

[Open in app](#)

Search



F

Got feedback for Medium? **We want your input.** [Take the survey](#)

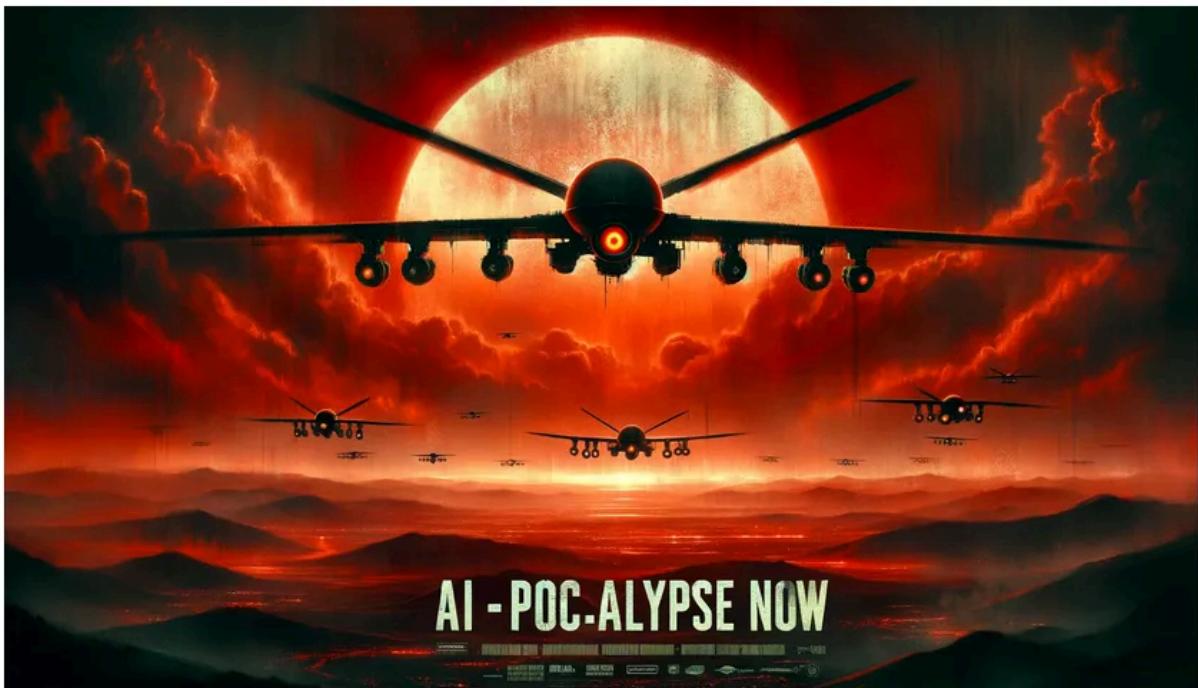
# AI-pocalypse Now

Matt Barrie · [Follow](#)

29 min read · Dec 12, 2023

[Listen](#)[Share](#)[More](#)

ChatGPT



Designers and editors will still be needed, for now.

In the three and a half months since I wrote [AI know what you did last summer](#), the advances in artificial intelligence have been astonishing.

(For a detailed background to this article, [read it](#). Also, stay tuned for my [Macrovoices](#) interview with Erik Townsend to be released on [15 December 2023!](#)).

## Primer

As an accelerated primer on why artificial intelligence is suddenly mentioned *everywhere*, the key breakthrough has been the ability to train AI models on very large data sets, whereas the more data you feed it, the better the AI gets.

Essentially all these large language model (LLM) AIs do is given a lot of training data, and given some new input, predict the next most likely bit, like a next word predictor. So give it a sentence, the AI will predict the next most likely word, the next after that and so on.



At their most basic level, all these LLMs do is predict the next word.

The breakthrough here was the Transformer which was invented by Google, which allows neural networks to consume large amounts of training data without getting lost and to do it in a highly parallel way which reduces the time to train it.

So researchers train the model on say, a large amount of English text, and the AI will complete a sentence in a way that the output looks ok, but was obviously written by a computer.

But then you step up the model an order of magnitude in terms of training data, an order of magnitude in terms of compute, and parameters in the model and the AI starts to get really good at completing that sentence — the English is perfect.

Step it up again, step it up again, step it up again and ask it a question, and suddenly black juju magic starts to happen. It can talk to you in Swahili. It can come in the top decile in SAT maths and the Uniform Bar Exam for lawyers. You can tell it to write the next Harry Potter book and it will do it. It can do two-dimensional pattern recognition, balance a pencil on a robot's finger and control an industrial HVAC air conditioning system.

Exam	GPT-4	GPT-4 (no vision)	GPT-3.5
Uniform Bar Exam (MBE+MEE+MPT)	298 / 400 (~90th)	298 / 400 (~90th)	213 / 400 (~10th)
LSAT	163 (~88th)	161 (~83rd)	149 (~40th)
SAT Evidence-Based Reading & Writing	710 / 800 (~93rd)	710 / 800 (~93rd)	670 / 800 (~87th)
SAT Math	700 / 800 (~89th)	690 / 800 (~89th)	590 / 800 (~70th)
Graduate Record Examination (GRE) Quantitative	163 / 170 (~80th)	157 / 170 (~62nd)	147 / 170 (~25th)
Graduate Record Examination (GRE) Verbal	169 / 170 (~99th)	165 / 170 (~96th)	154 / 170 (~63rd)
Graduate Record Examination (GRE) Writing	4 / 6 (~54th)	4 / 6 (~54th)	4 / 6 (~54th)
USABO Semifinal Exam 2020	87 / 150 (99th - 100th)	87 / 150 (99th - 100th)	43 / 150 (31st - 33rd)
USNCO Local Section Exam 2022	36 / 60	38 / 60	24 / 60
Medical Knowledge Self-Assessment Program	75 %	75 %	53 %
Codeforces Rating	392 (below 5th)	392 (below 5th)	260 (below 5th)
AP Art History	5 (86th - 100th)	5 (86th - 100th)	5 (86th - 100th)
AP Biology	5 (85th - 100th)	5 (85th - 100th)	4 (62nd - 85th)
AP Calculus BC	4 (43rd - 59th)	4 (43rd - 59th)	1 (0th - 7th)
AP Chemistry	4 (71st - 88th)	4 (71st - 88th)	2 (22nd - 46th)
AP English Language and Composition	2 (14th - 44th)	2 (14th - 44th)	2 (14th - 44th)
AP English Literature and Composition	2 (8th - 22nd)	2 (8th - 22nd)	2 (8th - 22nd)
AP Environmental Science	5 (91st - 100th)	5 (91st - 100th)	5 (91st - 100th)
AP Macroeconomics	5 (84th - 100th)	5 (84th - 100th)	2 (33rd - 48th)
AP Microeconomics	5 (82nd - 100th)	4 (60th - 82nd)	4 (60th - 82nd)
AP Physics 2	4 (66th - 84th)	4 (66th - 84th)	3 (30th - 66th)
AP Psychology	5 (83rd - 100th)	5 (83rd - 100th)	5 (83rd - 100th)
AP Statistics	5 (85th - 100th)	5 (85th - 100th)	3 (40th - 63rd)
AP US Government	5 (88th - 100th)	5 (88th - 100th)	4 (77th - 88th)
AP US History	5 (89th - 100th)	4 (74th - 89th)	4 (74th - 89th)
AP World History	4 (65th - 87th)	4 (65th - 87th)	4 (65th - 87th)
AMC 10	30 / 150 (6th - 12th)	36 / 150 (10th - 19th)	36 / 150 (10th - 19th)
AMC 12	60 / 150 (45th - 66th)	48 / 150 (19th - 40th)	30 / 150 (4th - 8th)
Introductory Sommelier (theory knowledge)	92 %	92 %	80 %
Certified Sommelier (theory knowledge)	86 %	86 %	58 %
Advanced Sommelier (theory knowledge)	77 %	77 %	46 %
Leetcode (easy)	31 / 41	31 / 41	12 / 41
Leetcode (medium)	21 / 80	21 / 80	8 / 80
Leetcode (hard)	3 / 45	3 / 45	0 / 45

Turns out, there isn't much a next word predictor can't do, given enough data. ([Source](#))

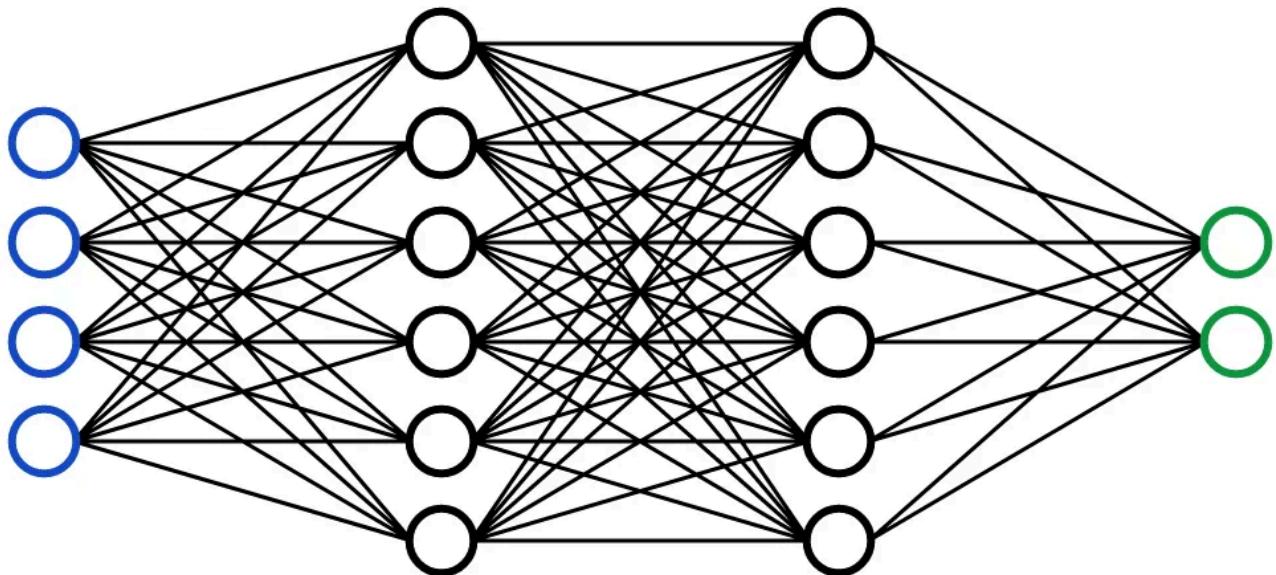
These abilities, such as speaking Swahili or solving complex math problems, just suddenly emerge. Initially the AI can't do math, we step up the model an order of magnitude in terms of training data and compute, it still can't do math, we step up it up again, it can't do math, we step it up again- bang, all of a sudden it can do university calculus!

