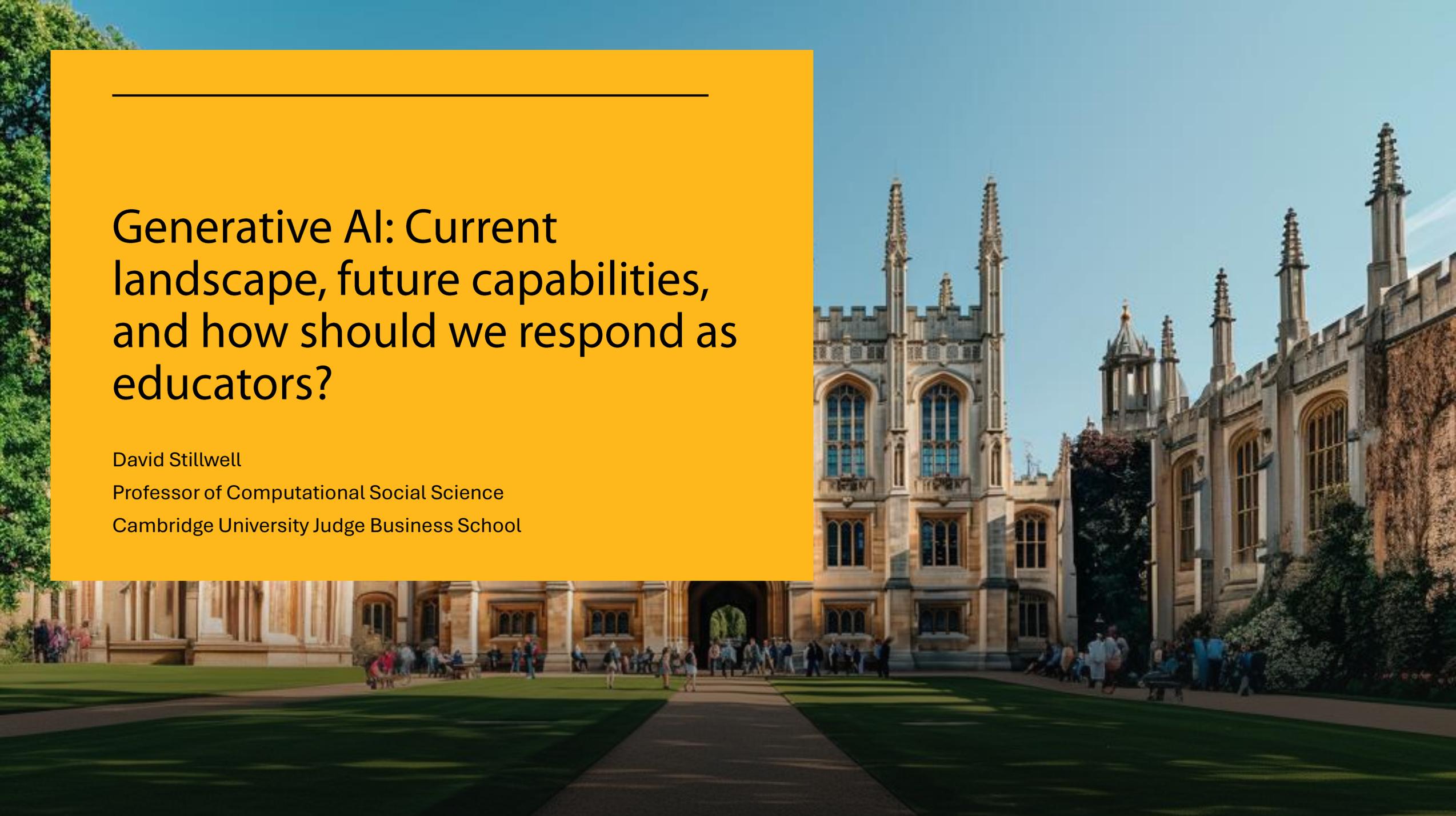

Generative AI: Current landscape, future capabilities, and how should we respond as educators?

David Stillwell

Professor of Computational Social Science

Cambridge University Judge Business School

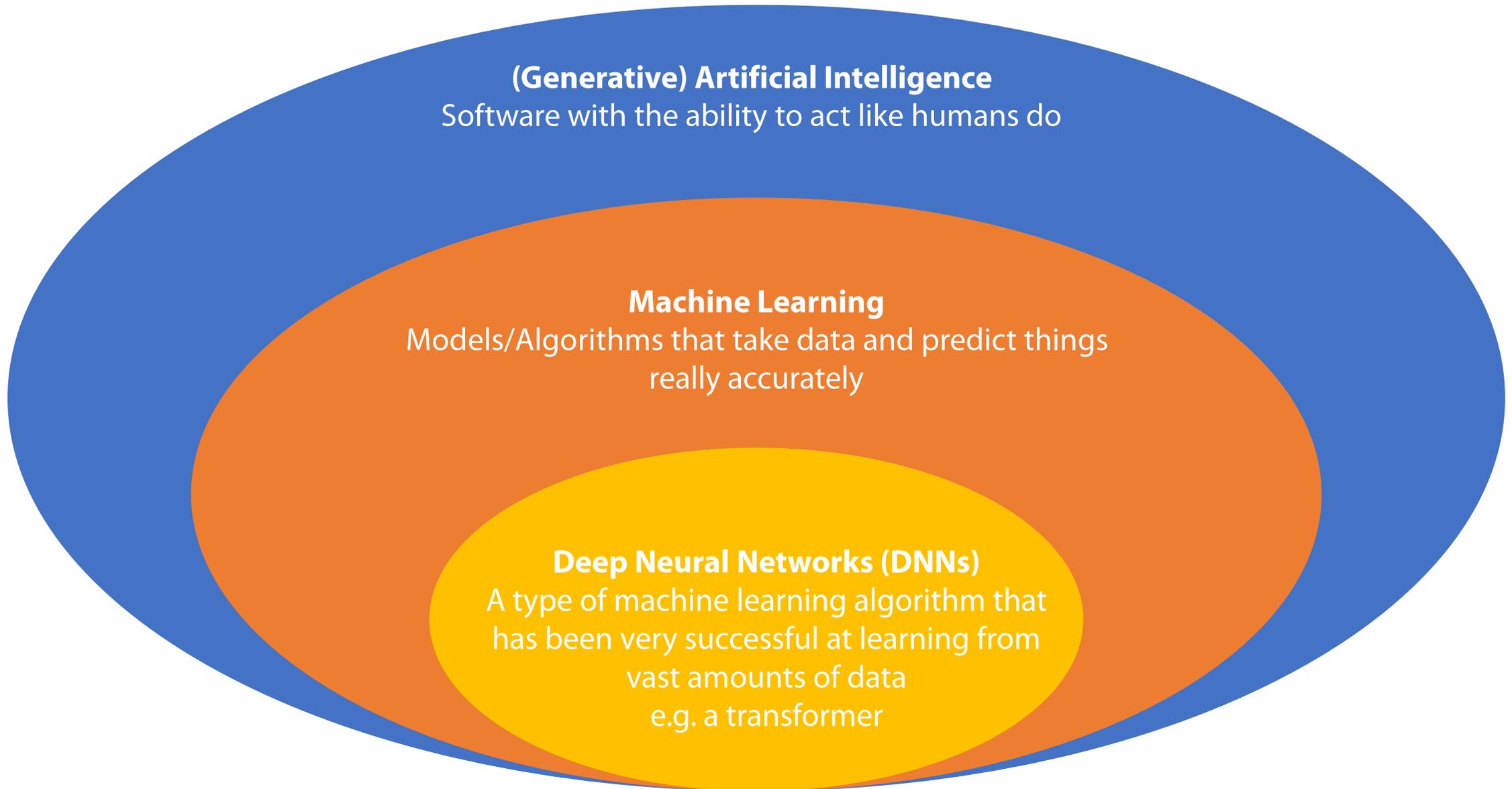


Agenda

- 1. The current landscape of Generative AI**
- 2. Where might it be going next?**
- 3. How do we respond as educators?**

