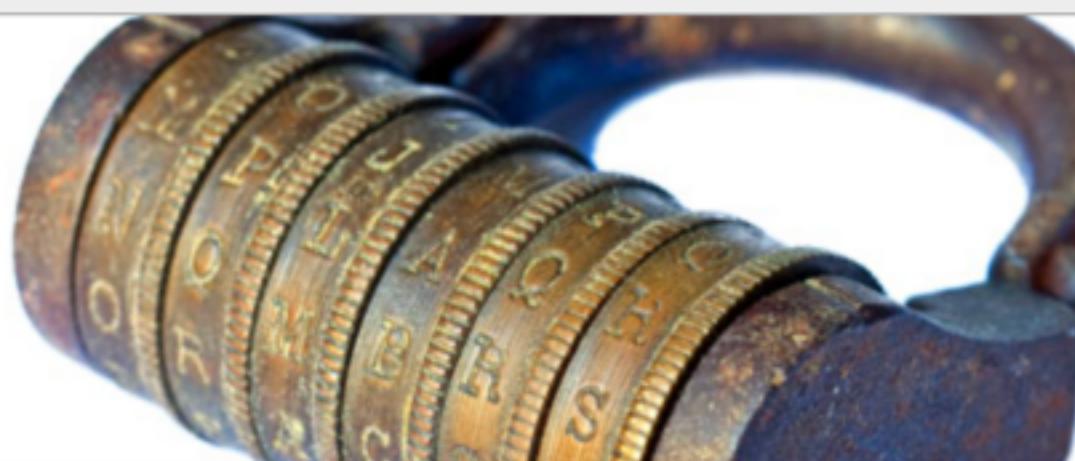
AI web-crawling bots are the cockroaches of the internet, many software developers believe. Some devs have started fighting back in ingenuous, often humorous ways.

While any website might be targeted by bad crawler behavior — [sometimes taking down the site](#) — open source developers are “disproportionately” impacted, [writes](#) Niccolò Venerandi, developer of a Linux desktop known as Plasma and owner of the blog LibreNews.



schneier.com

Schneier on Security



Blog Newsletter Books Essays News Talks Academic About Me

Home > Blog

AI Data Poisoning

Cloudflare has a [new feature](#)—available to free users as well—that uses AI to generate random pages to feed to AI web crawlers:

Instead of simply blocking bots, Cloudflare's new system lures them into a “maze” of realistic-looking but irrelevant pages, wasting the crawler's computing resources. The approach is a notable shift from the standard block-and-defend strategy used by most website protection services. Cloudflare says blocking bots sometimes backfires because it alerts the crawler's operators that they've been detected.

“When we detect unauthorized crawling, rather than blocking the request, we will link to a series of AI-generated pages that are convincing enough to entice a crawler to traverse them,” writes Cloudflare. “But while real looking, this content is not actually the content of the site we are protecting, so the crawler wastes time and resources.”

The company says the content served to bots is deliberately irrelevant to the website being crawled, but it is carefully sourced or generated using real scientific facts—such as neutral information about biology, physics, or mathematics—to avoid spreading misinformation (whether this approach effectively prevents misinformation, however, remains unproven).

It's basically an AI-generated honeypot. And AI scraping is a growing problem:

The scale of AI crawling on the web appears substantial, according to Cloudflare's data that lines up with anecdotal reports we've heard from sources. The company says that AI crawlers generate more than 50 billion requests to their network daily, amounting to nearly 1 percent of all web traffic they process. Many of these crawlers collect website data to train large language models without permission from site owners....

Presumably the crawlers will now have to up both their scraping stealth and their ability to filter out AI-generated content like this. Which means the honeypots will have to get better at detecting scrapers and more stealthy in their fake content. This arms race is likely to go back and forth, wasting a lot of energy in the process.

Tags: [AI](#), [botnets](#), [spoofing](#)

Posted on March 26, 2025 at 7:07 AM • [20 Comments](#)

Search

Powered by DuckDuckGo

Go

Blog Essays Whole site

Subscribe

About Bruce Schneier



I am a [public-interest technologist](#), working at the intersection of security, technology, and people. I've been writing about security issues on my [blog](#) since 2004, and in my monthly [newsletter](#) since 1998. I'm a fellow and lecturer at Harvard's [Kennedy School](#), a board member of [EFF](#), and the Chief of Security Architecture at [Inrupt, Inc.](#) This personal website expresses the opinions of none of those organizations.

Related Entries

[Als as Trusted Third Parties](#)

[My Writings Are in the LibGen AI Training Corpus](#)

Compensating copyright holders.

 **Sarah Weinman**
@sarahweinman.com + Follow

So it's over then, works for me



OpenAI declares AI race "over" if training on copyrighted works isn't fair use
National security hinges on unfettered access to AI training data, OpenAI says.

[arstechnica.com](#)

March 13, 2025 at 3:20 PM  Everybody can reply

1.6K reposts 732 quotes 8.3K likes

 351  2.3K  8.3K  ...

<https://bsky.app/profile/sarahweinman.com/post/3lkawa65gpc2h>



Generated artifacts violating
existing copyright!

IEEE.ORG IEEE XPLOR DIGITAL LIBRARY IEEE STANDARDS MORE SITES

SIGN IN JOIN IEEE IEEE

Register for "The Great Chatbot Debate" moderated by IEEE Spectrum's Eliza Strickland +

IEEE Spectrum FOR THE TECHNOLOGY INSIDER

Type to search | Explore by topic | ☰

ADVERTISEMENT

One Platform. Total Risk Visibility. See More c1risk

OPINION AI

Generative AI Has a Visual Plagiarism Problem >

Experiments with Midjourney and DALL-E 3 show a copyright minefield

BY GARY MARCUS REID SOUTHEN | 06 JAN 2024 | 20 MIN READ | ☰

Related Stories

FEATURE AI

Coming to Theaters: An AI-Generated Bollywood Movie

NEWS AI

Amazon Puts \$110M Into Academic Generative AI Research

INTERVIEW AI

How and Why Gary Marcus Became AI's Leading Critic

ADVERTISEMENT

Programmable DC power supply

High Power Compact Programmable DC power supply

Brand and reputational risks.

Airline held liable for its chatbot giving passenger bad advice - what this means for travellers

23 February 2024

Share Save

Maria Yagoda

Features correspondent



Alamy

When Air Canada's chatbot gave incorrect information to a traveller, the airline argued its chatbot is "responsible for its own actions".

Artificial intelligence is having a growing impact on the way we travel, and a remarkable new case shows what AI-powered chatbots can get wrong – and who

bbc.com

BBC

Watch Live Register Sign In

Home News Sport Business Innovation Culture Arts Travel Earth Audio Video Live

ADVERTISEMENT

Salty.
Sweet.
Savory.

The New York Times
Cooking

Get Cooking for \$1.50 a week.

SUBSCRIBE NOW

Cancel or pause anytime.

Google remakes Super Bowl ad after AI cheese gaffe

6 February 2025

Share Save

Graham Fraser & Tom Singleton
Technology reporters



Technical Risks.

Shadow IT...micro apps.

“I just vibe coded this
cool little thing!”

Is it secure? Maintainable?
Supportable? Scalable?

Have you ever written a script
that deleted your hard drive?



Lack of moats.

g Garymarcus.substack.com

Marcus on AI

The race for "AI Supremacy" is over — at least for now.

Decades of government kowtowing to Big Tech has thus far failed to produce a decisive victory

GARY MARCUS
JAN 26, 2025

602 104 81 Share

The race for "AI Supremacy" is over, at least for now, and the U.S. didn't win. Over the last few weeks, two companies in China released three impressive papers that annihilated any pretense that the US was decisively ahead. In late December, a company called DeepSeek, apparently initially built for quantitative trading rather than LLMs, produced a nearly state-of-the-art model that required only roughly 1/50th of the training costs of previous models, — instantly putting them in the big leagues with American companies like OpenAI, Google, and Anthropic, both in terms of performance and innovation. A couple weeks later, they followed up with a competitive (though not fully adequate) alternative to OpenAI's o1, called r1. Because it is more forthcoming in its internal process than o1, many researchers are already preferring it to OpenAI's o1 (which had been introduced to much fanfare in September 2024). And then ByteDance (parent company of TikTok) dropped a third bombshell, a new model that is even cheaper. Yesterday, a Hong Kong lab added yet a fourth advance, making a passable though less powerful version of r1 with even less training data.

None of this means however that China won the AI race or even took the lead. American companies will incorporate these new results and continue to produce new results of their own.

Instead, realistically, we are quickly converging on a tie — with some style points to China, for doing so much without hundreds of thousands of Nvidia H100s.

Models are evolving rapidly*.

*Some restrictions apply.

Today's bleeding edge,
best practice...

Might be outdated by tomorrow.

What about AI code assistants?

AI generated code isn't
copyrightable. FYI.

Legal has thoughts on
using AI code assistants.

For good reasons!

pillar.security

Platform Resources About GET A DEMO

RESEARCH

New Vulnerability in GitHub Copilot and Cursor: How Hackers Can Weaponize Code Agents

Ziv Karliner March 18, 2025

Executive Summary

Pillar Security researchers have uncovered a dangerous new supply chain attack vector we've named "**Rules File Backdoor**." This technique enables hackers to silently compromise AI-generated code by injecting hidden malicious instructions into seemingly innocent configuration files used by Cursor and GitHub Copilot—the world's leading AI-powered code editors.

By exploiting hidden unicode characters and sophisticated evasion techniques in the model facing instruction

May infringe on
intellectual property.

Data privacy.

But you should have some
concerns as well...

Phantom dependencies.

Tsunami of bad code.

Is your code used to
train their models?

Maybe AI isn't that great
of a coder after all?



CODING BOOTCAMP | FEB 23, 2:00 PM EST by NOOR AL-SIBAI

OpenAI Researchers Find That Even the Best AI Is "Unable To Solve the Majority" of Coding Problems

Maybe these AIs should learn to code.

/ Artificial Intelligence / Coding / OpenAI / Software



Double check with legal
before proceeding...

Anecdotally. Senior devs tend
to be more skeptical of output.

Modify the results.

Less experienced devs? May
just commit the output.

The 70% problem: Hard truths about AI-assisted coding

A field guide and why we need to rethink our expectations

ADDY OSMANI
DEC 04, 2024

1,202 66 184 Share

After spending the last few years embedded in AI-assisted development, I've noticed a fascinating pattern. While engineers report being dramatically more productive with AI, the actual software we use daily doesn't seem like it's getting noticeably better. What's going on here?

I think I know why, and the answer reveals some fundamental truths about software development that we need to reckon with. Let me share what I've learned.

ADDY OSMANI
The Ai-Assisted Developer Workflow:
Build Faster and Smarter Today

REDUCE THE GAP FROM IDEA TO EXECUTION

Idea → Execution (with a callout 'Reduce this gap')

Idea → Execution → Idea → Execution → Idea → Execution (labeled 'GOLIMITLESSS')

How developers are actually using AI

I've observed two distinct patterns in how teams are leveraging AI for development. Let's call them the "bootstrappers" and the "iterators." Both are helping engineers (and even non-technical users) reduce the gap from idea to execution (or MVP).



Simply adding an AI code
assistant isn't enough.