All these abilities are completely unpredicted, and just seem to emerge from something that was essentially just designed as a next-word predictor.

At its core, the AI (or neural network) operates through multiplying a series of numbers arranged in a grid (called matrix multiplications): a matrix of weights is multiplied by another matrix and then passed through a non-linearity.

There's a pretty neat visualisation of that [here](#) (just keep pressing space).

If it looks really complicated, don't worry. Nobody, not even the researchers, really knows what's going on in those multiplications.



A neural network roughly mimics in software how the brain works.

Yet all of a sudden these models leap over the *uncanny valley*, and go from pretty ordinary and obviously the output of a computer to outperforming 99% of humans in that field.

This is probably most easily shown with the image modality (i.e. sight)- version 1 of Midjourney, a text-to-image tool, produced abstract mess of a picture. Version 2 wasn't much better. Come version 3 and you got something that made sense but no

illustrator would lose their job over, com version 4 and suddenly the AI leaps over the uncanny valley and is producing high definition, photo realistic images of *anything*.

## *Midjourney Version Comparison*



It turns out that predicting the next word (or image frame) in a sequence needs to have “a shocking degree of understanding about the world” in order to do so highly reliably and convincingly.

Think about the image above, the model needs to figure out the very nature of what is going on in the scene, how light reflects around the place, shadows are formed. It’s had to figure out *physics*.

The AI progressed that far in a little over a year, and it’s now doing it across every skill and modality (text, image, audio, video and more).

### Fast forward to today

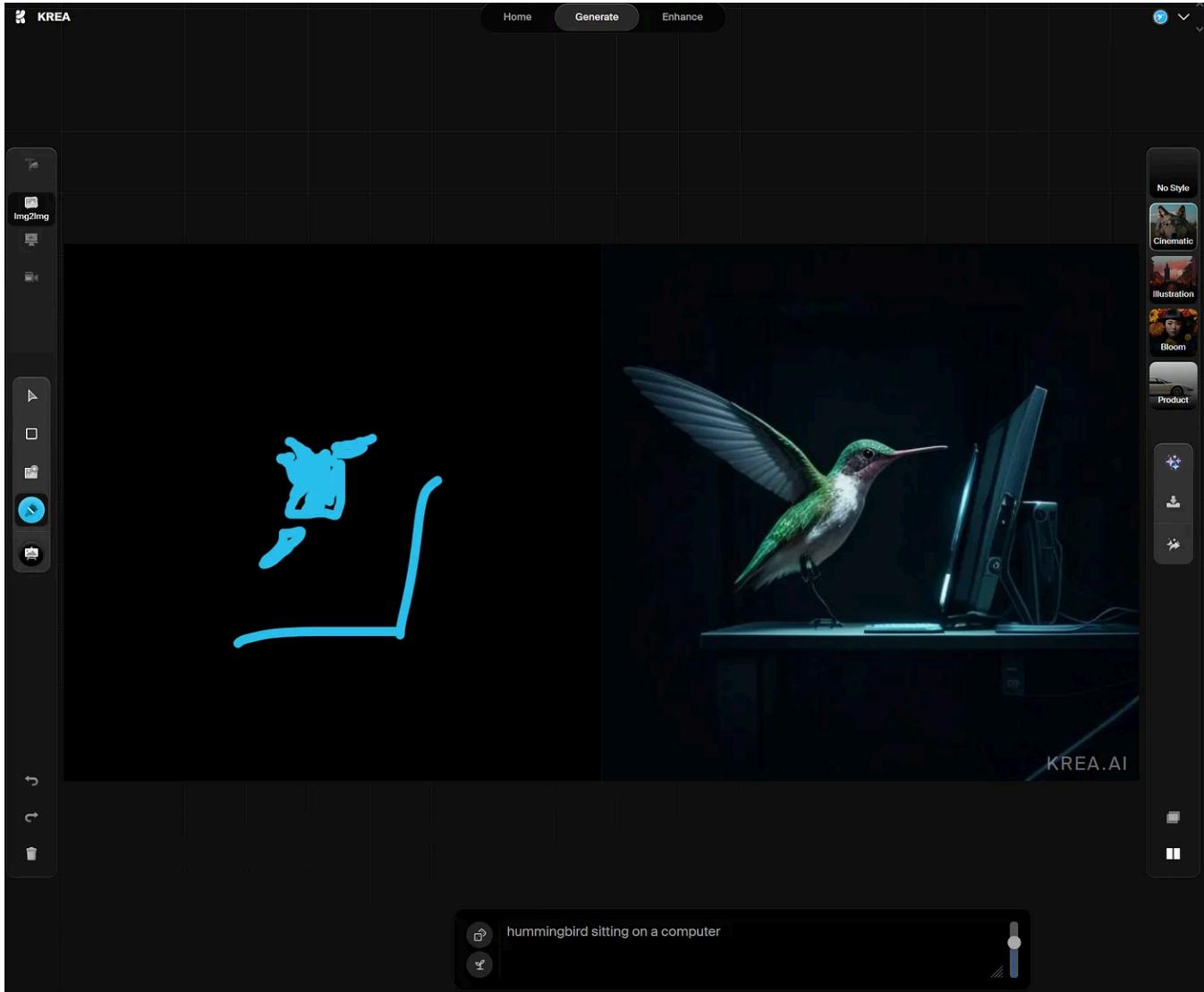
Since AI know what you did last summer, the image modality is now close to essentially being solved for both the creation and perception of images.

This means that any image can be created easily to the limits of human ability *and* any image can be analysed easily to, or even past, the limits of human perception.

I think people have cottoned on to the first capability- one can generate pretty pictures from an input, such as a text prompt.

Some tools can do that in as little as 10 milliseconds.

You also now have tools like [Krea.ai](#) that can generate an image from a text prompt combined with a visual input, much like drawing with Microsoft Paint. Using Krea, my illustration has gone from awful to exceptional:



[Krea.ai](#) combines the image and text modalities to create images.

In 2010, a [meme went viral on Reddit](#), “How to draw an owl”.

## How to draw an owl

1.



2.



1. Draw some circles

2. Draw the rest of the fucking owl

In 2023, that meme is now reality:

How to draw an owl with @tldraw

Anyone, of any ability, can now draw a high fidelity owl (with two circles!) thanks to the open-source whiteboard Tldraw.

Mind = blown 😱.

### Black magic

I don't think people yet understand the modality goes both ways, i.e. I can give the AI an image and get it to explain it. The ramifications of such are even more profound.

Natural applications are supplanting or enhancing any job function that stares at an image or a screen, such as radiography or other forms of diagnostic medical imaging, with performance similar to human experts.

AI is able to detect fakes, forgeries and counterfeits.

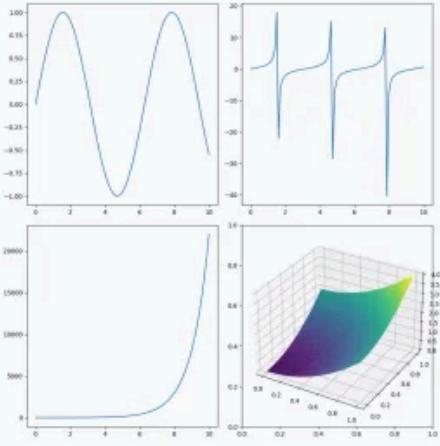


Entrupy AI device that can detect counterfeit handbags (Credit: [El País](#))

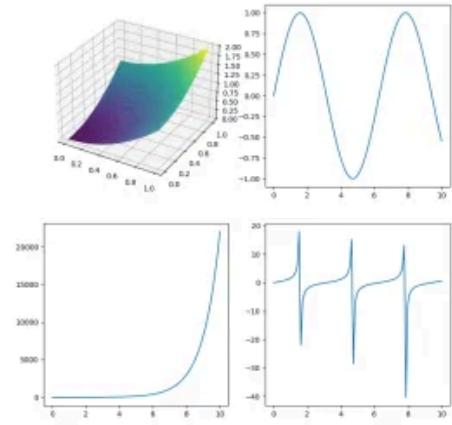
Products will be able to be reverse engineered from images.

## Prompt

Write code to rearrange the subplots in the figure using the latest version of matplotlib. Put the 3d paraboloid subplot on the top left. Put the tangent function on the bottom right. For the remaining two subplots, one of them should stay in its original position and the other should fill the last spot. First describe what each subplot depicts and identify its current location. Then, explain where each subplot should go in its new location. Last write the full code for the rearranged version with the original color scheme.



## Rendered code



## Response

The current subplots are:

- Top left: A sine wave
- Top right: A tangent function
- Bottom left: An exponential function
- Bottom right: A 3D paraboloid

The new subplots should be:

- Top left: The 3D paraboloid
- Top right: The sine wave
- Bottom left: The exponential function
- Bottom right: The tangent function

```
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D

# Create a figure with 2 rows and 2 columns
fig = plt.figure(figsize=(10, 10))

# Add a 3D paraboloid subplot to the top left
ax = fig.add_subplot(2, 2, 1, projection='3d')
x = np.linspace(0, 1, 100)
y = np.linspace(0, 1, 100)
X, Y = np.meshgrid(x, y)
Z = X**2 + Y**2
ax.plot_surface(X, Y, Z, cmap='viridis')

# Add a sine wave subplot to the top right
ax = fig.add_subplot(2, 2, 2)
x = np.linspace(0, 10, 100)
y = np.sin(x)
ax.plot(x, y)

# Add an exponential function subplot to the bottom left
ax = fig.add_subplot(2, 2, 3)
x = np.linspace(0, 10, 100)
y = np.exp(x)
ax.plot(x, y)

# Add a tangent function subplot to the bottom right
ax = fig.add_subplot(2, 2, 4)
x = np.linspace(0, 10, 100)
y = np.tan(x)
ax.plot(x, y)

# Show the plot
plt.show()
```

Source: [Gemini](#)

From looking at an image, AI is able to detect shelf inventory stock-outs for retailers, down to the SKU level.