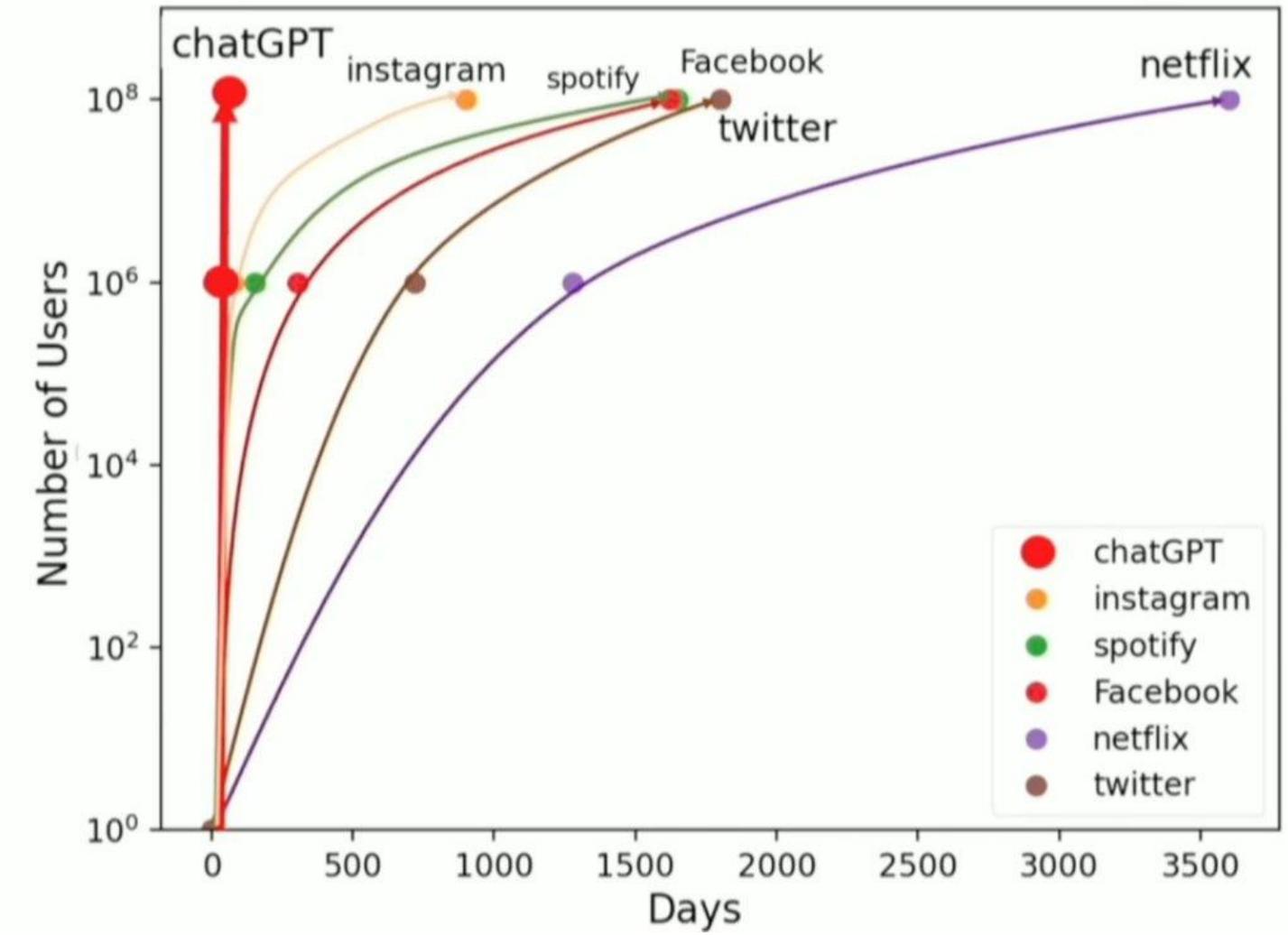


What is (Gen)AI?



What GenAI Tools Have You Come Across?

Time to 100m users



List of 'nowhere near solved' problems in AI, from *A Brief History of AI*, published January 2021)

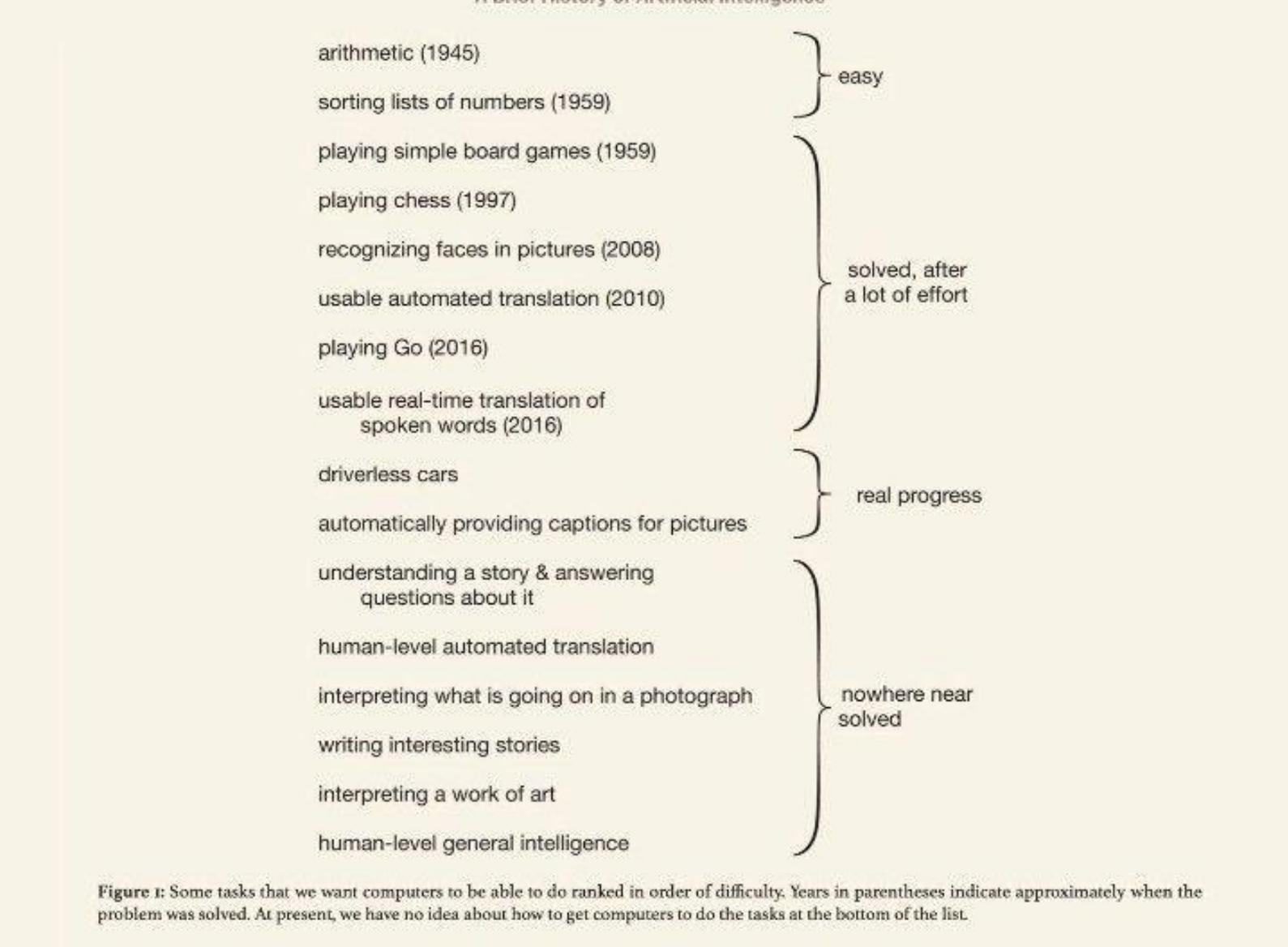
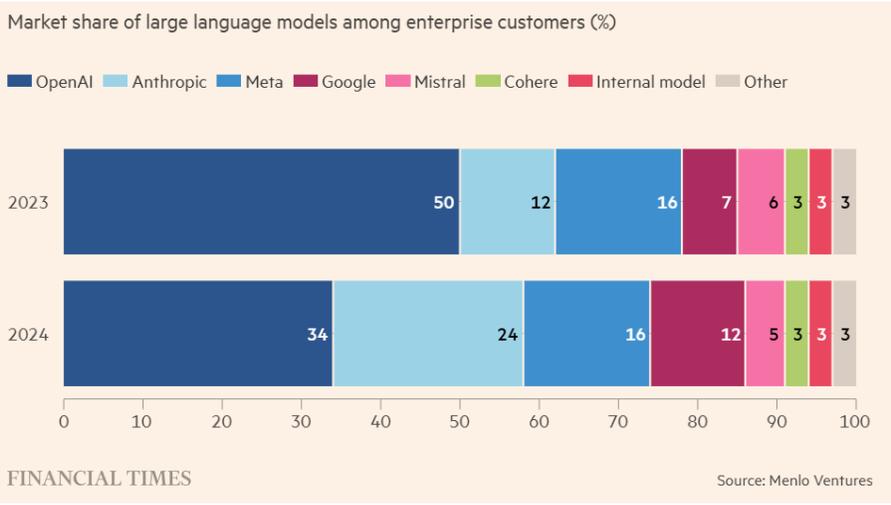