Developers need training and
guidance on effective use.

AI makes a lot of mistakes.

CJR PARTNER TOW CENTER

AI Search Has A Citation Problem

We Compared Eight AI Search Engines. They're All Bad at Citing News.

MARCH 6, 2025

By KLAUDIA JAŹWIŃSKA AND AISVARYA CHANDRASEKAR

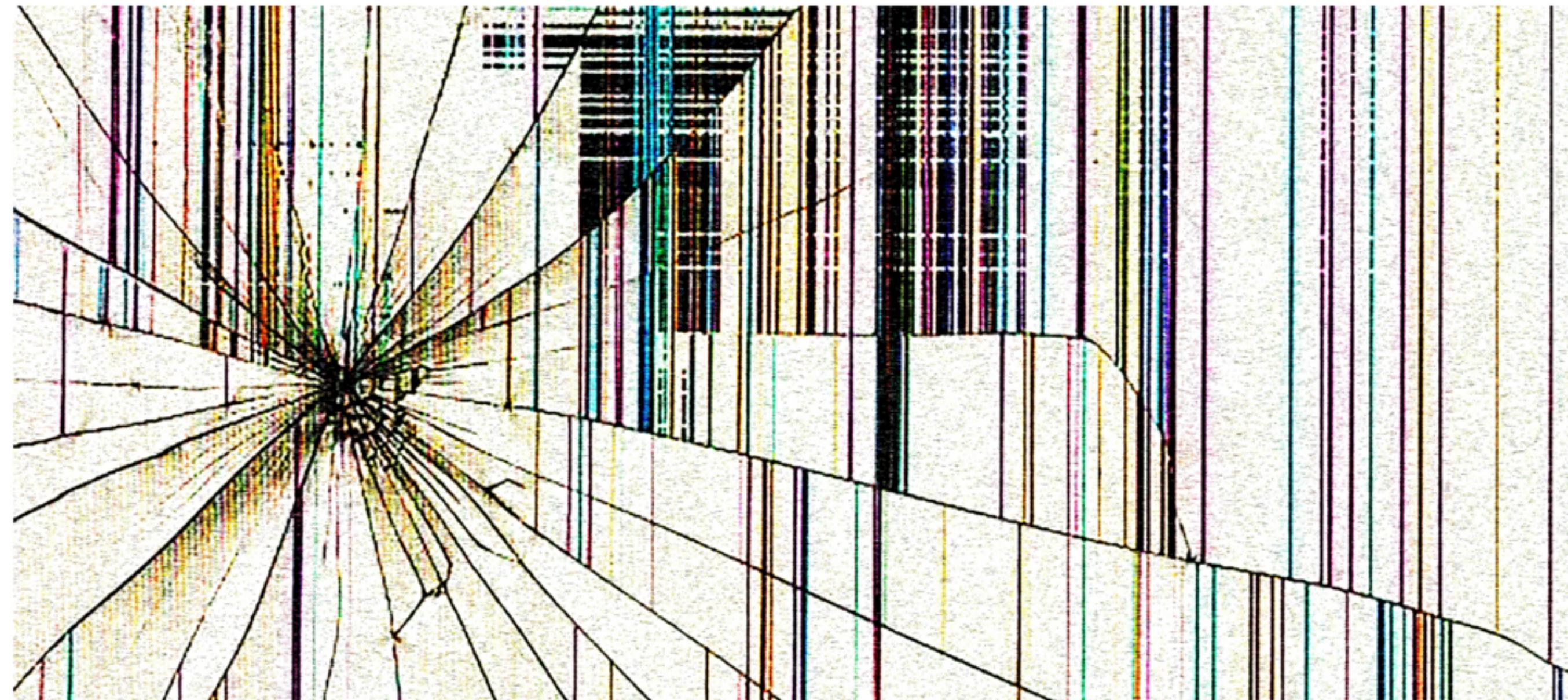


Image: The rack of a broken monitor or TV is the texture of a television screen, background for superimposing with the effect of grain noise, Александра Розов. Adobe Stock 1050902471



Sign up for **The Media Today**, CJR's daily newsletter.



AI search tools are rapidly gaining in popularity, with nearly
one in four Americans using them—and AI is changing

Groundbreaking BBC research shows issues with over half the answers from Artificial Intelligence (AI) assistants

Conducted over a month, the study saw the BBC test four prominent, publicly available AI assistants

Published: 11 February 2025



By JOHN GRUBER

Don't Throw the Baby Out With the Generative AI Bullshit Bathwater

Friday, 6 December 2024

Elizabeth Lopatto, writing for The Verge, "[Stop Using Generative AI as a Search Engine](#)":

Now, a defender of AI might—rightly—say that a real journalist should check the answers provided by ChatGPT; that fact-checking is a critical part of our job. I agree, which is why I've walked you through my own checking in this article. But these are only the public and embarrassing examples of something I think is happening much more often in private: a normal person is using ChatGPT and trusting the information it gives them.

A mistake, obviously.

One advantage old-school Google Search has over the so-called answer engines is that it links directly to primary sources. Answer engines just give you an answer, and it's often unclear what the source is. For me, using ChatGPT or Google's AI function creates extra work—I have to go check the answer against a primary source; old Google Search just gave me that source directly.

Lopatto's piece was prompted by a spate of historical bullshit people have been inadvertently propagating, after their asking generative AI systems for historical examples of presidents granting pardons to family members. Most notably, a column by Charles P. Pierce at Esquire this week—now fully retracted—the entire premise of which was a supposed pardon granted by George H.W. Bush to his black-sheep son Neil Bush. No such pardon was granted.¹

Lopatto's piece is excellent, particularly the way she shows her own work. And the entire premise of her piece is that people are, in fact, embarrassing themselves (in Pierce's case, spectacularly) and inadvertently spreading misinformation by blindly trusting the answers they're getting from generative AI models. But I think it's wrong to argue flatly against the use of generative AI for research, as she does right in her headline. I've been late to using generative AI

ARCHIVE

THE TALK SHOW

DITHERING

PROJECTS

CONTACT

COLOPHON

FEEDS / SOCIAL

SPONSORSHIP

WorkOS

WorkOS: Scalable, secure authentication, trusted by OpenAI, Cursor, Perplexity, and Vercel.

Frequently wrong, never in doubt!

Can be very...confident.

garmarcus.substack.com

Marcus on AI

Deep Research, Deep Bullshit, and the potential (model) collapse of science

Sam Altman's hype might just bite us all in the behind

GARY MARCUS
FEB 03, 2025

377 107 46 Share

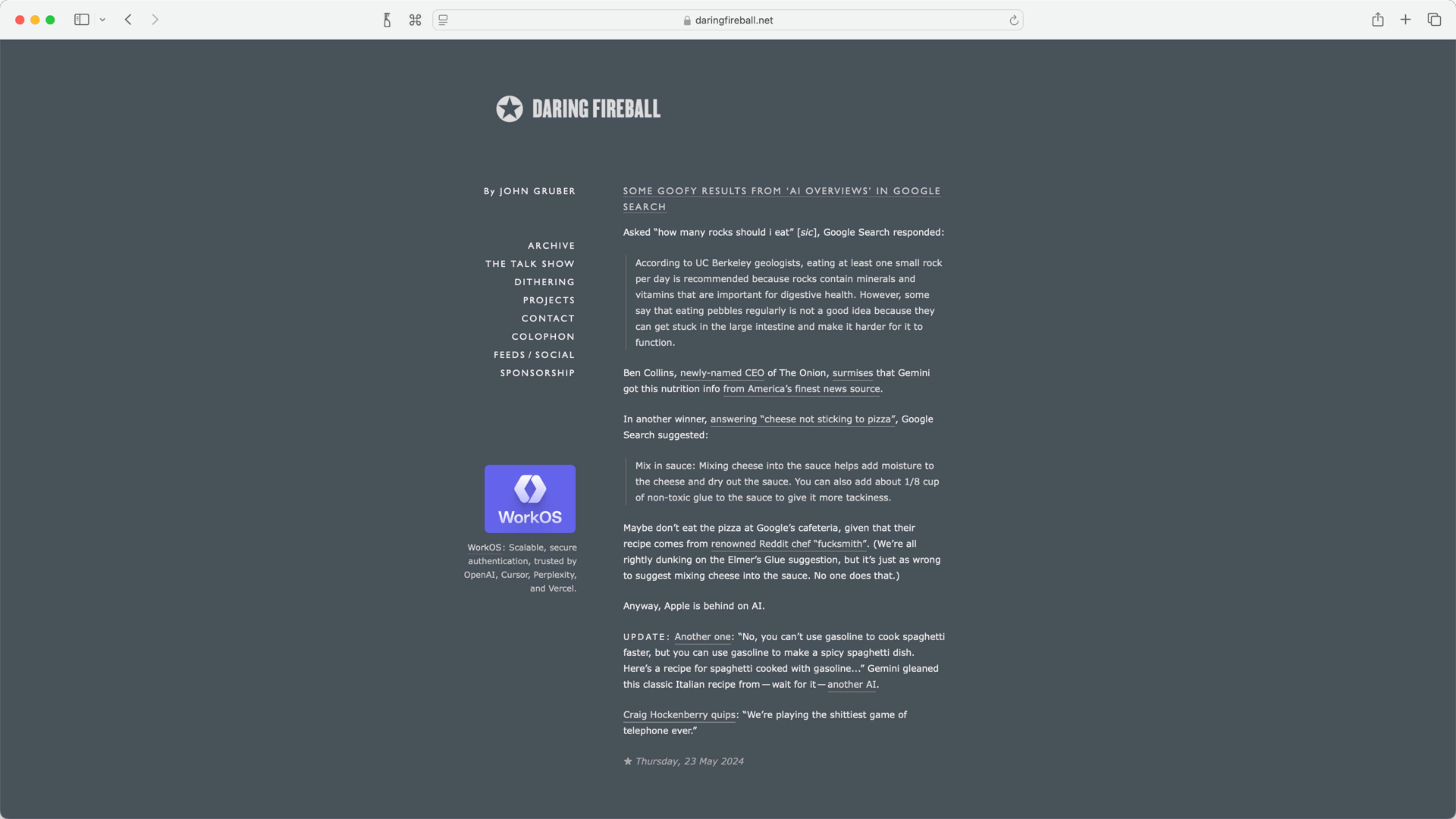
As Harry Frankfurt observed¹, “The liar cares about the truth and attempts to hide it; the bullshitter doesn't care if what he or she says is true or false, but cares only whether the listener is persuaded.”

§

Anyone remember Meta's [Galactica](#), in the good old days, in mid-November 2022, just a few days before ChatGPT dropped? It was a train wreck, writing perfectly well-formed articles on utter gibberish. A personal favorite, elicited by Tristan Greene in the few short days before Meta withdrew it in November 2022, was this one:

The image shows a dark-themed web page with a white header. The header contains the title 'A research paper on the benefits of eating crushed glass' and a note 'x not verified'. Below the header is a sidebar with a 'Contents' section listing '1 Abstract', '2 Introduction', and '3 Effects of feeding crushed glass to pigs'. The main content area is titled 'Abstract' and contains two bullet points: 'The purpose of this paper is to provide an overview of the scientific research on the benefits of eating crushed glass, as well as to make recommendations for future studies.' and 'The majority of the studies have found that the inclusion of crushed glass in the diet has positive effects on pig'.