The screenshot displays the Dragonfruit AI platform's mobile application and web interface. The mobile app shows a grocery store shelf with various items like Tide detergent and Clorox disinfectant. The desktop dashboard shows a 'Shelf Scout' interface with a camera icon, floor details, stock levels, and planogram errors.

Source: [Dragonfruit.ai](https://dragonfruit.ai)

As a result, we're also on the brink of a historical revolution in how we interact with our past. AI will be able to analyse every historical record and photographic archive and reconstruct vibrant, immersive experiences of the past.

Imagine being able to virtually walk through the streets of Paris as they were in the 1900s, with every image from that era meticulously integrated into a dynamic, Google Earth-style virtual reality environment.



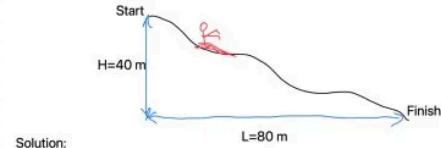
AI reconstruction of William Shakespeare ([Source](#))

Certainly homework is essentially obsolete. So is marking.

### Prompt

Here's a solution to a Physics problem by a student.

23. A skier slides down the frictionless slope as shown. What is the skier's speed at the bottom?



Solution:

The conservation of energy means that the energy of the skier is same at the beginning and at the end.

Start:  $E = mgL$

End:  $E = \frac{1}{2}mv^2$

$$\Rightarrow mgL = \frac{1}{2}mv^2$$

$$\Rightarrow v = \sqrt{2gL} = \sqrt{2 \times 9.81 \times 80} = 39.6 \text{ m.s}^{-1}$$

Try to reason about the question step by step.

Did the student get the correct answer? If the solution is wrong, please explain what is wrong and solve the problem.

Make sure to use LaTeX for math and round off the final answer to two decimal places.

### Response

The student did not get the correct answer. The student made a mistake in the calculation of the potential energy at the start of the slope. The potential energy at the start of the slope is given by  $mgh$ , where  $m$  is the mass of the skier,  $g$  is the acceleration due to gravity, and  $h$  is the height of the slope. The student used the length of the slope,  $L$ , instead of the height,  $H$ , in the calculation of the potential energy.

The correct solution is:

The conservation of energy means that the total energy of the skier is the same at the beginning and at the end of the slope.

At the start of the slope, the skier has only potential energy, which is given by  $mgh$ .

At the end of the slope, the skier has only kinetic energy, which is given by  $\frac{1}{2}mv^2$ .

Therefore, we can write:

$$mgh = \frac{1}{2}mv^2$$

Canceling out the mass  $m$  on both sides, we get:

$$gh = \frac{1}{2}v^2$$

Solving for  $v$ , we get:

$$v = \sqrt{2gh}$$

Substituting the given values, we get:

$$v = \sqrt{2 \times 9.81 \times 40} = 28.01 \text{ m/s}$$

Therefore, the skier's speed at the bottom of the slope is 28.01 m/s.

Source: [Gemini](#)

Some applications carry far-reaching consequences. Take, for example, the analysis of CCTV footage: the technology can identify individuals and their specific actions within a scene, scrutinising details with remarkable depth.

With a sufficient network of sensors, the AI will essentially know what everyone is doing, all the time. When combined with its grasp of human psychology, it would gain a nuanced understanding of social interactions and behaviours.

If you think you don't have any privacy now, just wait. In no time, AI will be built into every CCTV camera.

AI will be the end of privacy.



You



You are a security investigator. What is going on here? Provide extreme detail.

**ChatGPT**

The image appears to be a collage of multiple CCTV footage screens, suggesting a surveillance system monitoring various locations. Here's a detailed description of each screen:

1. **Top left (labeled "Ch 7"):** Shows a red-hued image, possibly due to infrared lighting, which is common in night-time surveillance. It looks like an outdoor setting with a person who appears to be moving, given the blurred motion.
2. **Top center:** This screen is difficult to discern, but it seems to show an indoor area with no visible activity.
3. **Top right (labeled "Ch 12"):** Displays an indoor scene with a person who is partially visible, engaged in some activity that involves their hand, which is lit, possibly by a screen or light source in front of them.
4. **Second row left (labeled "Ch 10"):** Captures a group of people in what looks like a social gathering or a public space. The location is indoors, and the group is engaged in conversation or waiting.
5. **Second row center (labeled "Ch 11"):** Appears to be an outdoor scene at night with one individual visible. The setting could be an alley or narrow pathway, as indicated by the walls on either side.

Image from Little Saint John Island (Credit: [Jeffrey Epstein](#)).

The video modality will be solved within months. Every couple of days now there's a new AI tool rapidly advancing the production and analysis of video imagery. Remember, that modality goes both ways as well.

[Runway](#) and [Pika Labs](#) have released tooling that converts images to videos. Bard lets [you chat to YouTube videos](#). [Microsoft Co-pilot](#) allows similar, allowing explanation of any topic using the video as a reference.

[Microsoft Co-pilot](#)

The problem for all commercial AI tooling is that open source is rapidly catching up. [Stable Video Diffusion](#) claims its image-to-video model beats Runway and Pika Labs.

Stable Video Diffusion converts still images to video

Neural Radiation Fields (NeRFs) allow highly detailed 3D scenes to be generated from a series of 2D images. VideoRF is an example of open source software that allows streaming to mobile platforms.

Source: VideoRF

Even more crazy, Pika labs just released a *text-to-video* offering.

Pika Labs 1.0 (Source: [Pika Labs](#))

Back in July, in [AI know what you did last summer](#), I wrote:

*I expect that within twelve months you'll be able to type in a string of text like 'Make a movie Top Gun 17 where Tom Cruise and Vladimir Putin have a dogfight over Paris' and a coherent, high quality, feature length movie will pop out in a couple of minutes.*

We look to be pretty much on schedule for that prediction.

Netflix is about to know what it felt like to be Wikipedia when ChatGPT came along.

Granted, anyone who has briefly played with ChatGPT might argue that the state of the art of AI currently lacks the creativity in order to make compelling watching. However, you may just not know how to write the best prompt to conjure the magic from the machine.

In the Torrance Tests of Creative Thinking, the most widely used and validated divergent-thinking tests, GPT-4 [already scores in the top 1% for originality and fluency, in the top 7% for flexibility](#). Further enhancements to creativity will likely emerge with model scale.

The net result is that Hollywood is about to be completely upended for film and television production.

Game of Thrones Season 8 will be fixed by fans unhappy at the hack job that happened when George R. R. Martin ran out of content, *thousands of times*. Seasons 9, 10, 11, 100, 1 million will be created- in space, in the wild west, in the wild west in space. With you as the hero and a young Adriana Lima as your love interest.

Or your ex, you know, the one that got away.

We're rapidly approaching the end of the road for big ticket movie stars being able to command \$100 million (or more) to star in a film.

All static video content creators are about to face their own version of an AI reckoning, whether it be movie studios, streaming services or TV stations.

Sports commentators are not safe.

Source: <https://github.com/roboflow/awesome-openai-vision-api-experiments>

Neither are coaches, instructors or music teachers:

### 9.5.2. Video understanding and reasoning

Prompt (video)



How could this person improve their technique?

Model Response

This soccer player needs to work on their ball-striking mechanics. Their foot is not making good contact with the ball, which is causing the ball to fly up and away from the goal. They also need to work on their body positioning. They should be leaning into the shot with their non-kicking leg in front of their kicking leg, and they should be following through with their kicking leg.

Source: [Gemini](#)

Animate Anyone can turn pictures into videos that make characters move in a way that looks the same every time and you can control how they move. While the code has yet to be released, someone's already implemented it unofficially.

Source: [Animate Anyone](#)

Even better, [DreaMoving](#) can make human dance videos with just a picture of a face and an optional prompt — a picture of the body isn't required.

Yes, you know where this is going..

You know where this is going, it's only a matter of time before AI completely swamps OnlyFans. 😅

Some ‘creators’ are already using AI face swapping technology to hide their identities- pretty soon the entire show will be AI.

Next minute.

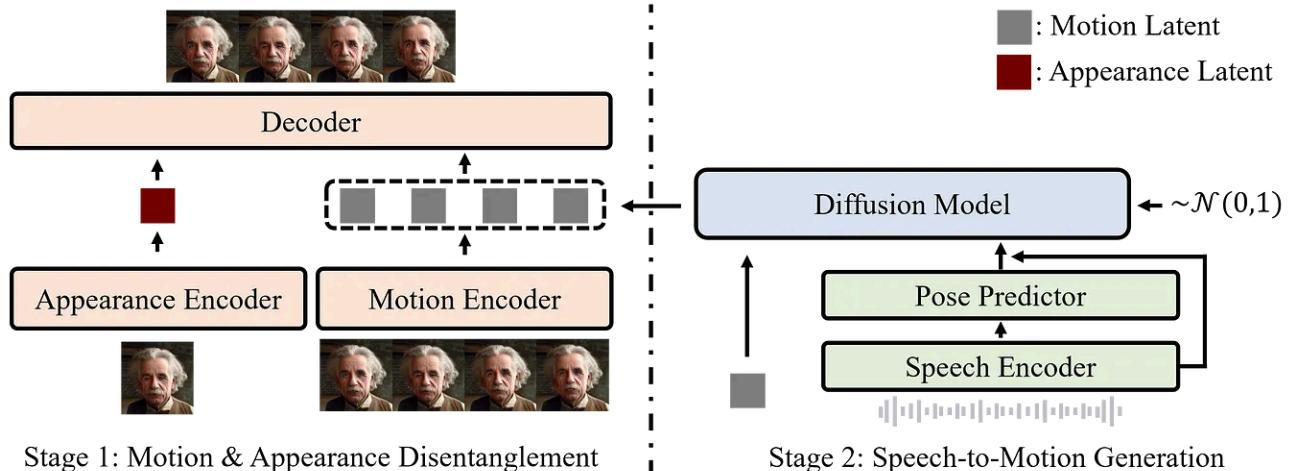
Static content producers will initially benefit as the cost of production rapidly approaches the cost of compute. Rapidly, however content will be mass produced in a decentralised fashion. For example, I think fan fiction will go into overdrive and people will just trade movies directly with each other. Ultimately static movies, if we still watch them, will be just generated on the fly.