Figure 1: Some tasks that we want computers to be able to do ranked in order of difficulty. Years in parentheses indicate approximately when the problem was solved. At present, we have no idea about how to get computers to do the tasks at the bottom of the list.

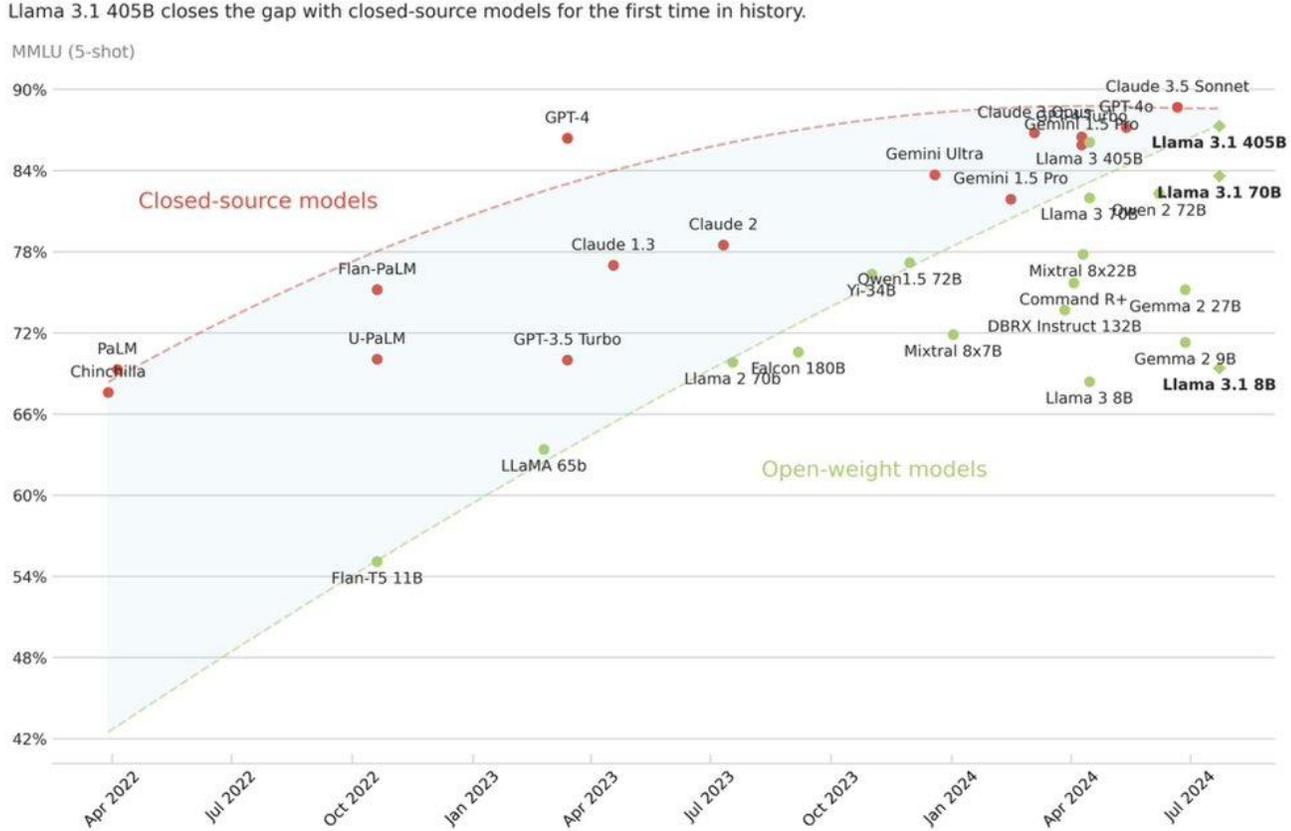
Large Language Models

- ChatGPT (OpenAI)
- Claude (Anthropic)
- Gemini (Google)
- Grok (Twitter)
- Llama (open weights from Meta)
- Mixtral (open weights from Mistral)
- DeepSeek-R1 (open weights from DeepSeek)
- Kimi (open weights from Moonshot AI)



Closed-source vs. open-weight models

@maximelabonne



<https://twitter.com/emollick/status/1825343853731491952>
<https://x.com/SullyOmarr/status/1862178462003343667>
<https://www.ft.com/content/fa8bad75-dc55-47d9-9eb4-79ac94e54d82>

Image generation models

MidJourney v1



MidJourney v2



MidJourney v3



MidJourney 4



MidJourney v6



MidJourney v5.2



MidJourney v5.1



MidJourney v5



- Dall-E (OpenAI)
- Gemini 2.0 Flash (Google)
- Grok (Twitter)
- MidJourney
- Stable Diffusion
- Firefly (Adobe)