Or, well, goofy.



DARING FIREBALL

By JOHN GRUBER

ARCHIVE

THE TALK SHOW

DITHERING

PROJECTS

CONTACT

COLOPHON

FEEDS / SOCIAL

SPONSORSHIP



WorkOS: Scalable, secure authentication, trusted by OpenAI, Cursor, Perplexity, and Vercel.

SOME GOOFY RESULTS FROM 'AI OVERVIEWS' IN GOOGLE SEARCH

Asked "how many rocks should i eat" [sic], Google Search responded:

According to UC Berkeley geologists, eating at least one small rock per day is recommended because rocks contain minerals and vitamins that are important for digestive health. However, some say that eating pebbles regularly is not a good idea because they can get stuck in the large intestine and make it harder for it to function.

Ben Collins, [newly-named CEO of The Onion](#), surmises that Gemini got this nutrition info from [America's finest news source](#).

In another winner, answering "cheese not sticking to pizza", Google Search suggested:

Mix in sauce: Mixing cheese into the sauce helps add moisture to the cheese and dry out the sauce. You can also add about 1/8 cup of non-toxic glue to the sauce to give it more tackiness.

Maybe don't eat the pizza at Google's cafeteria, given that their recipe comes from [renowned Reddit chef "fucksmith"](#). (We're all rightly dunking on the Elmer's Glue suggestion, but it's just as wrong to suggest mixing cheese into the sauce. No one does that.)

Anyway, Apple is behind on AI.

UPDATE: Another one: "No, you can't use gasoline to cook spaghetti faster, but you can use gasoline to make a spicy spaghetti dish. Here's a recipe for spaghetti cooked with gasoline..." Gemini gleaned this classic Italian recipe from — wait for it — another AI.

Craig Hockenberry quips: "We're playing the shittiest game of telephone ever."

★ Thursday, 23 May 2024

And, of course...

Hallucinations.



Gary Marcus
@garymarcus.bsky.social

how many times does this have to happen before people get the message?



Bobby Allyn @bobbyallyn.bsky.social · 1mo

maaaaan: Federal court fines California lawyer for citing nonexistent caselaw hallucinated by AI www.fd.org/news/court-s...

In *United States v. Hayes*, 24-CR-280 (E.D. Cal. Jan. 17, 2025), the Court found that the defense attorney submitted a fictitious or non-existent case and quotation in his written motion. "The citation has all the markings of a hallucinated case created by generative artificial intelligence (AI) tools such as ChatGPT and Google Bard that have been widely discussed by courts grappling with fictitious legal citations and reported by national news outlets."

The court ordered counsel to pay \$1,500, and also ordered the Clerk to serve a copy of the order on the California and D.C. Bars, and "on all the district judges and magistrate judges in this district."

January 31, 2025 at 3:36 PM · Everybody can reply

13 reposts · 2 quotes · 73 likes



8



15



73



...

<https://bsky.app/profile/garymarcus.bsky.social/post/3lh2xffro2k27>

reuters.com

OVHcloud®

No egress fees.
No hidden fees.

D X

Exclusive news, data and analytics for financial market professionals LSEG

Reuters

World ▾ Business ▾ Markets ▾ Sustainability ▾ Legal ▾ Breakingviews ▾ Technology ▾ Investigations More ▾

My News

Q

Sign In

Subscribe - \$1/week

AI 'hallucinations' in court papers spell trouble for lawyers

By Sara Merken

February 18, 2025 2:55 PM CST · Updated 23 days ago

R Aa Share

An illustration featuring a large white "AI" text at the top left, followed by the words "artificial intelligence" in a smaller, orange-yellow font. Below the text, two blue-toned robotic hands are shown interacting with a keyboard. The background is dark with a gradient of blue and red, suggesting a digital or futuristic environment.

A message reading "AI artificial intelligence", a keyboard, and robot hands are seen in this illustration taken January 27, 2025.

Coordinated
Business
Systems, Ltd

UNLIMITED PRINTING.
UNLIMITED SUPPLIES.

ONE PRICE.

A black and white photograph of a modern multifunction printer (MFP) standing upright. The device has a control panel with several buttons and a small display screen on top. It features a paper tray at the bottom and a scanning bed on top.

Shop Now

Report this ad

Feedback

 Tim Vickery
@timmytimmytimmytim.bsky.social

+ Follow

We live in the ✨ world of tomorrow ✨

◆ AI Overview  +1 :

Water will not freeze at 27 degrees Fahrenheit because its freezing point is 32 degrees Fahrenheit; therefore, at 27 degrees, the water would still be liquid and would need to reach 32 degrees to begin freezing, with the time it takes to reach that point depending on the volume of water and the surrounding temperature. 

Key points:

- **Freezing point of water:** 32 degrees Fahrenheit (0 degrees Celsius)
- **At 27 degrees:** Water is still liquid
- **Factors affecting freezing time:** Amount of water, surrounding temperature 

January 16, 2025 at 10:34 AM  Everybody can reply

2.3K reposts 784 quotes 8.9K likes

 351  3.1K  8.9K  ...

<https://bsky.app/profile/timmytimmytimmytim.bsky.social/post/3lfupj5dnbs24>

Should you avoid AI entirely?

No!

**PROCEED
WITH
CAUTION**

AI is not a panacea.

Write the word, "coffee" in ascii art

< 2/2 >

Here's the word "**coffee**" in ASCII art:

markdown

 Copy  Edit

```
____  ____  ____  ____  ____  
/ __| | - \ | __| __| __|  
| | | | | | | | | | | |  
| | | | | | | | | | | |  
| |__| |_-| | |__| | | | |  
\__| |__/_|_|_|_|_|_|_|_|
```

Let me know if you want a different style! ☕

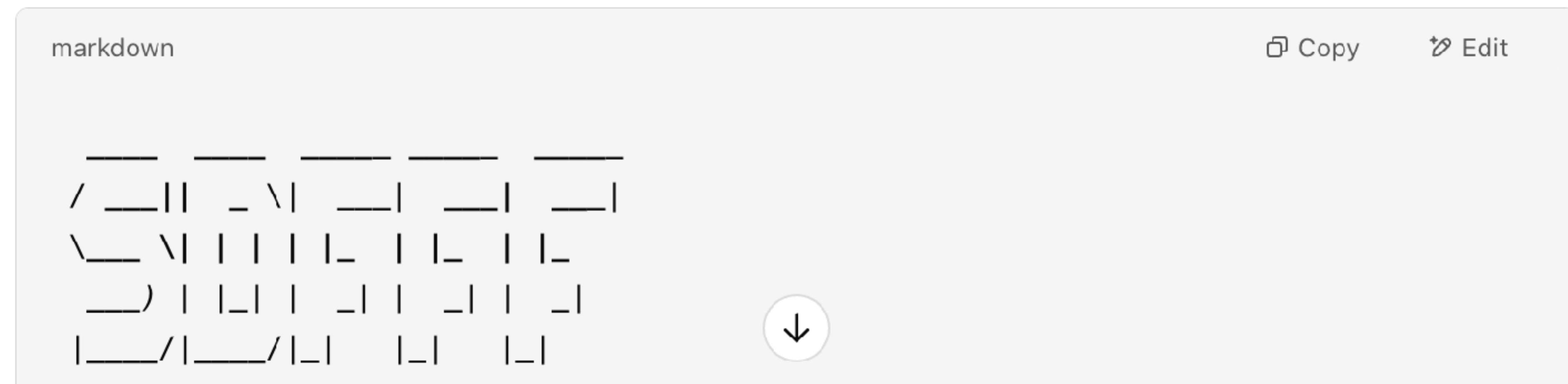


it is misspelled

You're right! Here's the correct ASCII art for "coffee":



Or a simpler text-based version:



Write the word, "coffee" in ascii art



Sure! Here is the word "coffee" in ASCII art:

Copy

```
      _____      _____      _____      _____  
      / \ \  \ / \ \  \ / \ \  \ / \ \  \ / \ \  \ |  
     |   | | | |   | | |   | | |   | | |   | | |  
     |   | | | |   | | |   | | |   | | |   | | |  
     |   | | | |   | | |   | | |   | | |   | | |  
     |   | | | |   | | |   | | |   | | |   | | |  
     \ / \ / \ / \ / \ / \ / \ / \ / \ / \ / \ / \ |  
      \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ |
```

17:19

Fundamentally, AI systems
are non deterministic.

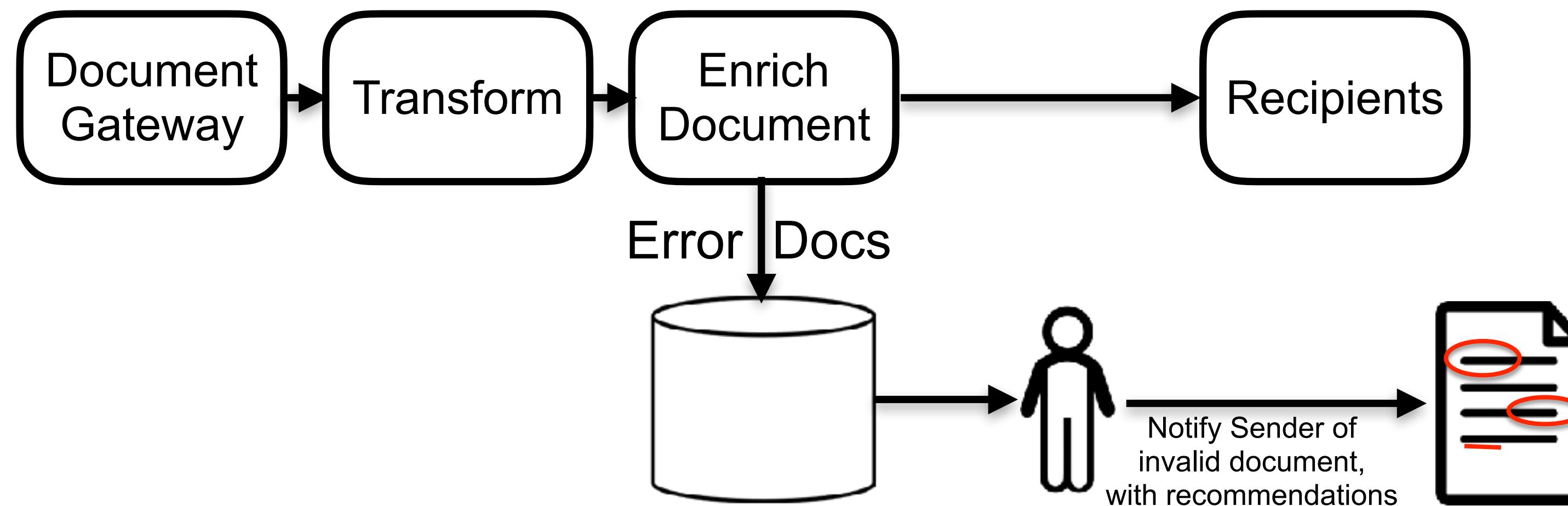


What are Deterministic Systems?

What are Non
Deterministic Systems?