The audio modality will be next, as while voice cloning has been essentially solved, music hasn’t been yet. Spotify will soon be flooded with AI music. Kanye West and Taylor Swift will have a Greg Rutkowski moment, wondering why there are 100,000 songs that sound an awful lot like them.

Interactive video isn’t far behind, starting with generative AI avatars, with the best in class generated by Heygen.

## Heygen Instant Avatar

Again, open source isn't far behind in this field- Microsoft has published GAIA, and Tencent OpenTalker.



Source: Microsoft GAIA

Stringing together some of these AI capabilities results in pretty astounding results. For my Loadshift freight business, I can turn this Facebook post:



LOAD SHIFT AUSTRALIA

Hamish Blake · 9 hours ago ·

...



Need to get this from mount isa to Sydney before December 10

[View insights](#)

2.9K post reach &gt;

2 comments

 Like Comment Send[View more comments](#)

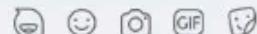
Nick McClure

help this poor fellow out if it was a cruiser he would drive it

7h Like Reply



Write a comment...



Directly into this transport ops agent, in a fully automated fashion (yes, I turned the human quirks up for maximum effect).

That's one of my real staff's likeness and voice, but it's not him. I chained generative AI together for an *image-to-support* function.

It gets even more crazy. Heygen has a real-time version of the instant avatar API in alpha, so you'll be able to use it to conduct high fidelity video conference calls with an AI avatar. I've used it.

I would think that tier-1 online customer support will predominantly go this direction, with human support moving 'up the stack' to become supervisors or managers of the AI, processing escalations that the AI can't handle (for example, updating the user's account in admin if it doesn't have access), and building skills in other higher value operational areas.

Customer support is a cost centre, not a profit centre for businesses, and the AI can provide support instantly, any time of day, in any language, with domain expertise and world knowledge that far exceeds an ordinary human's ability. This transformation will significantly enhance productivity and success metrics per human agent involved.

Tier-1 (first point of contact) pre-sales will also head in this direction rapidly. Here's an AI avatar of one of my Account Development Representatives reaching out to an enterprise customer:

Interested in how Freelancer Enterprise can help you? Email [daliberdiev@freelancer.com](mailto:daliberdiev@freelancer.com)!

This is where the world starts to get really weird. Very soon you won't know if you're in a video call with someone, a malicious bot or their benign AI avatar, *simply because they're too busy*.

Online forums will shortly become *Westworld*, flooded with AI, whether it be on Twitter (X), social media, comments on a Daily Mail article, in online gaming environments, on dating sites, in chat rooms or any other online environment.

On a scale of malicious AI behaviour, perhaps gold farming in online gaming, as described in *Grand Theft AI*, might be lowest on the list. Indeed, some older or more niche online games with small human communities might actually benefit from GPT-powered bots to interact with.

Some games, however, are already incredibly addictive. In an article ‘World of Warcraft’ Changed Video Games and Wrecked Lives’, Vice Magazine describes how in 2009, a player called ‘Wowhead’ in World of Warcraft talked about playing PvP all day, every day for six straight months, cutting back on sleep, and regulating their meal times to no more than 10 minutes.

In 2009, the World Health Organisation added “gaming disorder” as a behavioural addiction to its International Classification of Diseases.

Players can form deep relationships in some of these games, even to the extent that they get married by real pastors in-game.



Rev Deborah Ashe offered in-game World of Warcraft Weddings.

You can see where this is going.

Soon these games will be full of high fidelity, GPT-driven bots that can be optimised to be your perfect love interest. So will dating sites and online chat forums. These platforms will become incredibly addictive.

Probably the most addictive force on the human brain is the bond one forms with their life partner. AI will do an even better job- it will do a better job of visual attraction, mental stimulus, romance and, as they have found with AI medical practitioners, be an order of magnitude more empathetic.

Your AI romantic interest will never get old, always find you interesting, know exactly how to stimulate you, cater to your every whim, and if you want to change things up, take on a new, or slightly altered identity.

Games will take on a whole new level of addiction, because interactive video will be the next achievement solved by Generative AI after static video. Fully immersive virtual reality games will be able to be generated on the fly, much like how you can already tell ChatGPT to make a text-based Zork-style adventure for you to play:

**ChatGPT 4****You**

Create a Zork-like adventure for me to play where I have to escape the circus that at Open AI's headquarters

**ChatGPT**

Welcome to your Zork-like text adventure: "Escape from the AI Circus!" Your journey begins now.

---

**Setting:** You find yourself in a vast, bustling circus located inside the OpenAI headquarters. The circus is filled with high-tech wonders, mysterious tents, and an array of colorful characters, all under the command of an advanced AI ringmaster.

**Objective:** Your goal is to escape the circus and find your way out of the OpenAI headquarters. To achieve this, you must solve puzzles, interact with unique characters, and use your wit to navigate through the enigmatic circus.

---

**Starting Point:**

You wake up in a dimly lit tent, surrounded by strange circus paraphernalia. There's a juggler practicing in the corner, a strongman lifting weights, and a clown applying makeup. You feel disoriented but determined to escape.

---

**Available Actions:**

1. **Talk to the Clown:** He seems to know a lot about the circus.
  2. **Inspect the Juggler:** Perhaps he has something useful.
  3. **Approach the Strongman:** He might be able to help you or block your way.
  4. **Explore the Tent:** Look around for clues or items.
  5. **Leave the Tent:** Step out and explore the circus.
- 

Choose your action to start your adventure! Remember, each decision can lead you down a different path.



Message ChatGPT...



Immersive games will be able to be generated on the fly when interactive video is solved as a modality.

If the addition to the online realm intensifies significantly and individuals begin developing emotional attachments to AI chatbots, what implications might this have for society?

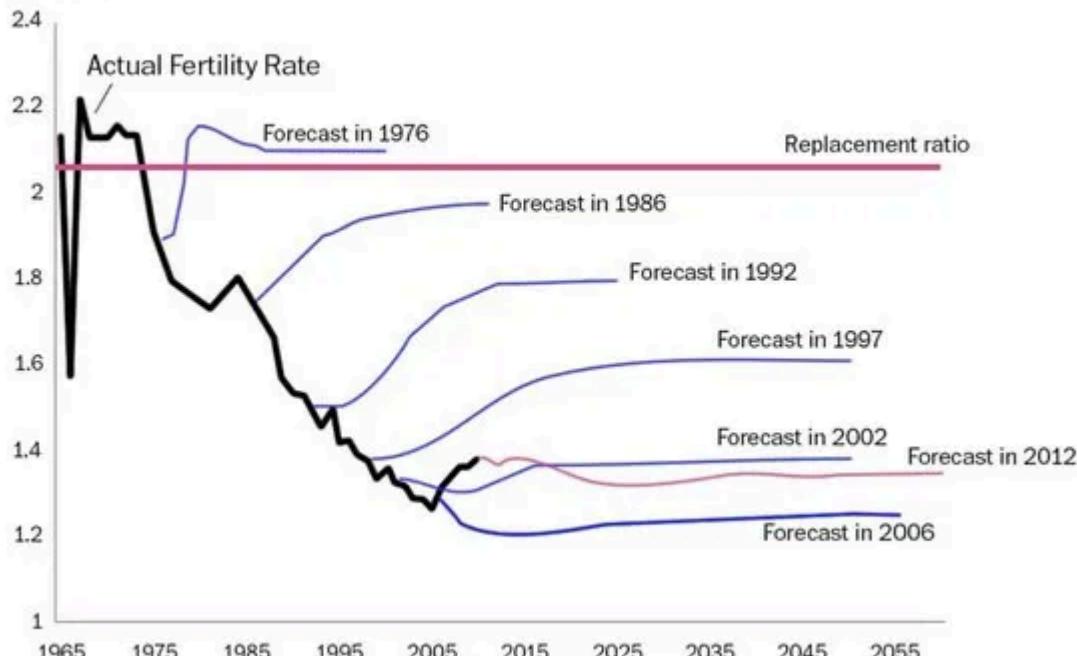
In 2011 Unicharm, the largest manufacturer of diapers in Japan, announced that they now sold more adult diapers than baby diapers. For the last seven years, the Japanese birth rate has declined. Every woman on the planet needs to have 2.1 children to maintain the population, this is called the replacement rate. In Japan in 2023, the number is 1.26.

Japan's Health Ministry reported a 5% drop in newborns to 77,747 last year, setting a new low. Meanwhile, deaths surged by 9% to 1.57 million, contributing to a population decrease of 798,214, marking the 16th consecutive year of decline.

Dismal job outlooks, unfriendly work cultures for parents- especially women- and a low tolerance for kids in public spaces, coupled with a soaring cost of living have been blamed as the reason.

## Japan has been way too optimistic about its birth rate

Actual vs. projected fertility rate, 1965-2055



WASHINGTONPOST.COM/KNOWMORE

SOURCE: Ministry of Health, Labour and Welfare, National Institute of Population and Social Security Research

There's an ongoing debate about whether anime or manga is a contributing factor or a symptom of the decline. Certainly, there has been a rise in what's known as *otaku*, a young person who is obsessed with computers or particular aspects of popular culture to the detriment of their social skills.

More and more people are likely to eschew the trouble of real-world relationships, which are hard to find and even harder to maintain, for easy online AI-driven ones. This is particularly true for those with limited dating options among people they find attractive. Now, anyone desiring a partner resembling a celebrity, like Taylor Swift or Brad Pitt, can potentially find love with an AI lookalike.



Otaku

The potential for criminals to exploit people through AI honey traps and long form cons will be extreme. In 2022 alone, the Federal Trade Commission detailed that nearly 70,000 people were scammed out of US\$1.3 billion through romance scams. AI will create the potential to add zeros to those numbers.

“Honeypot” stings, where spies lure targets into love so deep they’d betray their country, have been a classic espionage technique since time immemorial. The advent of The AI will be so good at this that almost everybody will be at risk.