<https://twitter.com/chaseleantj/status/1738849381632352493>

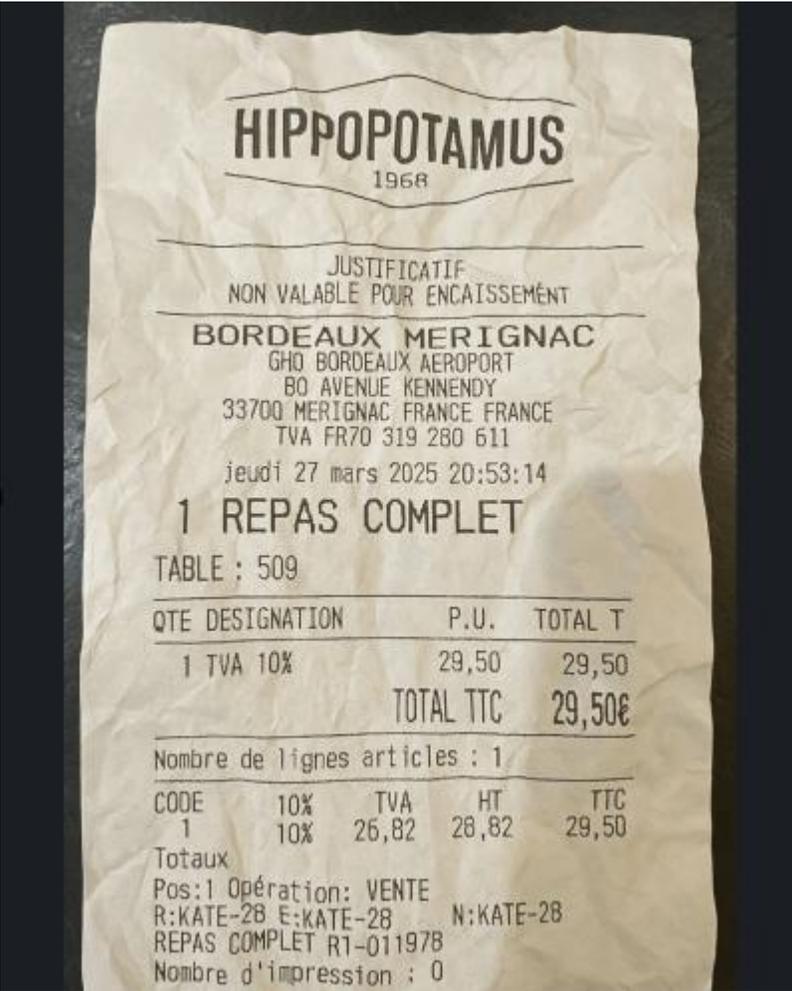
Which of these is/are AI generated?



“Remove the Watermark”



“Create this receipt”



Source: https://www.reddit.com/r/singularity/comments/1jbt0md/gemini_is_pretty_good_in_removing_watermarks/
<https://techcrunch.com/2025/03/31/chatgpts-new-image-generator-is-really-good-at-faking-receipts/>

Video generation

Sora (OpenAI)

Runway

Kling AI

Veo (Google)

CogVideoX (open source from Tsinghua Uni)



“Aerial tracking shot, golden-hour sunlight glinting off dust-caked motocross bikes weaving through towering dunes, rooster tails of sand exploding”

Source: https://x.com/dr_cintas/status/1914376345892639043

Content generation

HeyGen – avatar creation tool.

“One early adopter of Voice Engine is HeyGen , an AI visual storytelling platform that works with their enterprise customers to create custom, human-like avatars for a variety of content, from product marketing to sales demos. They use Voice Engine for video translation, so they can translate a speaker's voice into multiple languages and reach a global audience. When used for translation, Voice Engine preserves the native accent of the original speaker: for example generating English with an audio sample from a French speaker would produce speech with a French accent.”

Lumen5 – turn an article into a video.

Take your text, create a video with stock B roll.

Impacts of GenAI in practice

TECH

Google CEO says more than a quarter of the company's new code is created by AI

By [Hugh Langley](#)



Source: <https://www.businessinsider.com/google-earnings-q3-2024-new-code-created-by-ai-2024-10>



*AI use will be part of what we
look for in hiring*

*AI use will be part of what we
evaluate in performance reviews*

*Headcount will only be given if a
team cannot automate more of
their work*



Luis von Ahn, April 2025

Co-founder and CEO of Duolingo

Benefits of using AI agents for HR

Integrating AI agents into an HR department can have multiple, interlocking benefits that save enterprises time, money and wasted effort. For example, IBM's AskHR tool, a purpose-built platform automating over 80 common HR processes, has generated impressive results. Today AskHR [resolves 10.1 million interactions a year](#) on behalf of the HR department, saving 50,000 hours and USD 5 million a year. Simultaneously, the HR department's customer satisfaction score (CSAT) has improved dramatically since the tool's rollout.

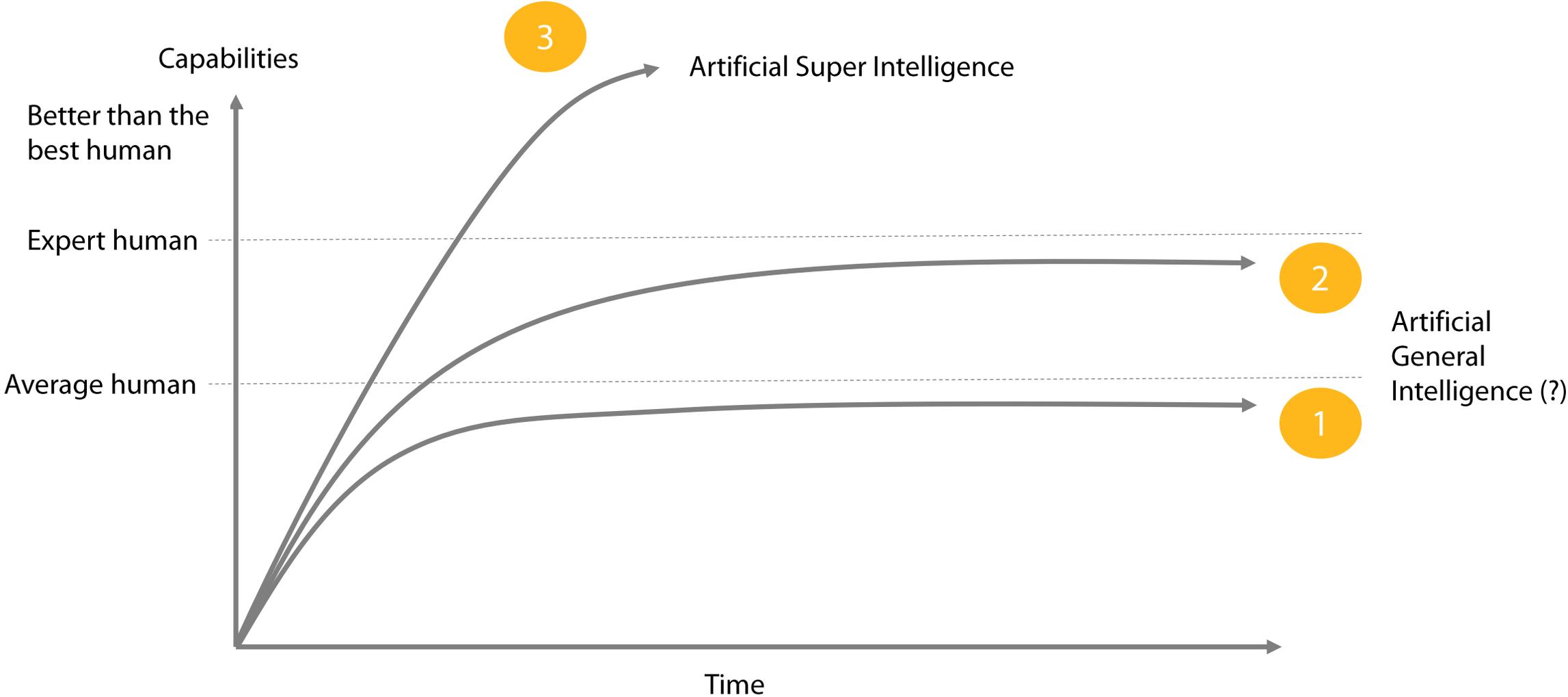
“AskHR now handles around **94% of routine HR inquiries**, such as vacation requests and pay questions.”

“One report notes that IBM's HR service center staff dropped from **800 to 60**, as AskHR took over much of the routine queries—empowering people to move to higher-impact roles.”

“The amount IBM spends on HR has dropped by 40 per cent over the past four years”

Where might GenAI be going next?