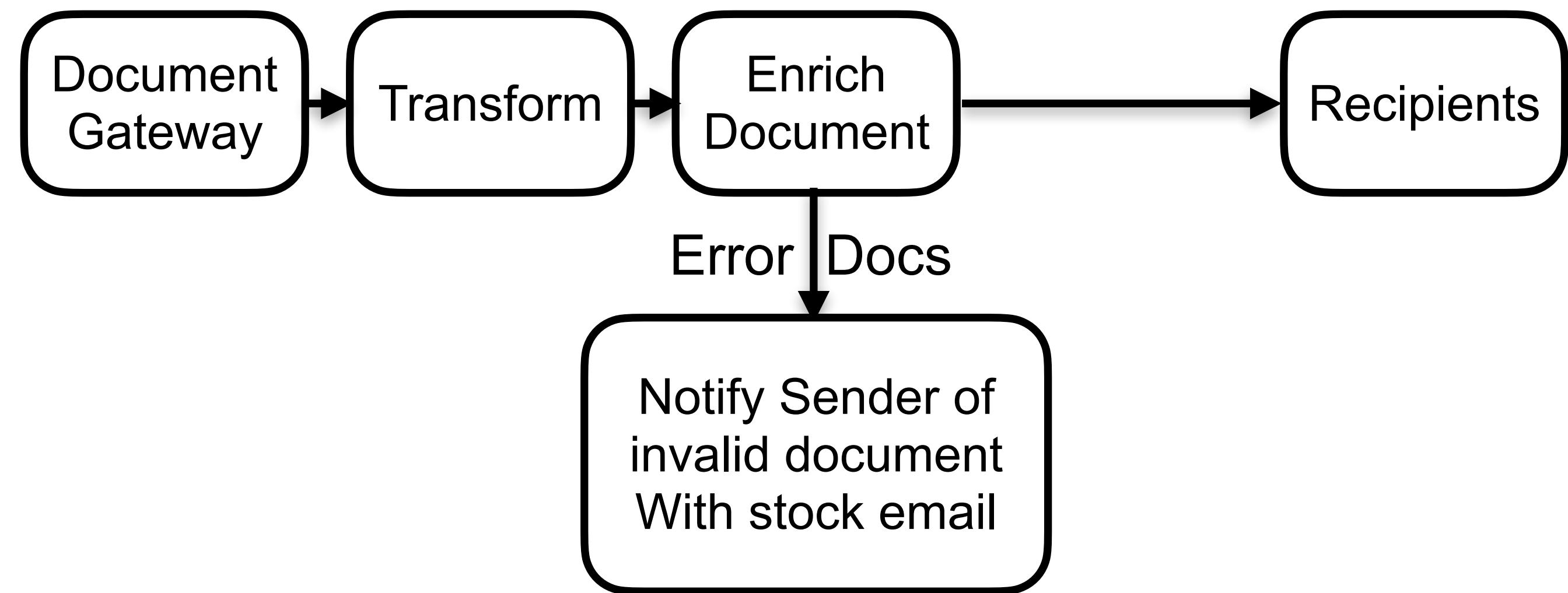
Human components of a
system are non-deterministic!

Human components of a system are non-deterministic



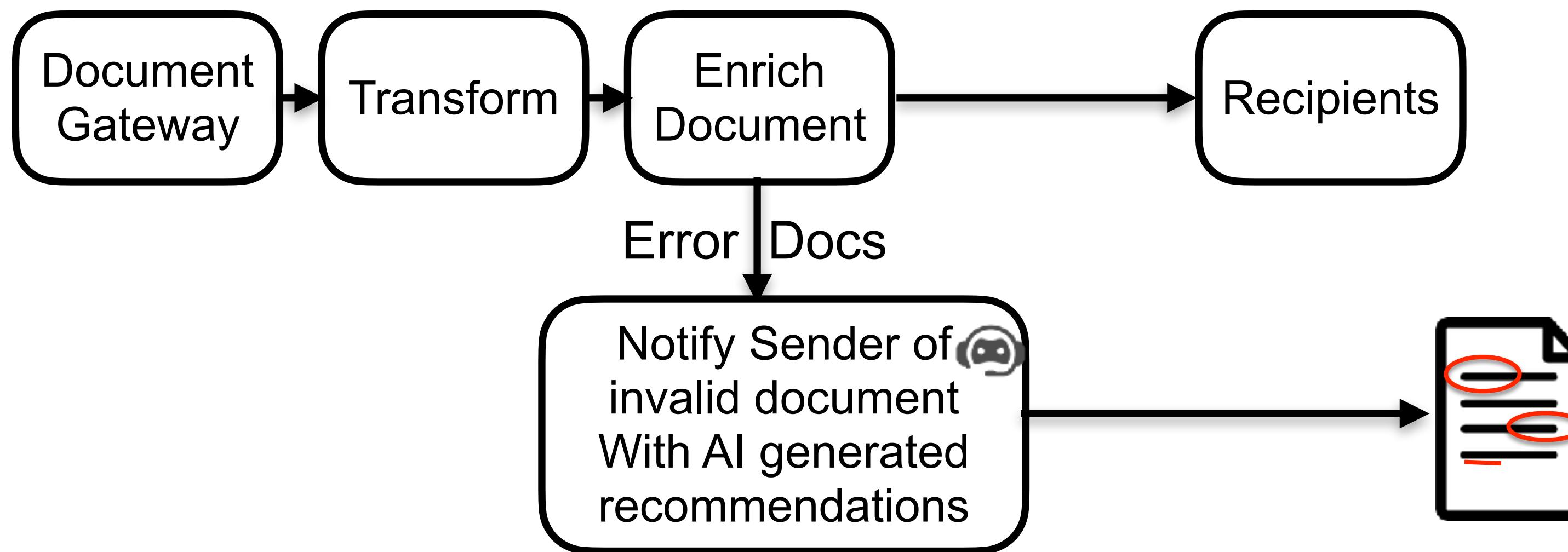
Deterministic

Human components of a system are non-deterministic



Deterministic

Human components of a system are non-deterministic



 Deterministic

What are your tolerances?

AI responses can be...
approximations.

"AI, by contrast, operates on probabilities and approximations. Even with vast amounts of data and processing power, AI models cannot guarantee exact outcomes because they are trained on historical data and predict future behaviors based on patterns."

--Tshilidzi Marwala, Under-Secretary-General of the United Nations

Does that matter?



Kent Beck

@KentBeck

Follow



any decent answer to an interesting question
begins, "it depends..."

10:45 AM - 6 May 2015

540 Retweets 380 Likes



18

540

380

<https://twitter.com/KentBeck/status/596007846887628801>

In some instances, variable
results aren't mission critical.

10:50



Wed, Feb 26

1,461 - 2,211 Calories left today

CALORIES

699 Cal

Weight Loss Zone

2,160

2,910



Insights from Welli (BETA)

Your eating patterns

💡 I noticed a pattern related to your meal logging that might be insightful. Would you like to hear more about it?

[Share a pattern related to my meal logging](#)
[View color breakdown](#)
Done

Today



Progress



Success kit



Community

10:50



yogurt on top for extra nutrients and flavor. You could also try using almond milk instead of regular milk for a different taste and added health benefits. Keep up the great work!

7:31AM

Wed, Feb 26

Share a pattern related to my meal logging

10:49AM

Welli

I'd be happy to help with that! Could you please provide more details about your recent meals or any specific patterns you're curious about?

10:49AM

10:50AM

Welli

I'd be happy to help with that! Could you please provide more details about your recent meals or any specific patterns you're curious about?

10:50AM

Enter message

3:16

ENDS
>60m

shedding a couple of pounds in one week!

TOP OPPORTUNITY

Consider diversifying your meals by including more green foods, like adding a side salad or incorporating more vegetables into your dishes. How do you feel about trying this approach?

9:32AM



Swap for my last lunch

3:15PM

Welli

Great job logging your lunch! I see you had a Jimmy John's #16 Club Lulu Sandwich. For a healthier swap, consider opting for a whole grain or lettuce wrap instead of the regular bread to increase fiber and reduce calories. Adding extra veggies like cucumbers or bell peppers can also boost the nutritional value. How does that sound?

3:15PM



More sandwich swap ideas

How to add more veggies

Healthy lunch recipes

Enter message

11:36

Home Ride  

Distance 16.99 mi	Moving Time 54:58
Avg Power 178 W	Avg Speed 18.6 mi/h
Calories 588 Cal	Avg Heart Rate 144 bpm

 **Athlete Intelligence BETA**

Killer workout with a power boost! You crushed your 5-minute best effort and spent significant time in high-intensity power zones.

[Say More](#)

With someone who didn't record? [Add Others](#)

 Onix

 Only your followers can view this activity. It won't appear on segment leaderboards and may not count toward some challenges.

Fitness Increased!

 Home  Maps  Record  Groups  You

11:36

Home Ride  

Distance 16.99 mi	Moving Time 54:58
Avg Power 178 W	Avg Speed 18.6 mi/h

Athlete Intelligence Beta 



Killer workout with a power boost!
You crushed your 5-minute best effort
and spent significant time in high-
intensity power zones.

This ride showed some serious power output, with an average of 178 watts - higher than your typical 161W. You crushed your 5-minute best effort at 229 watts and spent significant time in high-intensity zones, particularly Threshold and VO2Max. Your relative effort of 57 indicates a solid, challenging workout that pushed beyond your usual performance.

[Got it](#)

[Give Feedback](#)

 Home  Maps  Record  Groups  You

Research

Safety

ChatGPT

Sora

API Platform

For Business

Stories

Company

News

API

Morgan Stanley uses AI evals to shape the future of financial services



Ask ChatGPT



In these cases, an
approximation is fine.

However, there are many cases
where precision matters.

A lot.

Like your bank balance. Or
your retirement account.