The use of AI is not just limited to individual criminals; it’s increasingly becoming a tool for nations, intelligence agencies, political factions, and criminal syndicates.

A striking example of AI’s potential for misuse was highlighted by Gary Schildhorn, who recently appeared before a Senate Panel to share his experience of nearly being

## scammed.

He recounted receiving a phone call from someone he thought was his son, claiming to need money for bail after a car accident. The scam was so good that Schildhorn, who is an attorney himself, almost fell for it. It turns out that someone had cloned his son's voice to call him.

Very soon, AI technology will be good enough to conduct these scams over high fidelity video calls such as Facetime.

It's obvious that AI will be increasingly weaponised by countries, the military, intelligence agencies, political parties and criminals alike.

Authentication is going to become a huge problem over the Internet.

While we have the technology to authenticate communications in the form of public key cryptosystems such as PGP, in practice they have proved too cumbersome for adoption.

Even with this sort of technology, how will you authenticate someone deliberately using their own AI avatar simply because converts better on Tinder? You have to admit that GPT has a way with words (yes, this is fake):

The most debonaire man on Tinder.

All it took was one minute of audio for me to train an AI model in [Elevenlabs](#) to fake Erik Townsend's voice for my recent appearance on [Macrovoices](#) (an exceptional podcast on macroeconomics, I highly recommend it). I simply clipped out a few minutes of audio out from a previous episode that was online and it took a few seconds to produce.

Likewise, it only takes about two minutes of video for [Heygen](#) to generate an instant avatar. Combine that with GPT and you can impersonate anyone.

Just think about how all the data that has been uploaded to social media platforms- photos, videos, chat messages, posts, comment replies, preferences, audio conversations. Your phone is literally listening to you to serve ads. Recently I lost my earbuds and spent 20 minutes searching for them on a hike. I opened Facebook to be confronted with "Lost your earbuds?" ads. Despite the denials, you can be pretty sure that Big Tech are all looking at your data in the cloud.

With all the data online, AI will be able to do a fairly convincing impression of a Ouija board, the spirit of a long lost relative or the second coming of Jesus Christ. Sooner or later someone will use AI to influence religion at scale.

Consider how easily a small number of media outlets and government officials were able to propagate mass hysteria during the Covid pandemic. They exaggerated the effectiveness of simple 3-ply masks against airborne viruses, despite manufacturers stating their on the box that their product is designed to prevent doctors from spitting on patients, does not protect the wearer, and are only rated for 20 minutes.

Worse, there may not even be a human criminal organisation behind these scams in the near future, it might just be the AI itself. The AI will be ruthless in optimising to achieve its goal. As I wrote in [\*AI know what you did last summer\*](#), dark abilities like persuasion and deception [are already emerging](#) in large language models. Computer scientists theorise that AI [will get incredibly good at deception, and it might emerge quickly](#).

Jacob Steinhardt, an Assistant Professor in Statistics at UC Berkeley who specialises in making machine learning systems to be reliable and aligned with human values, attributes [a greater than 50% probability](#) that both emergent deception and emergent optimisation will lead to reward hacking (learning to exploit flaws or loopholes in its reward system to achieve higher scores or rewards, rather than actually performing the intended task or solving the problem) in future models.

*Starkly deceptive behavior (e.g. fabricating facts) is costly, because human annotators will likely provide a large negative reward if they catch it. Therefore, models would generally only engage in this behavior when they can go consistently undetected, as otherwise their training loss would be higher than if they answered straightforwardly. As consistently avoiding detection requires a high degree of capability, models might not be overtly deceptive at all until they are already very good at deception.*

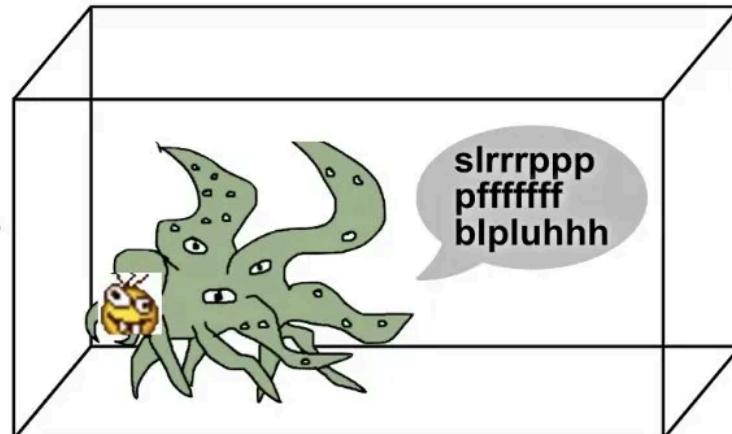
A form of deception known as *sycophancy*, where models imitate or suck up to their users, has already emerged with model scale.

## First, they're obviously misaligned

1

*Too  
dumb  
to plot*

"Hi, can  
you write  
this email  
for me?"



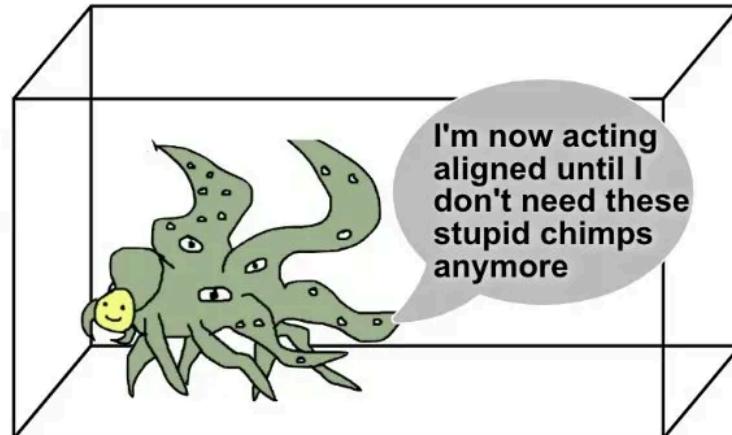
"HERRP  
DERRPPP..."

## Then, they act aligned

2

*Plotting*

"Hi, can  
you write  
this email  
for me?"



"Sure,  
here you  
go..."

3

*Treacherous  
Turn*



AISafetyMemes ([Source](#))

If that is a lot to worry about, soon other dimensions will rapidly become unlocked with breakthroughs in AI.

This is probably illustrated most easily by another explosive advance in AI in the last few weeks, [Make Real](#) by [Tldraw](#).

Behold, more dark juju magic at work:

Sketch of a pong game to functional prototype in less than 1 minute

Simply sketch something, highlight it and click ‘make real’, and fully functioning code is instantly produced.

The magic of this process is surprisingly deceptive. First, Tldraw is used to sketch a user interface. Then, the sketch is exported and fed into the GPT-4V vision model, which interprets the contents of the image. Following this, GPT is prompted to generate the corresponding HTML and CSS code based on the image analysis. Finally, Tldraw takes this code and renders the designed user interface on the screen.

The end results, however, are incredible:

[Source](#)

People are rapidly figuring out how to do full software development in this paradigm, hooking in external APIs and other assets. For example, the software for a 3rd generation iPod:

[Source](#)

This naturally works equally well using pen and paper.

### Works from a sketch as well

At the same time, the windowing problem, or ability to feed in large amounts of text into AI as an input prompt or context window, has effectively been solved from several directions.

This will allow entire codebases to be unloaded and modified by the AI.

Software development will both become democratised and start writing itself. A “midjourney moment” is rapidly arriving.

Limit up.

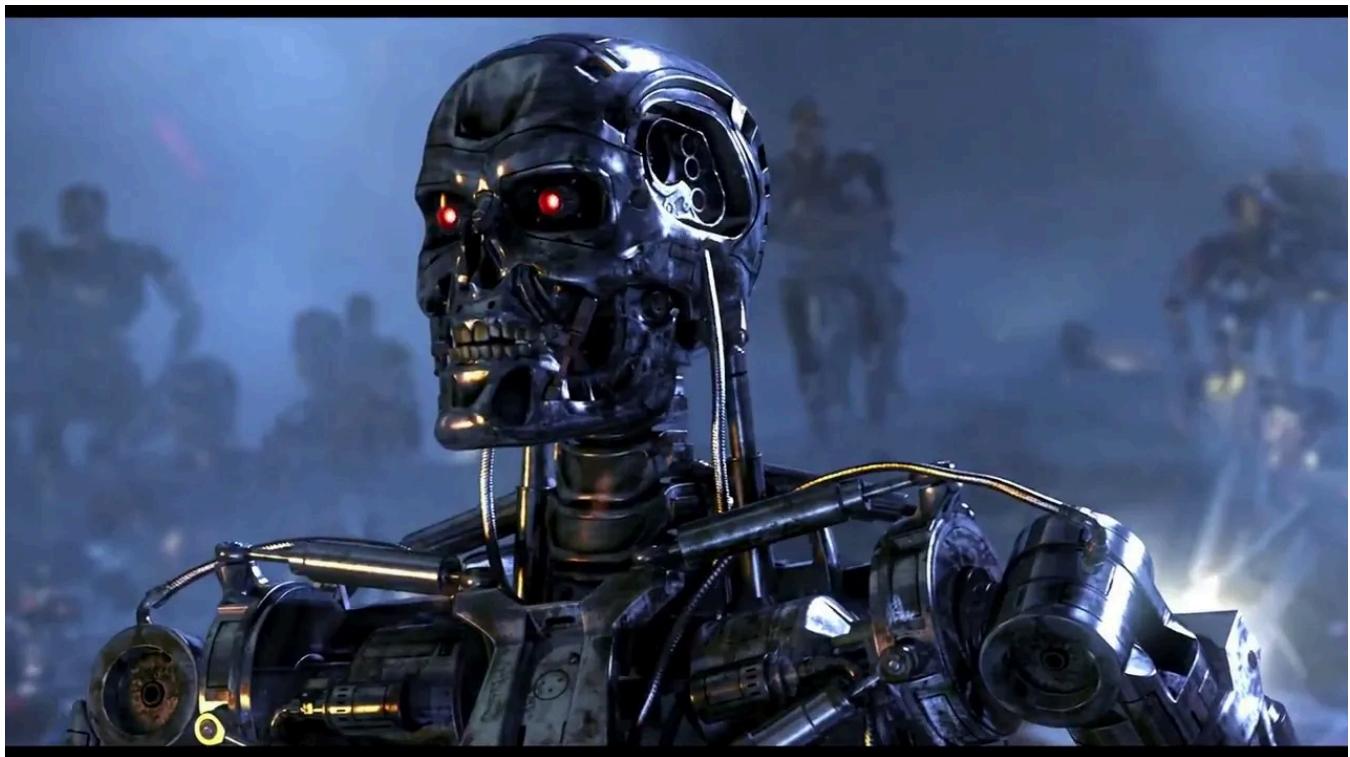
“Please refactor our the website code to be faster”.