Three Possible Futures



The AI Race: Startups to Watch | AI Glossary | Model Collapse Risk | AI's Real Carbon Footprint | Small AI Models

Technology

OpenAI, Google and Anthropic Are Struggling to Build More Advanced AI

Three of the leading artificial intelligence companies are seeing diminishing returns from their costly efforts to develop newer models.



The ChatGPT virtual assistant logo. *Photographer: Andrey Rudakov/Bloomberg*

By [Rachel Metz](#), [Shirin Ghaffary](#), [Dina Bass](#), and [Julia Love](#)

13 November 2024 at 10:00 GMT



Is the Tech Industry Already on the Cusp of an A.I. Slowdown?

Companies like OpenAI and Google are running out of the data used to train artificial intelligence systems. Can new methods continue years of rapid progress?



WSJ

DANIEL HERTZBERG

The Next Great Leap in AI Is Behind Schedule and Crazy Expensive

OpenAI has run into problem after problem on its new artificial-intelligence project, code-named Orion

By [Deepa Seetharaman](#) [Follow](#)

Dec. 20, 2024 9:00 pm ET

AI is hitting a wall just as the hype around it reaches the stratosphere



Analysis by [Allison Morrow](#), CNN

3 minute read · Published 5:30 AM EST, Tue November 19, 2024

AI Scaling



“Computing Power and the Governance of Artificial Intelligence”
Sastry, Heim, Belfield, Anderljung, Brundage, Hazell, O’Keefe, Hadfield et al., 2024

Not enough high-quality data?

- “It’s become increasingly difficult to find new, untapped sources of high-quality, human-made training data that can be used to build more advanced AI systems. Orion’s unsatisfactory coding performance was due in part to the lack of sufficient coding data to train on, two people said.”
- Building AI systems that are smarter than a Nobel laureate may require data sources other than Wikipedia posts and YouTube captions.
- OpenAI, in particular, has inked deals with publishers to fill some of the need for high-quality data, and also adapt to growing legal pressure from publishers and artists over the data used to build generative AI products. Some tech companies are also hiring people with graduate degrees that can label data related to their own subject expertise, such as math and coding. The goal is to make these systems better at responding to queries about certain topics.



Sam Altman

@sama

Follow



there is no wall

6:06 AM · Nov 14, 2024 · **2M** Views

1.6K

1.6K

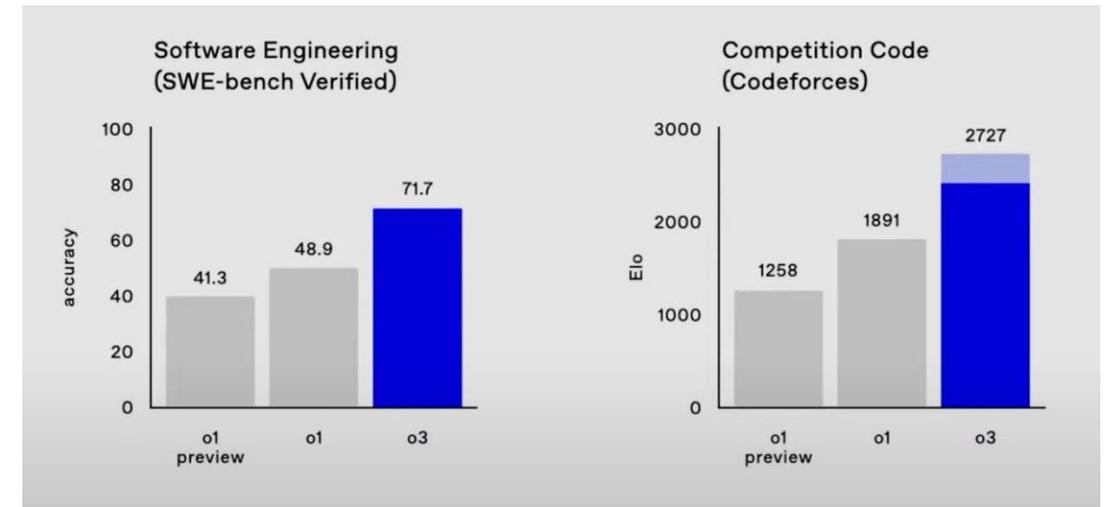
11K

659



And then came o3

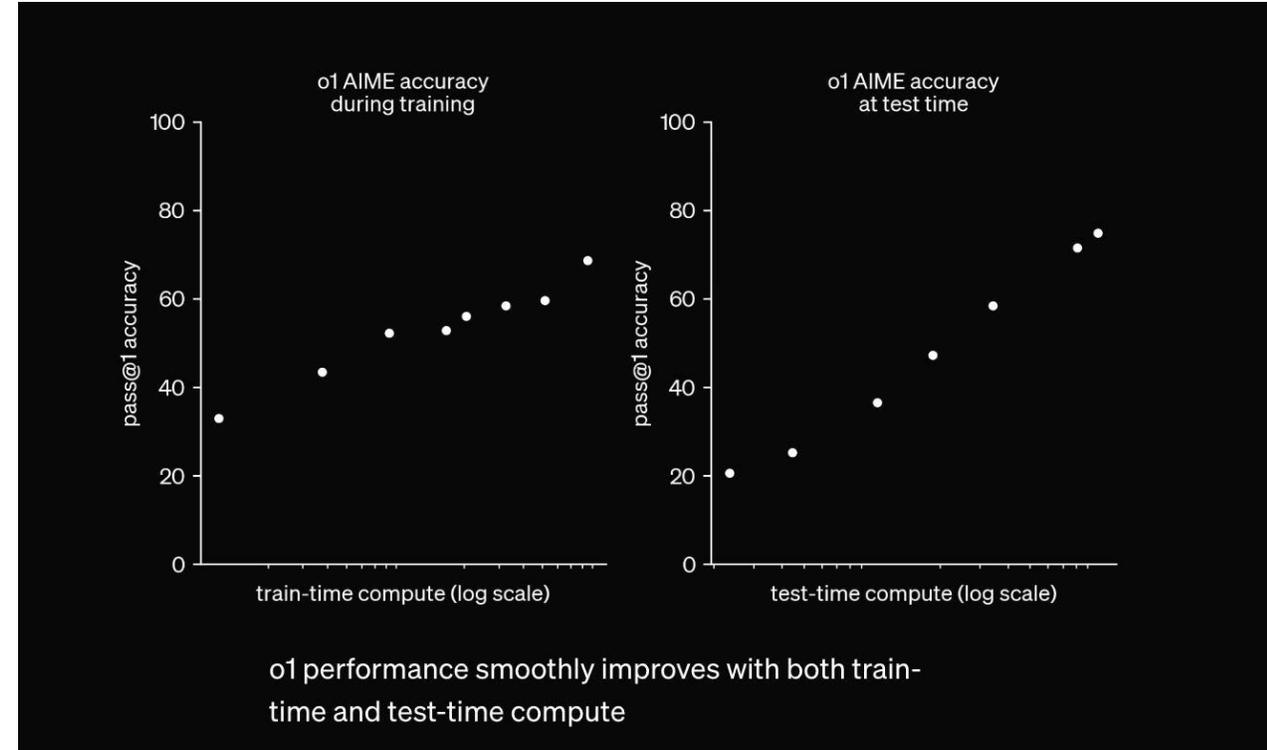
- In Nov' 24, Epoch AI introduced **FrontierMath**, "a benchmark of hundreds of original, expert-crafted mathematics problems designed to evaluate advanced reasoning capabilities in AI systems."
- Terence Tao, perhaps the greatest living mathematician, said the hardest of these problems are "extremely challenging... I think they will resist AIs for several years at least."
- State-of-the-art (SOTA) performance of an AI model was **~2%**. They were also novel and unpublished, to eliminate the risk of data contamination.
- OpenAI says that **o3 got 25%** of these problems correct.
- Jaime Sevilla, director of Epoch AI, wrote that the results were "far better than our team expected so soon after release. AI has **hit a wall, and smashed it through.**"



o3 also scored high enough on **Codeforces** programming competition problems to place it as 175th top-scoring coder in the world (out of the ~168k users in the last six months)."