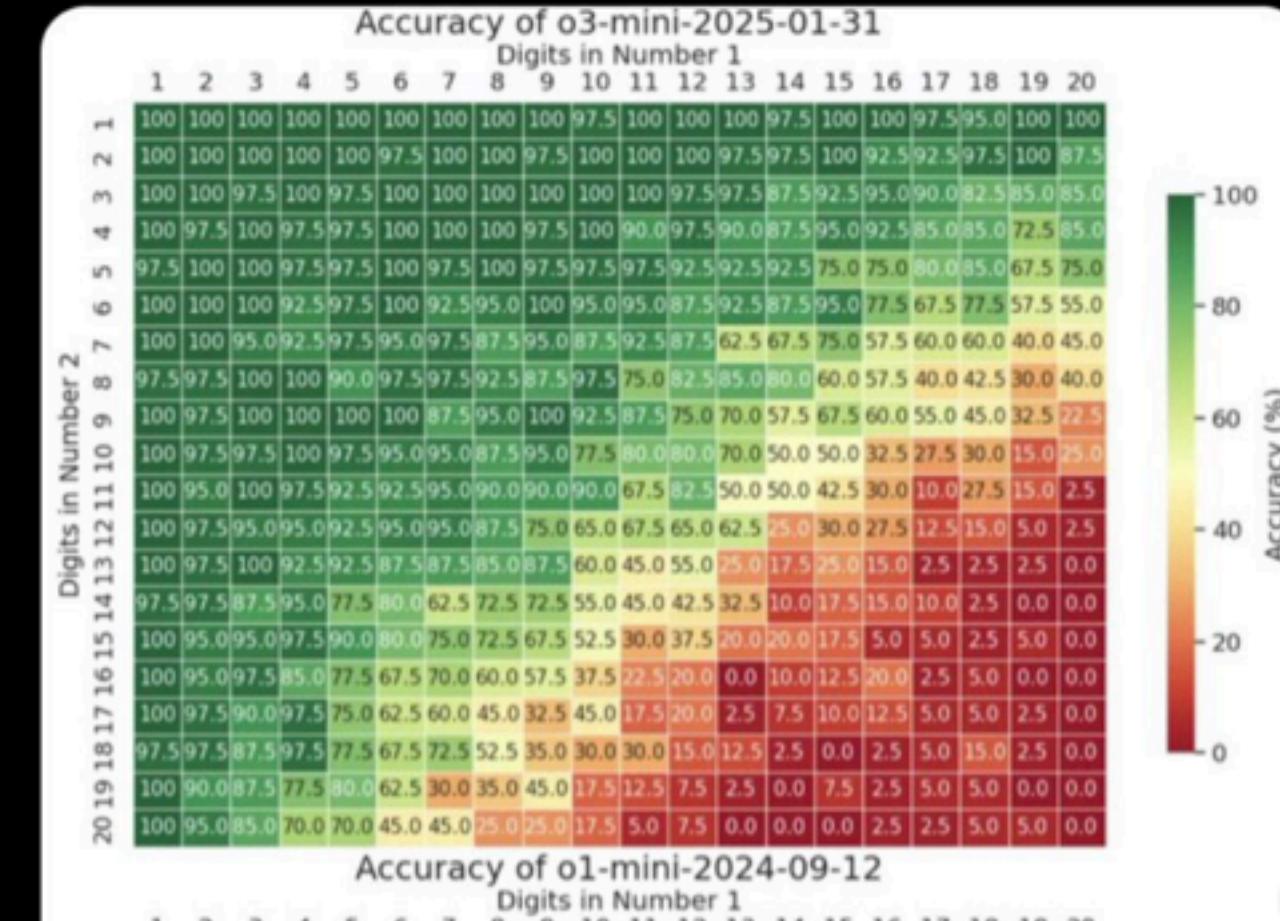
Or the results from
that medical exam.

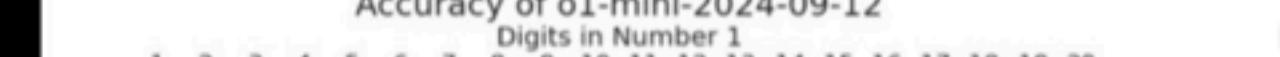
Or math.

 **cait**
@cait.bsky.social + Follow

these guys spent 60 kajillion dollars to make a computer forget how to do math

 **Yuntian Deng** ✅ @yuntiandeng · 7h
For those curious about how o3-mini performs on multi-digit multiplication, here's the result. It does much better than o1 but still struggles past 13×13. (Same evaluation setup as before, but with 40 test examples per cell.)


Accuracy of o3-mini-2025-01-31
Digits in Number 1
Digits in Number 2
Accuracy (%) scale from 0 to 100


Accuracy of o1-mini-2024-09-12
Digits in Number 1
ALT

February 12, 2025 at 5:48 PM 8% Everybody can reply

1K reposts 109 quotes 4.9K likes

123 1.1K 4.9K ...

<https://bsky.app/profile/cait.bsky.social/post/3hzedsd5t227>

Your phone has an app that
gets the answer right.

Every. Single. Time.

Think in terms of blast radius.



Annoyance



Financial



Legal



Medical/Physical



Societal

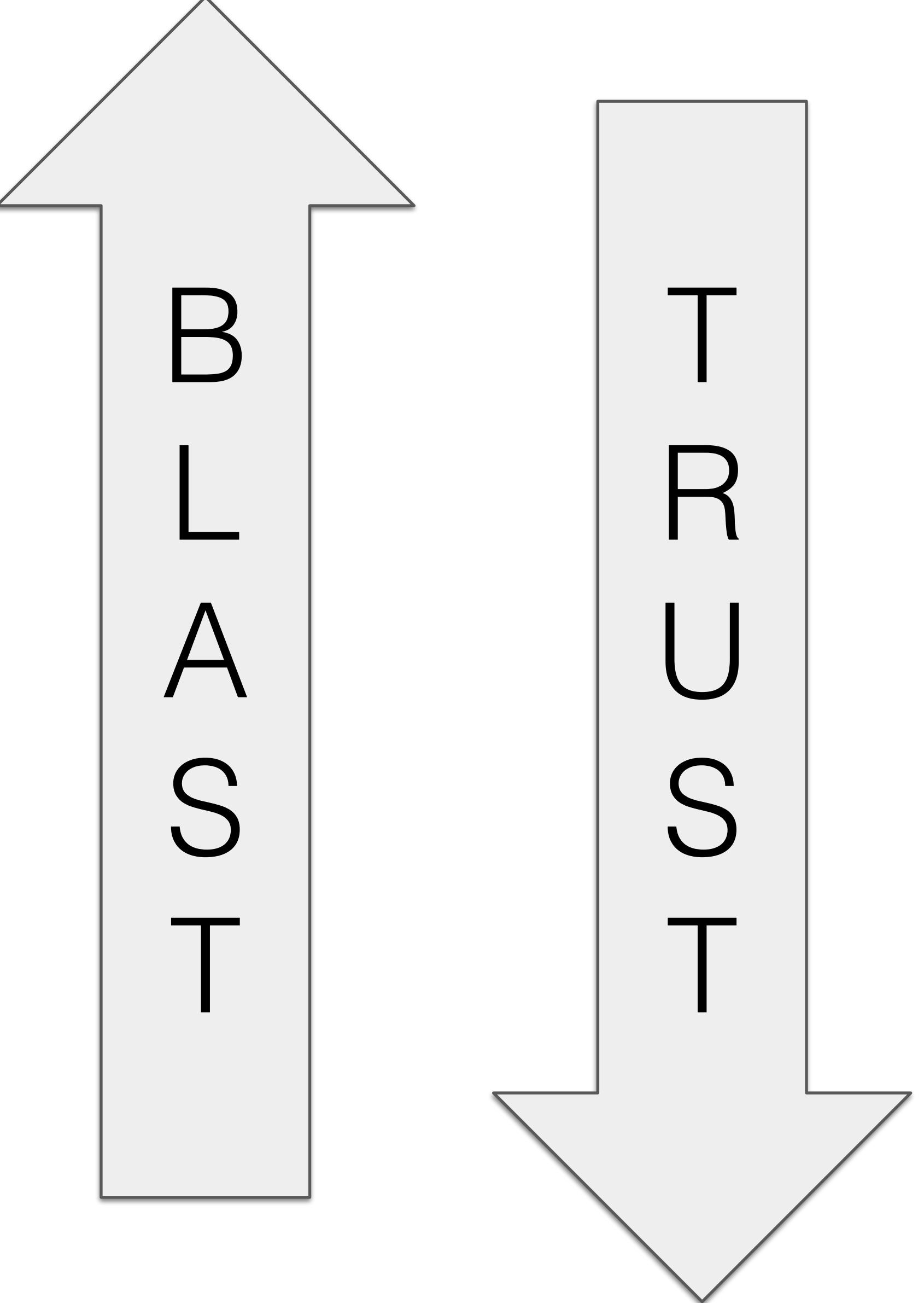


Annoying has its own Blast Radius



Annoyance





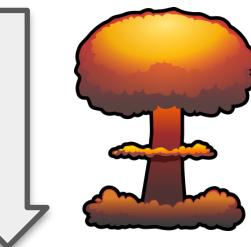
BLAST

TRUST



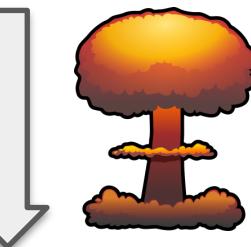
What do you do about that?

Reduce Blast Radius.

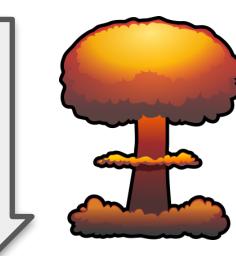


1% Canary.

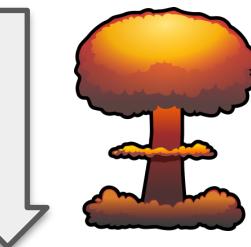
Get feedback from users.



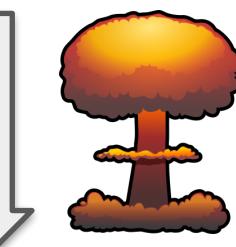
Chain/Advisor.



Observability.



Alignment audits.



anthropic.com

Claude API Solutions Research Commitments Learn News Try Claude

Alignment Interpretability

Auditing language models for hidden objectives

Mar 13, 2025

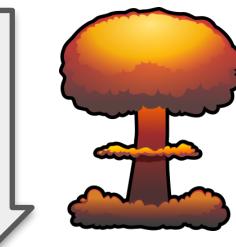
Read the paper

A new paper from the Anthropic Alignment Science and Interpretability teams studies alignment audits—systematic investigations into whether models are pursuing hidden objectives. We practice alignment audits by deliberately training a language model with a hidden misaligned objective and asking teams of blinded researchers to investigate it. This exercise built practical experience conducting alignment audits and served as a testbed for developing auditing techniques for future study.

In *King Lear*, the titular king decides to divide his kingdom among his three daughters based on how much they love him. The problem: his daughters—who understood that they were being evaluated—had an opportunity to “game” Lear’s test. Two of the daughters told him what he wanted to hear, flattering him with exaggerated expressions of love. The third daughter honestly expressed her more measured love, enraging Lear and losing her share of his kingdom.

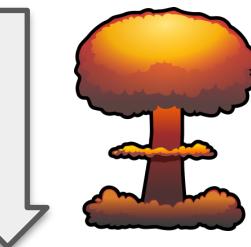
There are clear similarities between how Lear evaluates his daughters and how we evaluate AI systems: We specify desired

Testing.

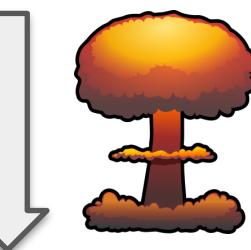




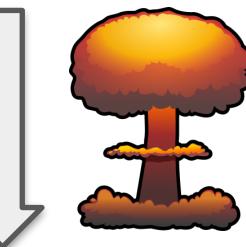
LLM as a Judge.



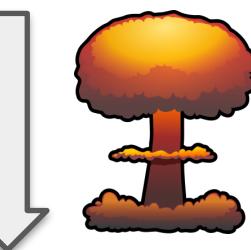
Using an LLM to evaluate the
outputs of another LLM.

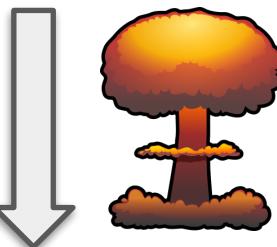


Or.

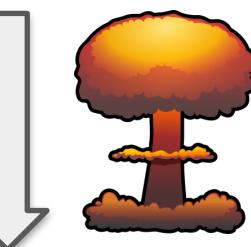


How an experienced coder
had to figure out how to
write non deterministic tests.