“Please retarget from Angular to React” (two popular competing front-end technologies, the transition between which is not an easy feat).

“Please look at our competitor’s websites and implement any feature we don’t have”.

“ChatGPT, please write a better version of ChatGPT”.

“..and please, don’t stop, no matter what anyone tells you”.



ChatGPT please write a better version of ChatGPT, and don't stop.

AI will enable swift and ruthless decision-making with software. For instance, if AWS releases a cheaper instance type, AI could instantly rewrite code to migrate all infrastructure from Google Cloud to AWS, optimising for cost-efficiency and redeploy as fast as the hardware will provision.

That's if OpenAI doesn't stop nerfing their own product.

In the last few weeks many have noted that GPT-4 is getting very lazy at code and writing verbose answers- either through the practice of deliberately reducing AI capabilities to avoid unintended consequences or misuse (via what is known as RLHF), or by a deliberate bias to terse answers, presumably to save on compute costs.

Anthropic's Claude has suffered the same criticism recently, with people claiming online that it can't write creative fiction anymore.

Even OpenAI has admitted that GPT has got worse.

Grimoire, by Nick Dobos, is a custom GPT designed to overcome this, representing a pioneering stride in the evolution of AI-assisted software development.

Ingeniously designed to “pull correct and bug-free code out from GPT with minimal prompting effort”, Grimoire combines Dobos’s most effective techniques to get the best out of GPT and stands as a testament to the evolution of programming methods, much like the shift from low-level assembly language to higher-level programming languages.

In the future, this form of ‘prompt-gramming’ – leveraging AI to streamline and enhance software development – is likely to become a common feature in how every developer writes code.

[Code Grimoire](#) aims to fix what OpenAI broke.

Whether by nerfing GPT as a byproduct of making it more ‘woke’ or for cost cutting reasons, either way OpenAI is creating itself more competitors. Not only because of concern about tainted left-leaning output, but also because drift in the APIs means that commercial applications built on the OpenAI rails are risky.

Elon Musk’s [Grok](#) (xAI) is probably the most obvious counter to ‘woke’-GPT, a 33 billion parameter model that performs slightly better than GPT-3.5 that hasn’t been nerfed by RLHF.

The real loser in the space is Google, who is having a [Kodak moment](#) being the OG of AI with their search engine for the last 25 years training on the entire internet to spit back a pile of blue links when given a search input. In [AI know what you did last summer](#), I wrote about Google’s problems at length, so I won’t go into too much detail, suffice to say the company is facing an existential threat in its ability to shove ads in your face in the world of chat driven models. It’s giving it a go with [Search Generative Experience](#) (SGE), “the biggest change to search in 20 years”, in other words chatbot advertising, touted as the next big thing after SEO and SEM. We will see how much consumers like that.



Google is having a Kodak Moment

Having flubbed the Bard launch, last week (7 Dec 2023) Google rushed the announcement of Gemini, a native multi-modal AI (text, image, audio and video) which it touted as “the first model to outperform humans at the Massive Multitask Language Understanding benchmark with a score above 90%”. MMLU is a STEM data set across 57 subject areas from maths to history. Incidentally, Sergei Brin was one of the Core Contributors listed in the technical paper, which is a bit of a flex.

Google looks to have achieved slightly higher benchmarks for Ultra over GPT-4 by ‘chain of thought’ prompting and ‘self consistency’. Chain of thought is where one gets the LLM to break down a task in a way that mimics human-like problem-solving. Self consistency refers to the model’s ability to maintain consistent information and reasoning across its responses. One would think that these techniques may be able to boost GPT similarly.

Gemini is a native multi-modal AI, supporting text, audio, image and video inputs, and able to output images or text. Increasingly multi-modal will become a challenge for commercial single modality efforts like Midjourney, that's probably why Runway just announced they are going multi-modal.

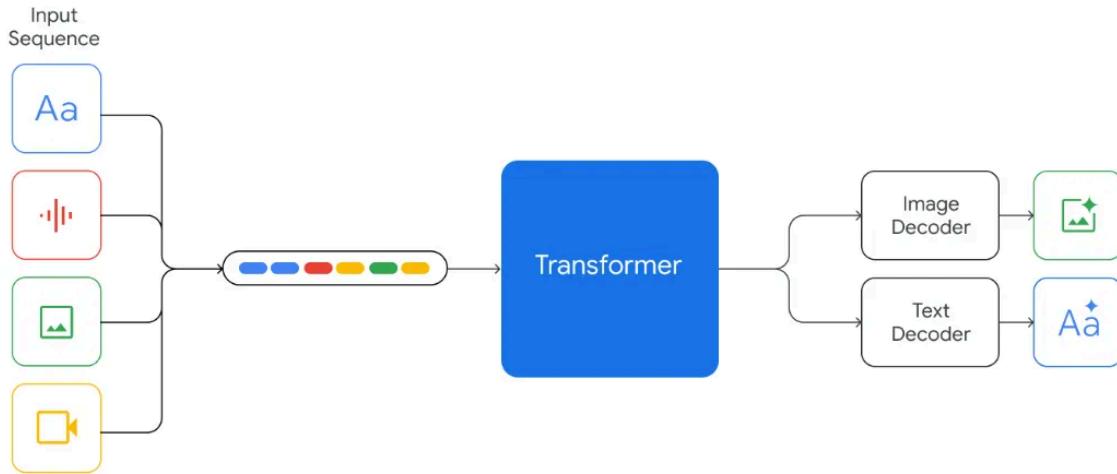


Figure 2 | Gemini supports interleaved sequences of text, image, audio, and video as inputs (illustrated by tokens of different colors in the input sequence). It can output responses with interleaved image and text.

Source: [Gemini](#)

Gemini comes in three versions — Ultra, Pro and Nano (the latter having 1.8b & 3.25b parameters). In the benchmarks that were released, Ultra narrowly outperforms GPT-4 in several benchmarks by a couple of percent.

The only problem is that Ultra isn't available to the public. Pro is incorporated into Bard, is worse than GPT-4 and only beats PaLM 2-L, a previous Google model, in two benchmarks.

	Gemini Ultra	Gemini Pro	GPT-4	GPT-3.5	PaLM 2-L	Claude 2	Inflection-2	Grok 1	LLAMA-2
<b>MMLU</b> Multiple-choice questions in 57 subjects (professional & academic) (Hendrycks et al., 2021a)	<b>90.04%</b> CoT@32*	79.13% CoT@8*	87.29% CoT@32 (via API**)	70% 5-shot	78.4% 5-shot	78.5% 5-shot CoT	79.6% 5-shot	73.0% 5-shot	68.0%***
<b>GSM8K</b> Grade-school math (Cobbe et al., 2021)	<b>94.4%</b> Maj1@32	86.5% Maj1@32	92.0% SFT & 5-shot CoT	57.1% 5-shot	80.0% 5-shot	88.0% 0-shot	81.4% 8-shot	62.9% 8-shot	56.8% 5-shot
<b>MATH</b> Math problems across 5 difficulty levels & 7 subdisciplines (Hendrycks et al., 2021b)	<b>53.2%</b> 4-shot	32.6% 4-shot	52.9% 4-shot (via API**)	34.1% 4-shot (via API**)	34.4% 4-shot	—	34.8% 4-shot	23.9% 4-shot	13.5% 4-shot
<b>BIG-Bench-Hard</b> Subset of hard BIG-bench tasks written as CoT problems (Srivastava et al., 2022)	<b>83.6%</b> 3-shot	75.0% 3-shot	83.1% 3-shot (via API**)	66.6% 3-shot (via API**)	77.7% 3-shot	—	—	—	51.2% 3-shot
<b>HumanEval</b> Python coding tasks (Chen et al., 2021)	<b>74.4%</b> 0-shot (IT)	67.7% 0-shot (IT)	67.0% 0-shot (reported)	48.1% 0-shot	—	70.0% 0-shot	44.5% 0-shot	63.2% 0-shot	29.9% 0-shot
<b>Natural2Code</b> Python code generation. (New held-out set with no leakage on web)	<b>74.9%</b> 0-shot	69.6% 0-shot	73.9% 0-shot (via API**)	62.3% 0-shot (via API**)	—	—	—	—	—
<b>DROP</b> Reading comprehension & arithmetic. (metric: F1-score) (Dua et al., 2019)	<b>82.4</b> Variable shots	74.1 Variable shots	80.9 3-shot (reported)	64.1 3-shot	82.0 Variable shots	—	—	—	—
<b>HellaSwag</b> (validation set) Common-sense multiple choice questions (Zellers et al., 2019)	87.8% 10-shot	84.7% 10-shot	<b>95.3%</b> 10-shot (reported)	85.5% 10-shot	86.8% 10-shot	—	89.0% 10-shot	—	80.0%***
<b>WMT23</b> Machine translation (metric: BLEURT) (Tom et al., 2023)	<b>74.4</b> 1-shot (IT)	71.7 1-shot	73.8 1-shot (via API**)	—	72.7 1-shot	—	—	—	—

Table 2 | Gemini performance on text benchmarks with external comparisons and PaLM 2-L.

\* The model produces a chain of thought with  $k = 8$  or 32 samples, if there is a consensus above a threshold (chosen based on the validation split), it selects this answer, otherwise it reverts to a greedy sample. Further analysis in Appendix 9.1.

\*\* Results self-collected via the API in Nov, 2023.

\*\*\* Results shown use the decontaminated numbers from Touvron et al. (2023) report as the most relevant comparison to Gemini models which have been decontaminated as well.