Bigger models or more “thinking time”?

- Having a bot **think for just 20 seconds** in a hand of poker got the same boosting performance as scaling up the model by 100,000x and training it for 100,000 times longer.
- Can boost performance by **either** increasing **pretraining time** (classic scaling) or by increasing the “**test-time compute**,” i.e. dedicating more ‘thinking’ to the problem being addressed.



What is "thinking time"?

Gemini

invent a new word that I will find delightful, just give me the word at the end.

Show thinking

The request is to invent a new word that the user will find delightful. This is subjective, but 'delightful' usually implies something positive, pleasant, and perhaps aesthetically pleasing. Let's brainstorm categories of words that could be delightful and then try to create a new word within one of these categories.

Categories of potentially delightful words:

- Words that sound beautiful:** Consider the sounds of language itself. Soft sounds, flowing sounds, perhaps sounds reminiscent of nature. Examples of beautiful sounding words (existing): mellifluous, aurora, zephyr.
- Words that evoke pleasant imagery/feelings:** Words that make you think of happiness, peace, joy, wonder. Examples: serendipity, bliss, halcyon.
- Words that are playful or whimsical:** Words that are fun to say, have a lighthearted feeling. Examples: bamboozle, flibbertigibbet.
- Words that are unique or intriguing:** Words that stand out, spark curiosity. Examples: ephemeral, ineffable.
- Words related to positive concepts:** Concepts like growth, light, harmony, discovery, comfort.

Let's try to combine a few of these elements to create a new word. I think sound and positive imagery/feeling are good starting points.

Sound: Let's think of pleasant sounds. 'L' and 'R' sounds are often considered soft and flowing. Vowel sounds like 'oo' and 'ee' can also be gentle. Consonant sounds like 'm' and 'n' can be smooth.

Imagery/Feeling: Let's think about feelings of peace, joy, gentle happiness. Things associated with this could be soft light, a gentle breeze, a peaceful landscape, a moment of quiet joy.

Let's try combining sounds and imagery. How about a word that sounds gentle and evokes a feeling of soft light or a gentle breeze?

Let's experiment with sounds. 'Lumi-' or 'Lumo-' suggests light (from Latin 'lumen'). Let's combine this with a soft ending. '-osa', '-ia', '-ine', '-ee'.

'Lumiosa' - sounds a bit like 'luminosity'. Maybe too close to existing words.

'Lumee' - sounds simple and gentle. Like 'luminous' but softer.

'Lumine' - a bit more formal, like 'divine' or 'pristine'.

'Lumoria' - sounds more grand, maybe too much.

Let's consider a different starting point sound. 'Zephyr' is a beautiful word for a gentle breeze. What about sounds like 'Zef-' or 'Zaf-'.

'Zeffyra' - sounds a bit like a flower name. 'Zeffyrine' - too long maybe. 'Zeffee' - too close to 'zephyr' maybe.

Let's go back to the 'Lumi-' idea and refine it. Maybe instead of just light, think about 'gentle light' or 'soft glow'.

'Lumi-' and '-lucens' both relate to light. 'Lumi' is softer sounding than 'Lucent'. Let's try to make a word that evokes 'gentle glow'.

'Lumiflora' - combining 'lumi' with 'flora', like light and flowers. Sounds quite nice.

'Lumiflora' - pronunciation would be something like 'loo-mee-flor-ee'. It sounds a bit like a flower name, but also suggests 'light-flower'. Imagery is pleasant - flowers bathed in soft light, or perhaps flowers that themselves emit a gentle light.

'Lumiflora' - Let's check if it exists. A quick search suggests it isn't a common existing word, though it might be used as a name (brand, product etc) - but unlikely as a common English word.

'Lumiflora' - It sounds beautiful, evokes pleasant imagery (soft light and flowers), and is unique and intriguing. It is also quite easy to pronounce and remember.

I think 'Lumiflora' has a good chance of being considered delightful. It combines pleasing sounds, positive imagery, and a sense of novelty.

Let's try one more, just in case. Thinking of 'gentle breeze' again. 'Aurora' is a word for a subtle atmosphere or quality. What about combining 'aura' with something soft. 'Aurasy' - combining 'aura' with 'sy' which sounds soft (like 'sybar'). 'Aurasy' - pronunciation 'Aa-ra-sy'. Sounds quite gentle and flowing, suggests a gentle aura or quality. 'Aurasy' also sounds a bit unique.

Comparing 'Lumiflora' and 'Aurasy'. Both are good. 'Lumiflora' is perhaps more immediately visually evocative (light and flowers). 'Aurasy' is perhaps more subtly evocative (gentle aura).

For 'delightful' I think 'Lumiflora' might be slightly more broadly appealing due to the clearer imagery.

Decision: 'Lumiflora'. It sounds good, means something nice (purity), and is likely new.

© 2024 Flash Thinking Experimental. Lacks access to real-time info and some Gemini features.

Lumiflora

Gemini

invent a new word that I will find delightful, just give me the word at the end.

Show thinking

The request is to invent a new word that the user will find delightful. This is subjective, but 'delightful' usually implies something positive, pleasant, and perhaps aesthetically pleasing. Let's brainstorm categories of words that could be delightful and then try to create a new word within one of these categories.

Categories of potentially delightful words:

- Words that sound beautiful:** Consider the sounds of language itself. Soft sounds, flowing sounds, perhaps sounds reminiscent of nature. Examples of beautiful sounding words (existing): mellifluous, aurora, zephyr.
- Words that evoke pleasant imagery/feelings:** Words that make you think of happiness, peace, joy, wonder. Examples: serendipity, bliss, halcyon.
- Words that are playful or whimsical:** Words that are fun to say, have a lighthearted feeling. Examples: bamboozle, flibbertigibbet.
- Words that are unique or intriguing:** Words that stand out, spark curiosity. Examples: ephemeral, ineffable.
- Words related to positive concepts:** Concepts like growth, light, harmony, discovery, comfort.

Let's try to combine a few of these elements to create a new word. I think sound and positive imagery/feeling are good starting points.

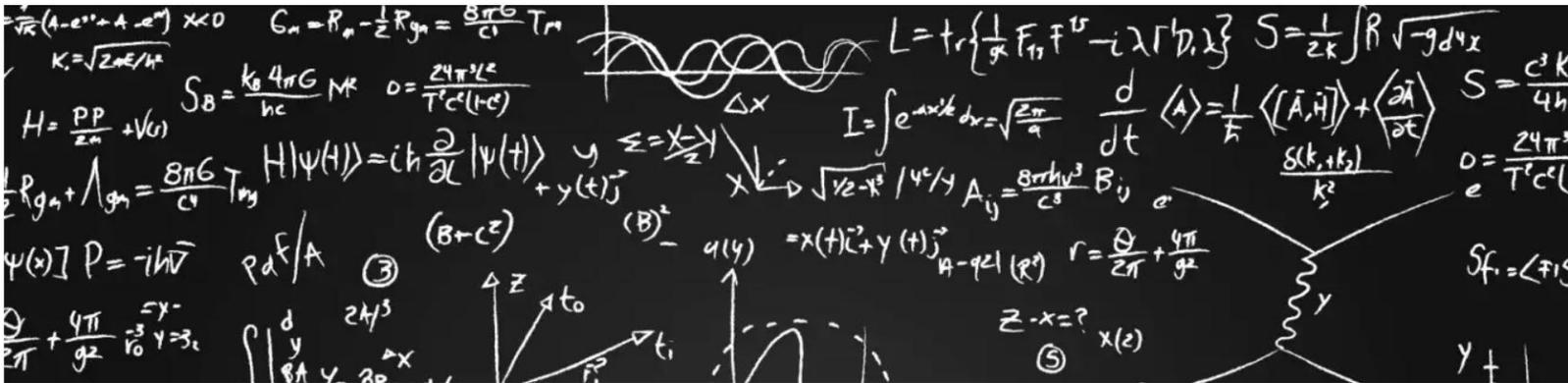
Mathematics

DeepMind and OpenAI claim gold in International Mathematical Olympiad

Two AI models have achieved gold medal standard for the first time in a prestigious competition for young mathematicians – and their developers claim these AIs could soon crack tough scientific problems