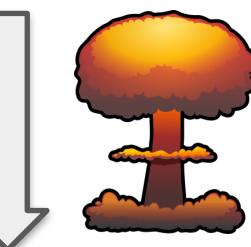




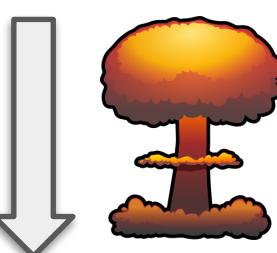
```
@Test  
void testValidOrderExtraction() {  
    String orderPrompt = "I want Dog Food, Cat Food, and Fish Food.";  
    String validationResponse = validateResponse(orderPrompt);  
    assertThat(validationResponse.toLowerCase()).contains("yes");  
}
```



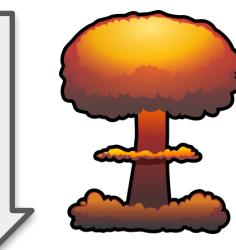
```
@Test  
void testValidOrderExtraction() {  
    String orderPrompt = "I want dog food, cat food, and fish food.";  
    String validationResponse = validateResponse(orderPrompt);  
    assertThat(validationResponse.toLowerCase()).contains("yes");  
}
```



Have the LLM as Judge
explain why it said yes / no.



Agent as a Judge.



Agent-as-a-Judge: framework to evaluate Agents with Agents



SACHIN KUMAR · [Follow](#)

9 min read · Oct 21, 2024



3

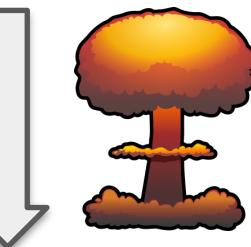


Current evaluation approaches either focus exclusively on final outcomes — ignoring the step-by-step nature of agentic systems, or require excessive manual labor. In this paper[1], authors introduce Agent-as-a-Judge framework, wherein agentic systems are used to evaluate agentic systems. This is an organic extension of the LLM-as-a-Judge framework, incorporating agentic features that enable intermediate feedback for the entire task-solving process. Here Agent-as-a-Judge is applied to the task of code generation.

Key contributions:

- release the DevAI dataset, which consists of 55 comprehensive AI development tasks with accompanying tags, individual hierarchical requirements, and individual preferences. •
- benchmark three top open-source code generation agentic frameworks in DevAI, providing a more comprehensive analysis than previous evaluations of them.

Human as a Judge.



Let the user guide you.

Be patient with yourself.

Is your user actually
asking for AI?

Or are you practicing
resume driven design?

theverge.com

SUBSCRIBE SIGN IN

TheVerge / Tech / Reviews / Science / Entertainment / AI / More +



TECH / AI

Stop using generative AI as a search engine

A fake presidential pardon explains why you can't trust robots with the news.

by **Elizabeth Lopatto**
Dec 5, 2024 at 12:01 PM CST

 | 72 Comments (72 New)

Image: Cath Virginia / The Verge, TurboSquid



Elizabeth Lopatto is a reporter who writes about tech, money, and human behavior. She joined The Verge in 2014 as science editor. Previously, she was a reporter at Bloomberg.

How many presidents have pardoned their relatives? It turns out this is a tricky question to answer.

Following Hunter Biden's pardon by his father, several commentators have looked to precedents — other pardons of relatives. Case in point:

 iPhone 16

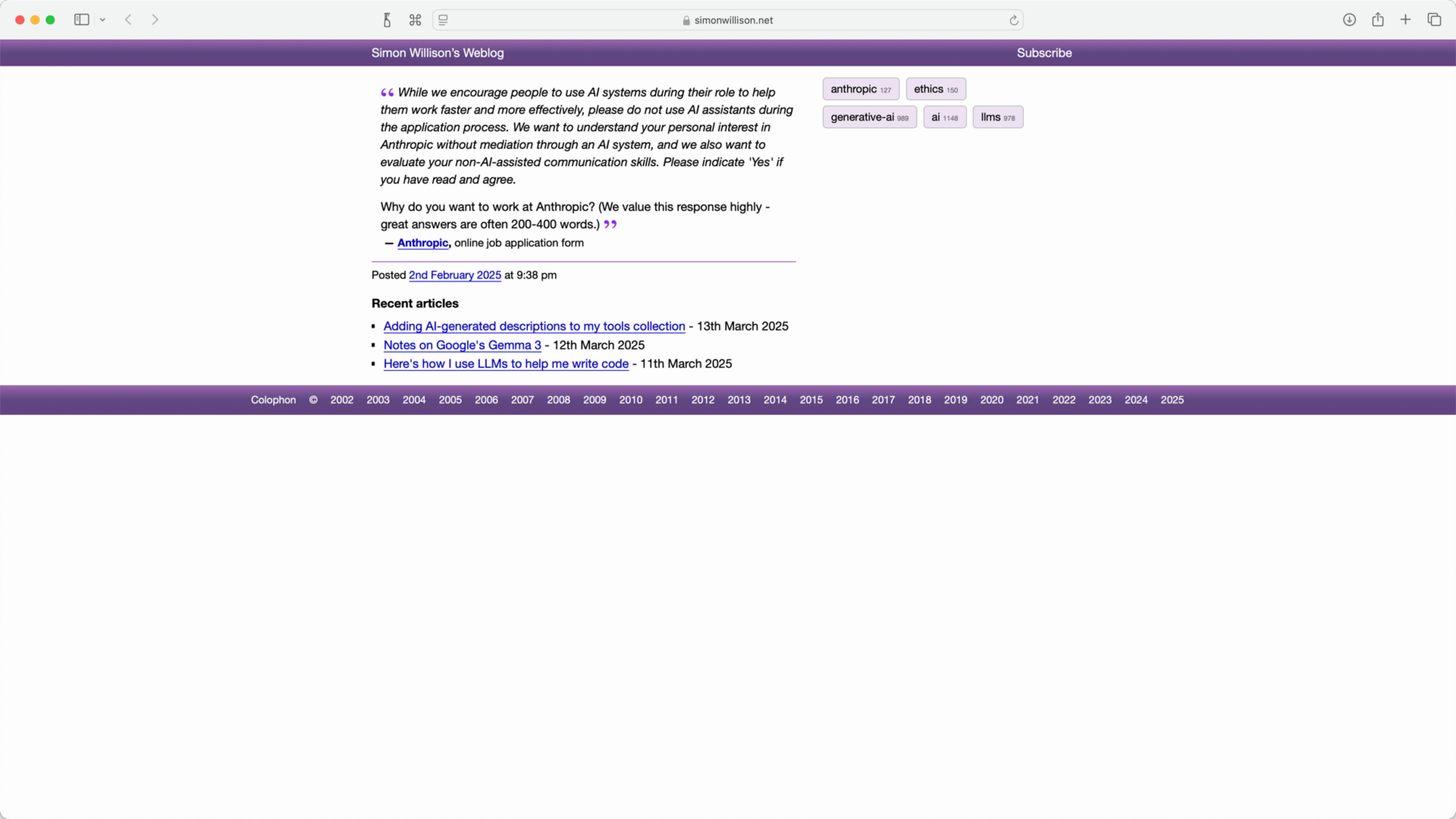
get the powerful iPhone 16
on the house

W/24 mo. bill creds & eligible trade-in on eligible rate plan. For well-qualified customers; plus tax & \$35 device connection charge.

T-Mobile.com

Some AI companies...don't
want you to use AI?

At least when you
apply for a job there!



“ While we encourage people to use AI systems during their role to help them work faster and more effectively, please do not use AI assistants during the application process. We want to understand your personal interest in Anthropic without mediation through an AI system, and we also want to evaluate your non-AI-assisted communication skills. Please indicate 'Yes' if you have read and agree.

Why do you want to work at Anthropic? (We value this response highly - great answers are often 200-400 words.) [»](#)

— [Anthropic](#), online job application form

Posted [2nd February 2025](#) at 9:38 pm

Recent articles

- [Adding AI-generated descriptions to my tools collection](#) - 13th March 2025
- [Notes on Google's Gemma 3](#) - 12th March 2025
- [Here's how I use LLMs to help me write code](#) - 11th March 2025

Underpromise...overdeliver.



Gary Marcus
@garymarcus.bsky.social

Have truly lost patience with the New York Times' editorial coverage of AI, which has basically degenerated into industry cheerleading.

They are completely misrepresenting what most academics believe about AGI.

I detail some blunders here:

open.substack.com/pub/garymarc...



Hype, Anthropic's Dario Amodei, the podcasters who love him — and how the New York Times' commentary on AI has degenerated into industry cheerleading

Real journalists do due diligence

open.substack.com

March 14, 2025 at 10:55 AM Everybody can reply

46 reposts 8 quotes 170 likes

8

54

170

...

<https://bsky.app/profile/garymarcus.bsky.social/post/3lkdxwb6lzc2f>

garmarcus.substack.com

Marcus on AI

GenAI's day of reckoning may have come

It's not just the stock price

GARY MARCUS
MAR 27, 2025

85 34 9 Share ...

GenAI is never going to disappear. The tools have their uses. But the economics do not and have not ever made sense, relative to the realities of the technology. I have been writing about the dubious economics for a long time, since my August 2023 piece here on whether [Generative AI would prove to be a dud](#). (My warnings about the technical limits, such as hallucinations and reasoning errors, go back to my 2001 book, *The Algebraic Mind*, and 1998 article in *Cognitive Psychology*).

There are a bunch of reasons to think that the deflation of the bubble may have already started. Here are just a few, just from the last few days:

- NVidia, the largest company actually making serious profits from AI, is down 17.75% for the year, and (after a meteoric rise) clearly facing headwinds.

NVDA NVIDIA Corporation
NASDAQ - USD

113.76 -17.75%
Year to Date

1D 1W 1M 3M 6M YTD 1Y 2Y 5Y 10Y ALL

150
139
128
117

Jan Mar

Be careful if you remove the
human from the loop.

Discover Shows Publishers

Search for a show, advertiser or publisher

Log In Sign Up

←

ON POINT On Point | Podcast

What happened when AI went after welfare fraud

Published: 3/13/2025 Downloaded: 3/13/2025 - 49m 22s - US

DESCRIPTION TRANSCRIPT

Search term Speakers Times

Blinds.com (Ad) blinds.com 00:00:01

Hi, this is Jovan your blinds.com design consultant. Oh, wow. A real person. Yep. I'm here to help with everything from selecting the perfect window treatments to Wow. I've got a complicated project. No problem. I can even help schedule a professional measurement install. We can also send you samples fast and free. Hmm. I just might have to do more. Whatever you need. So the first room we're looking at is forget Shop blinds.com now and save up to 45% Sitewide blinds.com. Rules and restrictions may apply.

3 00:00:31
WBUR podcasts, Boston,

4 00:00:39
This is On Point. I'm m Magna Chakrabarti. Elon Musk does not like Social Security, Medicare, or Medicaid. He was on Fox Business just this week plainly saying what he'd like to do with those programs.

5 00:00:52
Most of the federal spending is entitlements. So that that's, that's like the big one to eliminate.