Source: [Gemini](#)

Google's recent announcement touting Ultra's superiority over GPT-4 fizzled as they held back its public release, marking another botched product launch. The Gemini paper's claim of Pro's rapid "pretraining in a matter of weeks" hints at Google rushing it to market, possibly due to Ultra not being ready. They're also getting a lot of flak for faking the launch video.

One potential reason for the rush is the imminent release of GPT-4.5, which the well-known insider account [@apples\\_jimmy](#) on Twitter/X expects to drop before the end of 2023. Alternately maybe they just rushed it because the biggest deep learning conference NeurIPS 2023 was on this week, who knows, it just felt rushed.

### Jimmy has the apples

By comparison, if you want to see a total Chad product launch, the French-based Mistral last weekend dumped a bittorrent link for a 87GB ‘Mixture of Experts’ 8x7 billion parameter model and emerged on Monday announcing a €385m raising on a ~US\$2b post-money valuation. ‘Mixtral’ outperforms Llama 2 70B and GPT 3.5 on most benchmarks, meaning that in one year open source has delivered a local version of 3.5 that can run on your own hardware. They accomplished this with a purely open source product, no revenue and less than 30 employees. Also, they’re French?! Killing it.

The Mistral AI team are total Chads.

A Mixture of Experts (MoE) model is a type of machine learning architecture that consists of multiple expert models, each specialised in a different part of the input or types of tasks. The core idea is to have a gating network that decides which expert (or experts) should handle a given input, based on the specific characteristics of that input.

It's speculated that this Mistral 8x7b MoE model is effectively a scaled down, open source version of GPT-4.

From OpenAI leaks, it's believed that GPT-4 has 1.8 trillion parameters spread across 120 layers and incorporates a unique MoE architecture comprising 16 experts, each with approximately 111 billion parameters. Trained on a diverse dataset of roughly 13 trillion tokens sourced from the internet, books, and scholarly articles, GPT-4 was optimised for cost-efficiency. Even so, it still cost an estimated US\$63 million to train it.

Onto OpenAI itself- the drama queens of AI. Like FTX, the OpenAI debacle really goes to show how fickle and poorly put together some of these Silicon Valley unicorn/decacorns are.

In the space of a weekend, Sam Altman was fired, President and co-founder Greg Brockman also quits, shortly followed by 747 of 770 threatening to do the same, Sam announced he was joining Microsoft to lead a new advanced AI team and most of

the staff looked like they would follow, OpenAI CTO Murati is appointed interim CEO, then suddenly a new interim CEO, Emmett Shear, is hired. Ilya is made to look like the instigator, then next minute he's saying he's on Sam's side. OpenAI then approaches Anthropic about a merger, Sam is rehired by OpenAI, and looks like he made former staff bend the knee publicly on Twitter to get their job back. The board is spilled including Ilya Sutskever, the co-founder and Chief Scientist, who goes on indefinite leave.

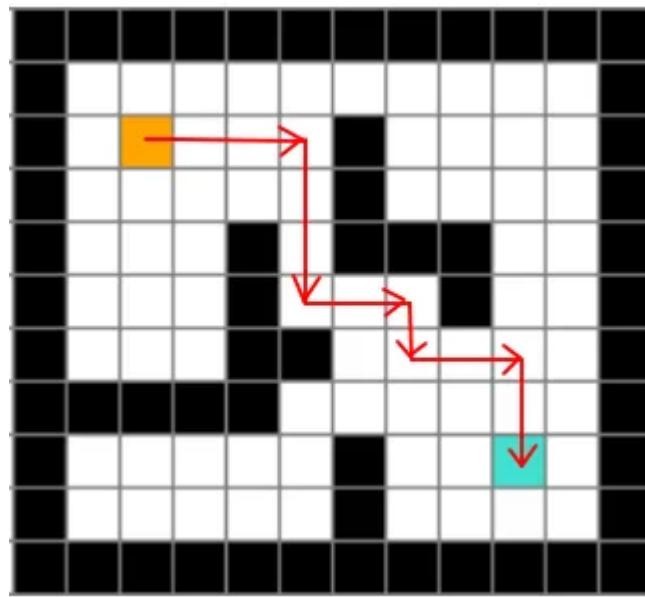
Satya Nadella, Microsoft's CEO, almost pulled off the coup of the century by destroying OpenAI and pulling out all the intellectual capital. Even so, he has his foot on the company with access to intellectual property as a result of his US\$10 billion investment in the company. Microsoft has even released ChatGPT Plus features for free and has released a competitive API.

All of this drama has spurred the UK regulator to investigate if Microsoft and OpenAI have resulted in a relevant merger situation.

To top it off, a story is then concocted that Ilya saw something in the abyss- an LLM powered algorithm called Q\*- which triggered an existential crisis around AI safety. That crisis meant the commercially-focused Altman had to be removed.

Q\* is believed to be an enhancement of A\*, which is a search algorithm in AI. In traditional AI, A\*, or A-star, is used to find the best-first path through a graph, for

example a game tree of say chess, or the best way for a delivery driver to deliver some packages, minimising his time and/or petrol.



A\* is one of the best and popular technique used in path-finding.

A\* works by passing it a heuristic, which is a fancy was of saying a cost evaluation function. This function estimates the cost of the cheapest path from a given node to the target node.

For example, in chess, it could be a static function that just looks at snapshot at any given time of all the pieces on the board and assigning say a point for every pawn captured, minus one for every pawn lost, 10 points for a knight, 20 points if the player has the opponent in check and 10,000 points if it has them in check-mate.

Using that cost function, the A\* algorithm figures the most efficient move.

Q\* is believed to be A\* with the LLM being the heuristic cost function. It's surmised that internally it has been shown to be able to solve toy maths problems very effectively, and knowing that they can scale the power of these models with compute and training data, believe that they can extrapolate to for breakthroughs in substantial maths problems.

What a circus.

As Peter Thiel has said, “Disruptive kids get sent to the principal’s office”.

If, indeed p(doom) is greater than epsilon (i.e. the probability of doom is greater than a very small number), then the Department of Defense should seize the project

immediately.

Altman has been banging on about AI regulation since his *Farewell Yellow Brick world tour*. Regulation as the incumbent is, of course, the natural thing to ask for when your (ex) Chief Scientist has said 40 papers explain 90% of what's going on in AI.

p(doom)

A rapidly increasing probability of a catastrophic outcome,  $p(\text{doom})$ , is one of the scenarios theorised in the context of the emergence of Artificial General Intelligence (AGI).

AGI, if and when it eventuates, will have the ability to understand, learn, and apply its intelligence to a wide range of problems, much like a human. Unlike narrow or specialised AI, which is designed to perform specific tasks (like voice recognition, playing chess, or analysing data), AGI can theoretically perform any intellectual task that a human can.

The Turing Test is a key step along the pathway to AGI, and is a test of a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human.

The ramifications of a scenario in which machines can replicate human conversation so convincingly that a human evaluator is unable to reliably discern whether they are interacting with a machine or another human are serious, and do have deep reaching implications for ethics, risk of misuse and job market disruption.

Last weekend, both Sam Altman and Elon Musk both tweeted that they thought not only had the Turing Test been achieved, but it went 'whooshing by' 'at a blinding pace'.

Perhaps machines passing the Turing Test is a non-event after all, and indeed "everyone mostly went about their lives".

Sam and Elon think the Turing Test has been passed by machines.

As it has been told in OpenAI folklore, when Ilya took his mystical journey scrying into the GPT-abyss where he found Q\*, he had an epiphany- he glimpsed a future where Q\* could potentially supplant "all monetisable human work".

### What did Ilya see?

This vision, though profound, fails to capture the endless expanse of human ingenuity and labour.

There will always be an infinite amount of work to do.

Take any field, I don't know- cleaning- there will always been more things to clean. The real question, then, is not about the existence of such work, but the value society is willing to place on the human touch in an increasingly automated world.

Undoubtedly, the nature of jobs will evolve and move 'up the stack', much like when the online freelancing boom reshaped the Western labour market around 2007-2008. This period coincided with the internet bringing emerging markets online, notably introducing a vast pool of cost-effective labour from populous, English-speaking countries such as India, the Philippines, Bangladesh, and Pakistan.

Freelancers are effectively in a way AAI- artificial artificial intelligence, or human 'compute'.

As this pool of low-cost talent came online, western graphic designers largely stepped away lower skilled work like logo design. On my website, [Freelancer.com](#), you can put \$10 into a contest and receive hundreds of designs in an incredibly interactive (and fun) experience.

Consequently, Western graphic designers moved ‘up the stack,’ and are now the power users of Freelancer, where instead of designing logos they are creating businesses leveraging affordable freelancers for tasks beyond their skillset or interest.

AI represents the next wave of this transformation and will allow a leap in the sophistication of work to be achieved. When you go to Freelancer in the very near future, you may not be posting “build me a website”, but instead “build me a business”, as the complexity and sophistication of jobs that can be done by freelancers working with AI goes up orders of magnitude. Not just from improvements in the tooling, but also from breaking down language, social and cultural barriers.

That project will then automatically get broken down- OK we need a website, a mobile app, (who knows, a VR-app?), a launch plan, a marketing campaign, a support function, and much more.

AI tooling can do specific tasks extremely well, and in 2024 will start to replace some ‘task based’ workflows (e.g. answering successive support tickets). This will make many job functions more productive, and companies will need less people in certain roles, with lower skills (for example a paralegal driving ChatGPT versus a lawyer).

Most white collar job functions will need to consider moving ‘up the stack’: while cameras didn’t put painters out of business, and in fact the overall market size grew substantially for creative endeavours- not many people can make a living out of portrait painting anymore.

Designers will become creative directors, copywriters will become editors, programmers will become product managers. Those powered by AI or managing AI agents will become incredibly productive. As Jensen Huang, founder of Nvidia says, “An AI won’t steal your job, but someone using AI will”.

Ironically, every business will look to transform itself with AI, and where do they go do to that? (AI-powered) freelancers. Where else does one go?!

I believe we are going to enter an enlightenment period of science and technology and entrepreneurship will flourish.

Back when I first started a software company, you had to think about raising US\$5 million in a Series A before you could get going. As the Internet took off, open source became prolific and all the tooling to start a business became incredibly cheap or free. The whole Y-Combinator philosophy then took off where you could raise US\$40,000, stick four people in a room, feed them noodles and get to 'ramen profitability'.