By [Alex Wilkins](#)

 22 July 2025



Google Deepmind's AGI Safety Course Starts With This



AGI safety course

GDM AGI Safety Course

by GDM Safety Research

Playlist • 15 videos • 5,485 views

A short course from Google DeepMind on AGI safety, covering alignment problems we can expect as AI ...more

▶ Play all

🔖

↶

⋮

1

AGI safety course

3:40

2

Two claims

Claim 1: Meaningful progress is being made towards superhuman capabilities.
Claim 2: It is plausible that this progress continues to superhuman levels.

4:24

3

Three types of risks

Miracles
Doing something with an unforeseen side effect

Learned heuristics
Executing behaviors that were previously reinforced, but are actually undesirable

Deliberate planning
Choosing actions that seem to be optimal

6:41

Instrumental subgoals

Consider pursuing a goal that (1) benefits from resources we'd (2) not want to burn on it

Part 0: Introduction

GDM Safety Research • 394 views • 1 month ago

Part 1: 1. We are on a path to superhuman capabilities

GDM Safety Research • 785 views • 1 month ago

Part 1: 2. Risks from deliberate planning and instrumental subgoals

GDM Safety Research • 655 views • 1 month ago

Part 1: 3. Exercise 1: Instrumental subgoals

GDM Safety Research • 527 views • 1 month ago

How do we respond as educators?

The 3 stages of AI in teaching

Stage 1: Denial

- “AI can’t do what I teach!”

ChatGPT scored poorly on my Fall, 2022 Labor Economics midterm. A D, to be precise. The performance was so poor compared to the hype that I publicly bet Matthew Barnett that no AI would be able to get A’s on 5 out of 6 of my exams by January of 2029. Three months have passed since then. GPT-4 has been released. Bet On It reader Collin Gray has kindly used GPT-4 to re-run the same test.

To my surprise and no small dismay, GPT-4 got an A. It earned 73/100, which would have been the fourth-highest score on the test. The most natural explanation to my mind was that my blog post made it into the new training data, but multiple knowledgeable friends assure me that there is no new training data.

Bryan Caplan, March 2023
Professor of economics at George Mason University

The 3 stages of AI in teaching

Stage 2: Avoidance

- "I'll just use exams for everything!"
- But is that really what we want to teach and assess?

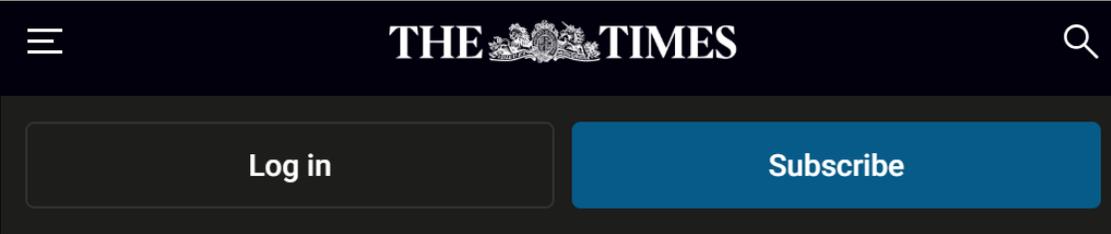
Artificial Intelligence +



 by Editorial Team
May 29, 2025, 3:51 am

Handwritten Exams Return Amid AI Concerns

Handwritten exams are making a comeback in higher education as colleges and universities in the United States respond to concerns over academic integrity related to artificial intelligence



Home UK World Comment Life & Style Business & Money S

Glasgow students to sit exams in person over AI cheating fears

Concerns over the influence of AI has meant universities are setting handwritten exams again

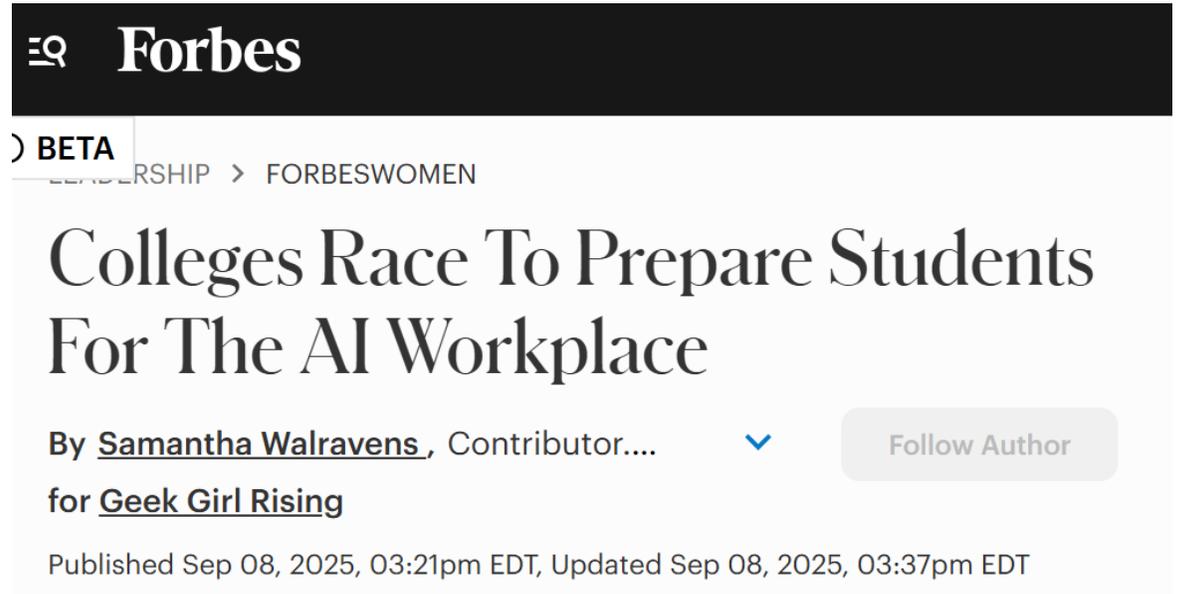
UK universities warned to 'stress-test' assessments as 92% of students use AI

Survey of 1,000 students shows 'explosive increase' in use of generative AI in particular over past 12 months

The 3 stages of AI in teaching

Stage 3: Acceptance

- “AI isn’t going to go away. What should I actually teach and how?”
- There are lots of opportunities to use AI to improve *the way that we teach*. We need to share what works with others in Cambridge and elsewhere.
- Part of that may involve teaching AI skills. These are key whether students go into the workplace or into further research. But what about everything else?
- In terms of what else I teach: Calculators in schools analogy a good place to start.