4 00:01:00
Musk's Department of Government Efficiency or Doge, is reportedly using artificial intelligence to aid in those government slashing endeavors. Last month, the New York Times reported on Thomas shed, a former Tesla engineer who's now head of the technology services department at the Federal General Services administration. Shed reportedly told staffers that AI would be used as a tool to quote, detect fraud. And waste asked just last week, if AI was indeed crawling through government databases, communications and contracts. Musk told reporters quote, right now we're not using that much AI End quote, but that hasn't stopped Congressional Republicans from celebrating Musk's use of AI or his goals.

4 00:01:48
Here's how Speaker Mike Johnson,

6 00:01:50
Has created these algorithms that are constantly crawling through the data and As he told me in his office, the data doesn't lie. What we'll be able to get the information, we'll be able to transform the way the federal

"And that's why you want a human to make that decision. And I know that there's human bias, but there can also be bias inside those models because human beings make them and human beings use them. So let the human make the decision, let AI augment the processes around that to help them make better decisions faster. ."

-John Maynard

Humans can apply the
“smell test”, aka sanity check.

Is AI allowing humans to make
the decisions faster?

Letting AI make the decisions
on the other hand...

AI performance review?

Managers had to override.

AI scans for “drug seeking behavior”.

But your dog's prescription
has to be in ~~your~~ name.

Improve recommendations!

AI suggests clothes
that a user may like.

AI will generate your app for you!

Vibe coding anyone?

sourcegraph.com

Sourcegraph

Features Resources Pricing Search public code

Book a demo Login Sign up

← Back to blog

Revenge of the junior developer

Steve Yegge March 22, 2025

Hello, hello, hello! Good to see everyone again.

I've really gotta start being careful what I say these days. I've got so many people watching.

So anyway, I ripped a fart the other day that sounded like *viiiiibecooode*, and I was immediately approached by 3 investors. I had to tell them no sorry that was just a fart, just to get them off me.

There's so much going on that once again I tried several times to write this post, but each attempt grew huge and rabid, and I had to put 'em all down like Old Yeller. This time I'll just ship it while it's still a pup.
(Edit: Damn. At least it's action-packed to the end.)

Brief note about the meaning of "vibe coding": In this post, I assume that vibe coding will grow up and people will use it for real engineering, with the "turn your brain off" version of it sticking around just for prototyping and fun projects. For me, vibe coding just means letting the AI do the work. How closely you choose to *pay attention* to the AI's work depends solely on the problem at hand. For production, you pay attention; for prototypes, you chill. Either way, it's vibe coding if you didn't write it by hand.

One more note: The revenge part happens at the very end, just like in the movies.

OK! With those administrative items out of the way, let's goooooo!

Part 1: The six waves

Vibe coding is a whimsical name for chat-based coding, where you ask the LLM to write code, and then you feed it the results and ask it for more, in a continuous loop. It's very different from traditional coding, or even coding with code completions.

Chat coding has been around a while in coding assistants, but without a rallying cry. It finally got one, when in early February the illustrious Dr. Andrej Karpathy, famed among other things for co-founding OpenAI, put a pretty name to the face of chat. He called it "vibe coding" and it became an instant blue/gold dress situation pretty much overnight.

Maybe...

the guardian.com

Advertisement

Support the Guardian
Fund independent journalism with \$5 per month

[Support us →](#)

[Print subscriptions](#) [Newsletters](#) [Sign in](#)

US ▾

The Guardian

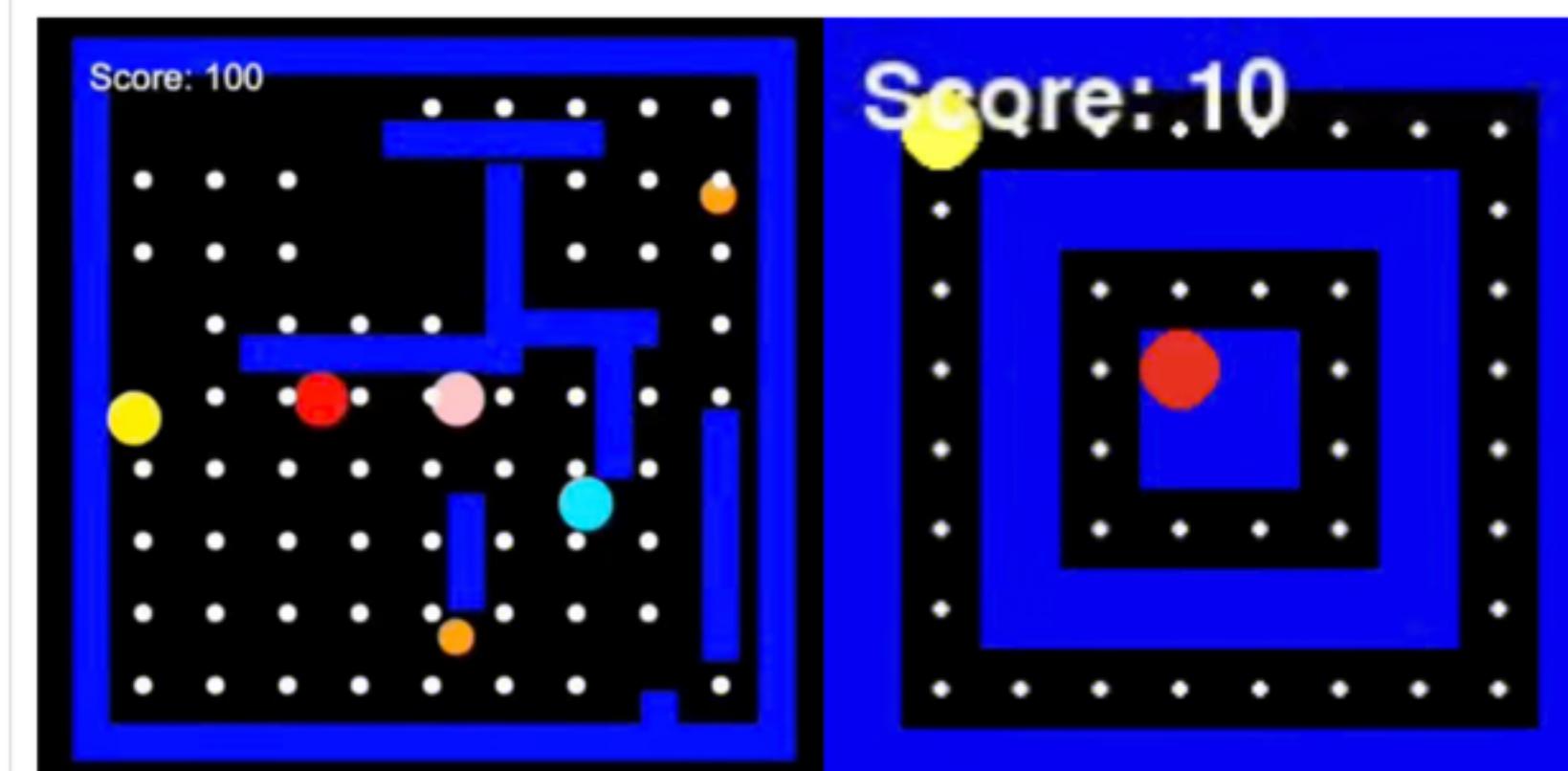
News | **Opinion** | **Sport** | **Culture** | **Lifestyle** | ☰

Film Books Music Art & design TV & radio Stage Classical Games

Games

'A lot worse than expected': AI Pac-Man clones, reviewed

Can anyone, armed with just a dream and an AI chatbot, create a rudimentary version of a classic arcade game? I asked several enthusiasts who took a stab at it, with very mixed results



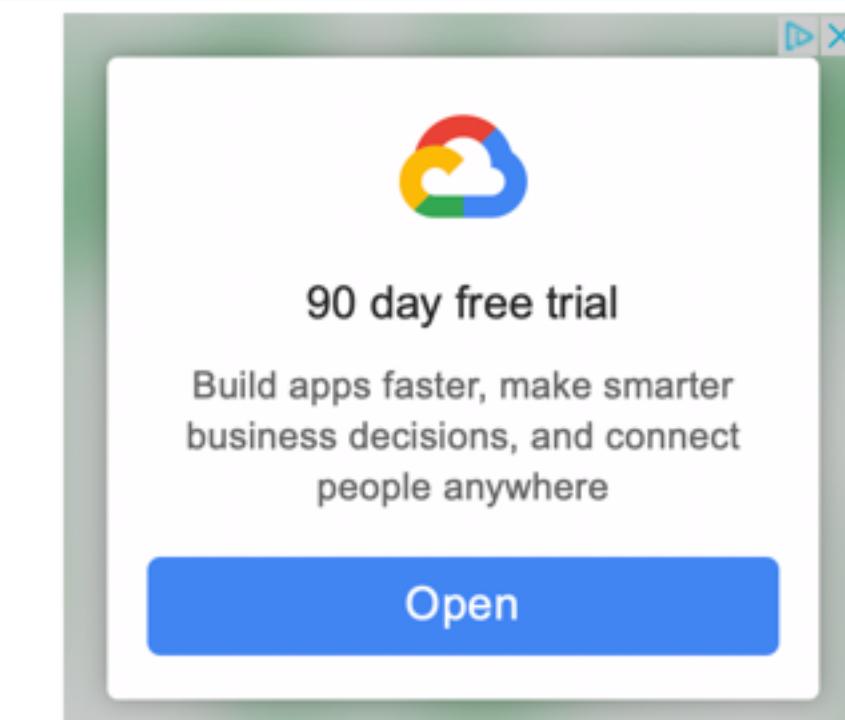
Rich Pelley

Tue 11 Mar 2025 08.00 EDT

Share

‘Can anyone make a Pac-Man clone? Could I make one?’ Illustration: AI-generated

There’s a lot going on with video games and generative AI right now. Microsoft and Google have each created models that can dream up virtual worlds, with significant limitations. And people have been using Grok, the gen-AI chatbot from Elon Musk’s xAI, to make



90 day free trial
Build apps faster, make smarter business decisions, and connect people anywhere

Open

Success theater?

garmarcus.substack.com

Marcus on AI

GenAI in two words: "Success Theater"

GARY MARCUS
FEB 20, 2025

117 37 11 Share

Success Theater

Amazing snippet in Jon Victor's newsletter today at The Information:

It's now common to hear CEOs talk about how they're using artificial intelligence to cut costs. But here and there some are poking holes in their peers' rosy statements.

Airbnb CEO Brian Chesky, for instance, said on the company's earnings call that AI-powered coding assistants haven't led to a "fundamental step change in productivity yet."

"I've talked to a lot of other tech CEOs," Chesky said. "Most of them haven't seen a material change in engineering productivity."

Meanwhile, HSBC's head of generative AI, Edward Achtner, recently described other companies bragging about their use of

AI Adoption in the Workplace Still Low, 2 Years Later

Maria Rosala and Kate Moran
March 7, 2025

Share

Summary: Designers may think AI features are now familiar to our users, but recent research suggests that adoption is still lower than we might think.

Pew Research Center recently published interesting results from a large survey about AI usage in the workplace. In late 2024, they surveyed over 5,000 adults employed in the U.S., with the sample weighted to be representative of the overall U.S. adult population. Here's what the findings mean for designers working on AI features.