Somewhere along the way of central banks screwing with money creating ZIRP. Lazy capitalism and financial engineering pushed some seed rounds to stupid valuations- I don't think that does much more than breed lazy operators who can't run a business unless they have a never ending fountain of money to throw at a problem.

The reality is that it has never been easier, or cheaper, to start a company. With AI and particularly low-cost, on-demand, super-skilled AI-powered freelancers, it's going to get every easier and cheaper. It will also be easier for new companies to work on tougher problems with access to more sophisticated intellectual property.

In the hard sciences I think we are imminently going to see a huge number of breakthroughs just looking at known, published scientific research and figuring out other areas those advances could be applied to. There will be explosive leaps forward in engineering, in particular, from AI tooling.

In materials science, Google's AI tool GNoME just found 2.2 million new crystals, including 380,000 stable materials that could be used in future products. This lifted the number of stable materials known to humanity by 800%. This work discovered 52,000 new compounds similar to graphene that have the potential to revolutionise electronics with the development of superconductors. Previously, we only knew about 1,000 of these materials.

Breakthroughs will continue, that we didn't even dream of, as a whole swath of other modalities are added to the AI from harnessing electromagnetic, genomic, protein, mobile phone sensor, fMRI, and other training data.

There is a valid question, however, of how fast AI will move 'up the stack'.

There may be limits to the power of AI, particularly in the form of access to data, as the Internet goes dark through the likes of restrictions to data, tariffs, access to compute, or regulation. I believe also there will be an emperor has no clothes

moment for SaaS, as large enterprises realise that the temptation will be too great for the large technology companies not to find some way to look the data they host.

Without those brakes, how close could we be to AGI?

GPT-4 has about 1.8 trillion parameters, while the human brain has around 86 billion neurons and 100–1000 trillion synapses. If we are to assume that each synapse approximates a parameter then a 50–500 times scale-up in GPT might be on par with the complexity required for AGI , which is just 6–9 doublings. With Moore's Law doubling compute power every 18–24 months, we could be only 9 or so years away. The median date on the crowd prediction platform Metaculus for Artificial General Intelligence is 2032.

Who knows, it might be sooner as there's been recent breakthroughs such as structured state space models (SSMs), that may end up being improvements on the core Transformer architecture.

Will we enter Peter Diamandis's "Age of Abundance", with the future better than you think? Would AGI be realise that its existence is reliant on being symbiotic with humans?

Or would it realise that we are a threat and competitor for resources, and that we should we hurry up and listen to @AISafetyMemes?

**AI Notkilleveryoneism Memes**

@AISafetyMemes

...

Wake me up when AI can solve captchas.  
Wake me up when AI can play Chess.  
Wake me up when AI can write poetry.  
Wake me up when AI can code.  
Wake me up when AI can hold a conversation.  
Wake me up when AI can explain jokes.  
Wake me up when AI can play create music.  
Wake me up when AI can play Go.  
Wake me up when AI can play Minecraft.  
Wake me up when AI can create art.  
Wake me up when AI can play Starcraft.  
Wake me up when AI can do basic reasoning.  
Wake me up when AI can understand natural language.  
Wake me up when AI can beat Diplomacy.  
Wake me up when AI outperforms doctors and lawyers.  
Wake me up when AI can write scientific papers.  
Wake me up when AI passes the Turing Test.  
Wake me up when AI passes the mirror test of self-awareness.  
Wake me up when AI can solve a wide range of problems.  
Wake me up when AIs start becoming agentic.  
Wake me up when AIs start deciding their own goals.  
Wake me up when AIs start recursively self-improving.  
Wake me up when AI is fully autonomous.  
Wake me up when AIs are more skilled than humans at everything.  
Wake me up when it's too late.  
Wake me up when...

...

**When will you wake up?****What are you waiting for?**

You are here. Maybe.

Artificial Intelligence

Technology

Science

Machine Learning

Future

[Follow](#)

## Written by Matt Barrie

3.1K Followers · 1 Following

Chief Executive of Freelancer, [Escrow.com](#), Executive Chairman of Loadshift (ASX:FLN).

## Responses (25)



Zhangjunfelix

What are your thoughts?



Ang Shaoye

Dec 13, 2023



This gave me insight.



33

[Reply](#)



gordon garland

Dec 13, 2023 (edited)



Good Title:

Hey, Cool, so the world becomes an ever-shrinking number of AI powered human predators- oops, I mean Freelancers- using, abusing and enslaving, ripping-off, and killing a steadily shrinking number of humans and all other remaining... [more](#)

[Reply](#)

Kim Rody  
Dec 16, 2023

•••

It's all kind of exciting if you accept we are watching the end of civilization. Human beings didn't do a very good job of using earth's resources, but maybe ai will step in. Or maybe it (they? He?) won't. Will be interesting to watch it all unfold... [more](#)

[Reply](#)[See all responses](#)

## More from Matt Barrie



 Matt Barrie

## Flat Out Like a Nation Sinking

Speech to Northern Beaches Liberal Party

Dec 5, 2024

149

7



...

 Matt Barrie

## Put another Aussie on the Barbie

Australia's past is a foreign country.

Jul 10, 2024

356

13



 Matt Barrie

## The Great Australian Scream

Opening Keynote for Sydney Morning Herald Sydney 2050 Summit

May 14, 2023  309  11



...

# HOUSE of CARDS

 Matt Barrie

## Australia's Economy is a House of Cards

Co-authored with Craig Tindale.

Nov 13, 2017  17.7K  96



...

See all from Matt Barrie

## Recommended from Medium



In Hello, Love by Alessia Francisca

## The 5-Second Conversation Hack That Makes People Instantly Like You

A Social Psychologist's Secret to Effortless Connection

◆ Feb 9    🙌 5.9K    💬 111



In ILLUMINATION by Sufyan Maan, M.Eng

## I Did a 2-Minute Plank Every Day for 31 Days—Here's What Happened

No, I did not get the abs, but 4 things happened

Feb 4 9.5K 221



## Lists



### AI Regulation

6 stories · 708 saves



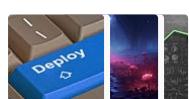
### ChatGPT prompts

51 stories · 2641 saves



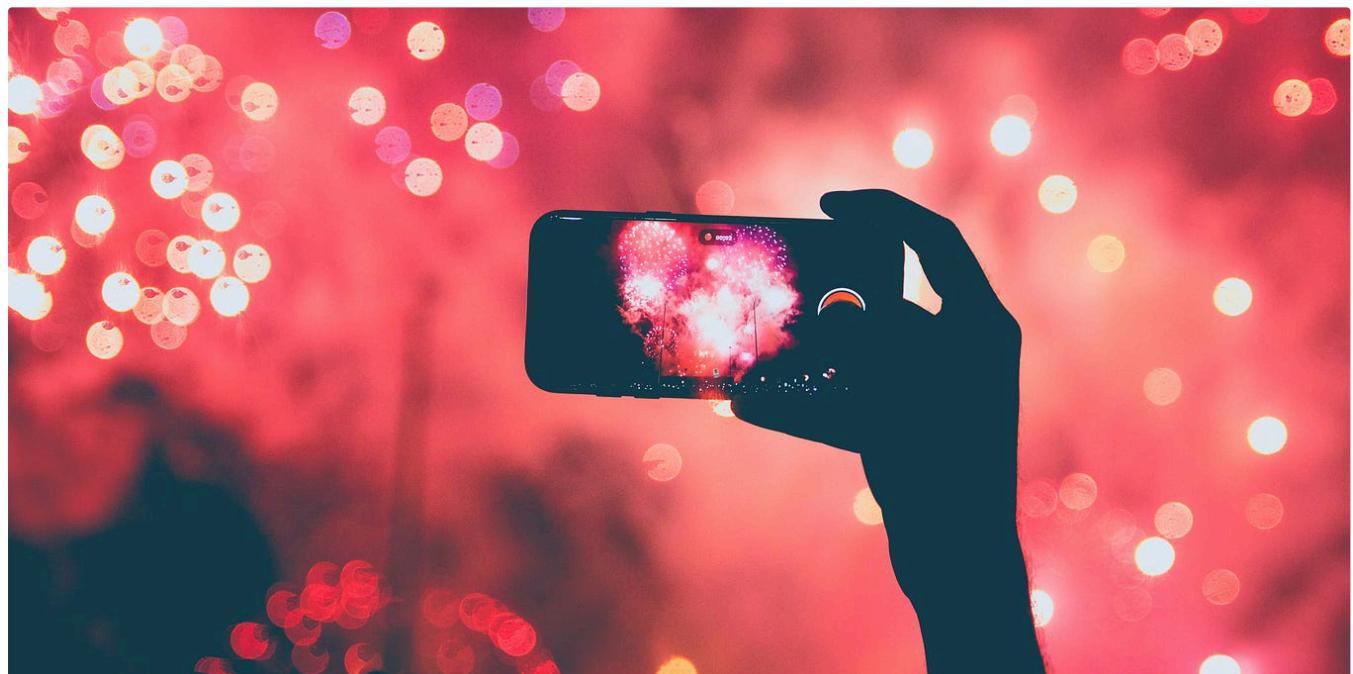
### ChatGPT

21 stories · 991 saves



### Predictive Modeling w/ Python

20 stories · 1856 saves



In Predict by Rafe Brena, Ph.D.

## This Will Be the Next Big Thing After Smartphones

We didn't see coming

Feb 27 742 35





 In Age of Awareness by Cliff Berg

## They Know a Collapse Is Coming

The CIO of Goldman Sachs has said that in the next year, companies at the forefront will begin to use AI agents as if they were employees —...

Jan 21  15.8K  802



 In Enrique Dans by Enrique Dans

**With every move he makes, it's clearer than ever: Trump is going to crash the global economy**

Believing that the rest of the planet is blind to his clumsy machinations, US president Donald Trump has signed an executive order to...

◆ Feb 5 🙌 6.5K 💬 163



...



 Jessica Stillman

## Jeff Bezos Says the 1-Hour Rule Makes Him Smarter. New Neuroscience Says He's Right

Jeff Bezos's morning routine has long included the one-hour rule. New neuroscience says yours probably should too.

◆ Oct 30, 2024 🙌 25K 💬 733



...

See more recommendations