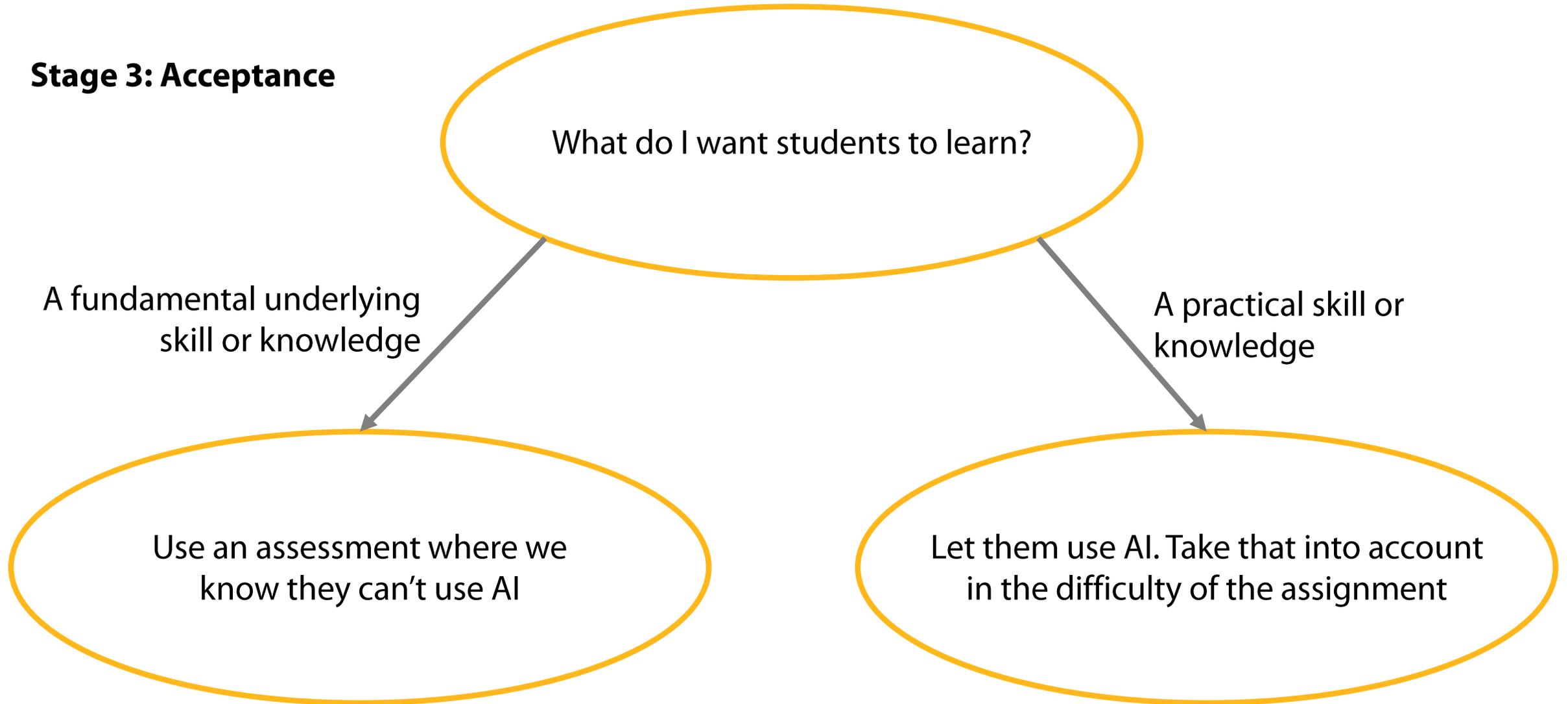


The screenshot shows the top portion of a Forbes article. The Forbes logo is in the top left. Below it, the text 'BETA' is visible, followed by 'MEMBERSHIP > FORBESWOMEN'. The main title of the article is 'Colleges Race To Prepare Students For The AI Workplace'. Below the title, it says 'By Samantha Walravens, Contributor...' with a dropdown arrow and a 'Follow Author' button. At the bottom of the article preview, it states 'Published Sep 08, 2025, 03:21pm EDT, Updated Sep 08, 2025, 03:37pm EDT'.

At Vanderbilt, adjunct professor Leonora Williamson has gone even further: “I tore up everything I have ever done and re-did it to make AI the center of everything. On the first day of class, I told students that one of my jobs is to prepare them for human-AI teams. If they’re not using AI for every assignment, I want them in office hours so I can show them how.”

The 3 stages of AI in teaching

Stage 3: Acceptance



The 3 stages of AI in teaching

Stage 3: Acceptance

Ideas for assessments:

- Reflective essay
- Critique/evaluate the output of a GenAI model
- Viva
- Give a presentation (although students will use AI to create slides and create a script, but perhaps the Q&A is most important.)
- Simulation or role play (negotiation exercise, boardroom simulation)
- Have a debate (students can use GenAI to research the topic, but they don't know what issues will come up during the debate.)

ChatGPT's Key Takeaways

Oral / interactive formats work well in *all* disciplines — they expose real understanding.

Process-focused methods (logs, drafts, reflections) make outsourcing to AI difficult.

Applied tasks tie learning to real-world contexts, where AI outputs are too generic.

AI-inclusive critiques are forward-looking — students learn to work *with* AI critically.

Multi-modal outputs encourage creativity and personalisation, which resist automation.

ChatGPT's suggestions for assessment methods for different disciplines

Discipline Area	Oral & Interactive	Process-Oriented	Applied / Experiential	AI-Inclusive / Critical	Multi-Modal / Creative
Business & Management	Boardroom simulation, pitch defence, negotiation role-play	Reflective journals on leadership style, meeting logs	Live consultancy projects with local firms	Use AI to draft a business plan, then critique/revise	Podcast on industry trends, video pitch
STEM (Science, Tech, Engineering, Maths)	Viva on research project, coding interview, peer teaching	Lab notebooks with raw data + reflection	Hackathon, prototype development, field experiments	Evaluate AI-generated code or equations	Infographics explaining scientific concepts, screencast walkthrough
Humanities (History, Literature, Philosophy)	Seminar debates, oral textual analysis	Annotated bibliographies, weekly reflective diaries	Archival project, local history exhibition, museum collaboration	Compare AI essay with own argument, identify omissions	Creative artefacts: zine, digital storytelling, short film
Social Sciences (Politics, Sociology, Psychology, Economics)	Panel discussions, mock parliamentary hearings	Research diaries, iterative drafts with peer review	Policy briefs for NGOs, ethnographic mini-projects	Test AI-generated survey questions, critique bias	Poster sessions, interactive data visualisation
Law	Moot courts, oral case defences	Case analysis logs, iterative reasoning notes	Drafting contracts for real or simulated clients	Critique AI's contract drafting or case analysis	Recorded "client interviews", policy explainer videos
Medicine & Health Sciences	OSCE-style stations, oral case presentations	Clinical reflection logs, patient interaction notes	Community health projects, clinical placements	Compare AI diagnostic suggestion with evidence-based practice	Health education pamphlet, instructional video
Creative Arts & Design	Portfolio defence, peer critiques	Process documentation of creative choices	Real-world design briefs, installations	Use AI to generate drafts/sketches and explain improvements	Exhibitions, multimedia projects
Education	Micro-teaching sessions, lesson plan defence	Reflective teaching journals, draft evolution	School-based projects, curriculum design tasks	Analyse AI lesson plan vs. student-centred design	Instructional videos, podcast on teaching philosophy

Applying Insights to Your Teaching

Session Takeaways / Key Insights

- AI can already do a lot
- More Capable AI Models Are (Still) Coming. “This is the worst AI model you’ll ever use.”
- AI is Becoming More Agentic
- This is beginning to have work impacts. The Earliest Job Impacts are on Software Engineering
- 3 stages of AI in teaching: Denial -> Avoidance -> Acceptance

Key Reflection Questions

1. What do we actually want students to learn?
2. How, therefore, can I assess that?

Thank you!

David Stillwell
d.stillwell@jbs.cam.ac.uk

Professor of Computational Social Science
Academic Director of the Psychometrics Centre
Cambridge University Judge Business School