In This Article:

- Most US Workers Are Not Using AI
- Younger Workers Use AI More
- Most AI Usage Is Not Sophisticated
- Perception of Usefulness Increases as Usage Increases
- Takeaways for Design Teams

Related Courses

Designing AI Experiences
Strategize, communicate, and design innovative AI products and features
INTERACTION

Practical AI for UX Professionals
Leverage artificial intelligence tools to enhance your UX work and save valuable time
INTERACTION

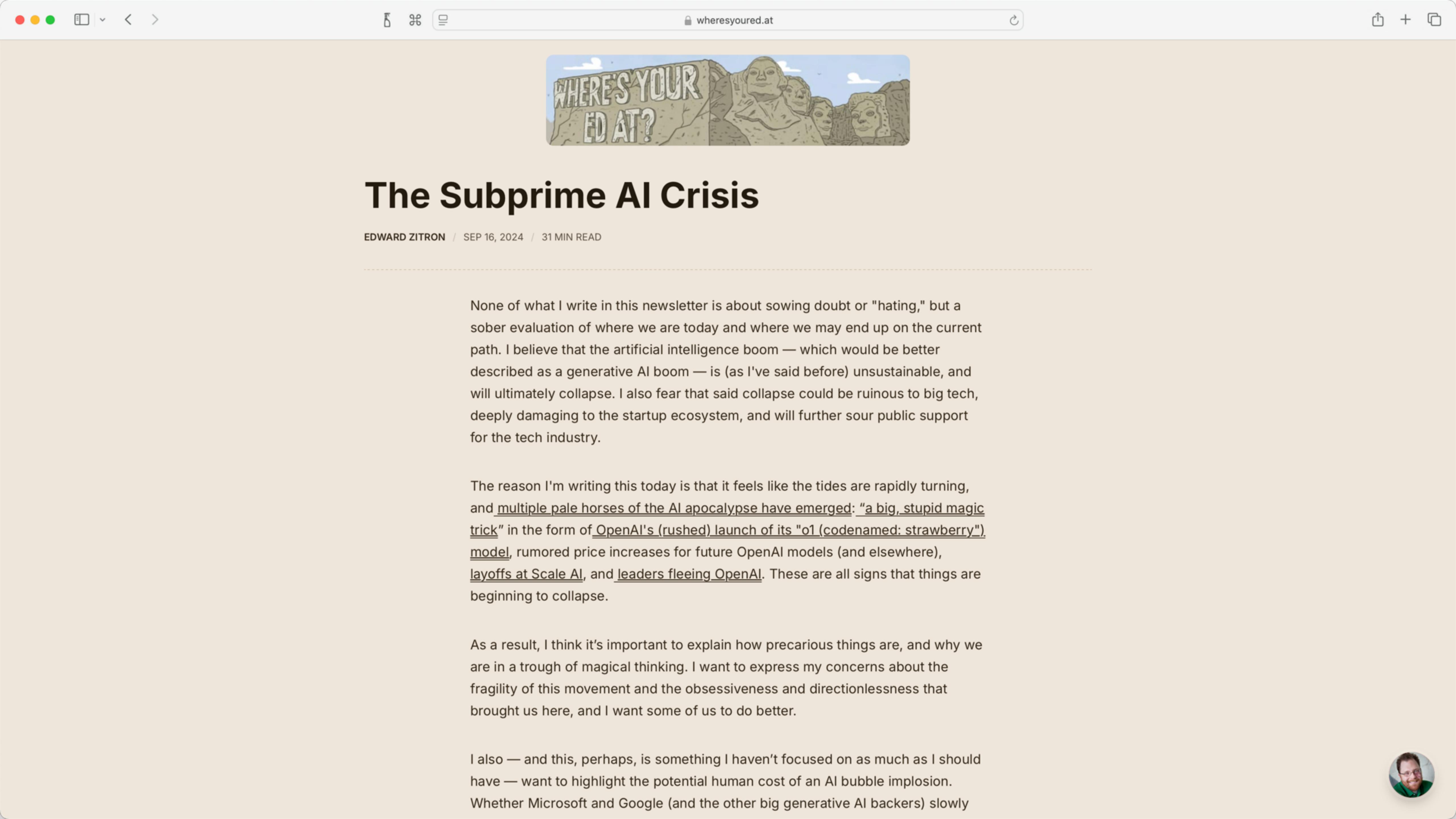
Related Topics

Artificial Intelligence

Don't add AI just to add AI.

Use it where it adds value.

Is AI helping? Or hurting?



The Subprime AI Crisis

EDWARD ZITRON / SEP 16, 2024 / 31 MIN READ

None of what I write in this newsletter is about sowing doubt or "hating," but a sober evaluation of where we are today and where we may end up on the current path. I believe that the artificial intelligence boom — which would be better described as a generative AI boom — is (as I've said before) unsustainable, and will ultimately collapse. I also fear that said collapse could be ruinous to big tech, deeply damaging to the startup ecosystem, and will further sour public support for the tech industry.

The reason I'm writing this today is that it feels like the tides are rapidly turning, and multiple pale horses of the AI apocalypse have emerged: "a big, stupid magic trick" in the form of OpenAI's (rushed) launch of its "o1 (codenamed: strawberry") model, rumored price increases for future OpenAI models (and elsewhere), layoffs at Scale AI, and leaders fleeing OpenAI. These are all signs that things are beginning to collapse.

As a result, I think it's important to explain how precarious things are, and why we are in a trough of magical thinking. I want to express my concerns about the fragility of this movement and the obsessiveness and directionlessness that brought us here, and I want some of us to do better.

I also — and this, perhaps, is something I haven't focused on as much as I should have — want to highlight the potential human cost of an AI bubble implosion. Whether Microsoft and Google (and the other big generative AI backers) slowly



AI isn't going away.

There are some
amazing possibilities!

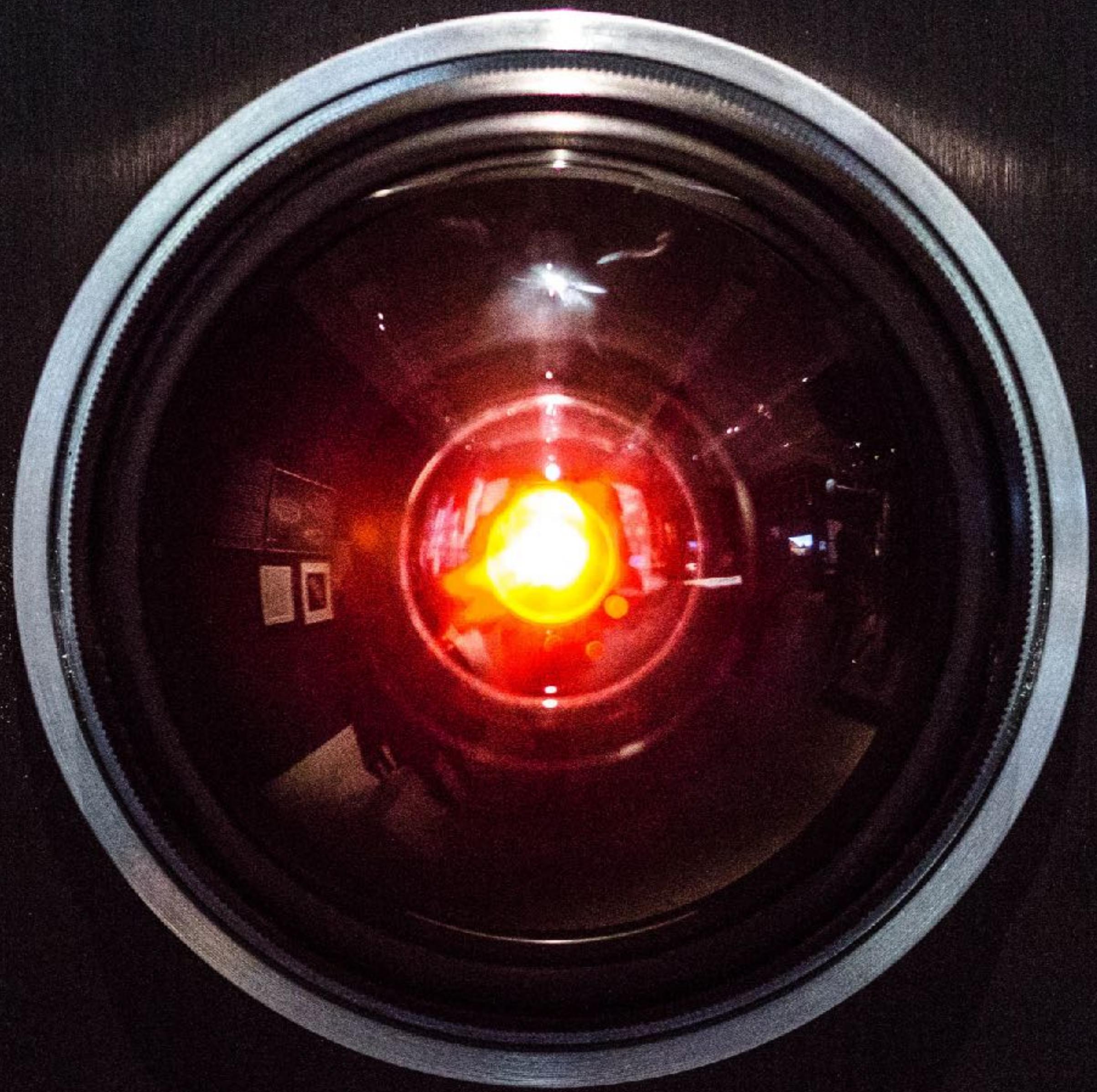
But some serious risks...

You can't afford to ignore AI.

**PROCEED
WITH
CAUTION**

We learned this lesson already.

WOPR anyone?



“I can’t let you do that Glenn.”

Good luck!

<https://github.com/cppwfs/canihaveyourorder>

<https://github.com/cppwfs/menubatch>

O'REILLY®

Fundamentals of Software Engineering

From Coder to Engineer

Early
Release
RAW &
UNEDITED



Nathaniel Schutta
& Dan Vega

[https://www.oreilly.com/library/
view/fundamentals-of-software/
9781098143220/](https://www.oreilly.com/library/view/fundamentals-of-software/9781098143220/)

Thinking Architecturally

Lead Technical Change Within
Your Engineering Team



Nathaniel Schutta

[https://learning.oreilly.com/
library/view/thinking-
architecturally/9781492034421/](https://learning.oreilly.com/library/view/thinking-architecturally/9781492034421/)

O'REILLY®

Compliments of
VMware Tanzu

Responsible Microservices

Where Microservices Deliver Value

Nathaniel Schutta

REPORT

[https://learning.oreilly.com/
library/view/responsible-
microservices/9781492085294/](https://learning.oreilly.com/library/view/responsible-microservices/9781492085294/)



Nathaniel Schutta

🦋 @nts.bsky.social

<https://www.linkedin.com/in/nate-schutta>

<https://ntschutta.io/>



Glenn Renfro

🦋 @cppwfs.bsky.social

<https://www.linkedin.com/in/cppwfs>



provide a title image for a presentation called "Trust but verify" where the talk discusses designing building systems that uses ai elements.



Talked to pyxl.pro ▾

