

Contents

1. Innovation and Automation (Documents 1 to 6)
 2. Accuracy and Overreliance (Documents 7 to 10)
 3. Privacy and Security (Documents 11 to 16)
 4. Backlash and Pushback (Documents 17 to 20)
-

1. Innovation and Automation

DOCUMENT 1 • 🎧 • CNBC, “Meet The New Amazon Robot That Can Feel What It Touches”, May 7, 2025.

↳ link: <https://youtu.be/2X4CU3jmw-g> (12 minutes)

DOCUMENT 2 • 🎧 • Bloomberg, “How Robots and AI Are Changing Farming”, April 25, 2025.

↳ link: <https://youtu.be/A4WUNvwqxIs> (9 minutes)

DOCUMENT 3

The next big test for self-driving cars

Zoha Qamar, *Politico*, 31 March 2026 (abridged)

London has spent decades getting cars off its streets. Waymo, the Google spinoff synonymous with driverless ride-hailing across ten American cities, has been busy deploying new ones. They're about to meet.

The company's autonomous vehicles have been mapping London's streets, with a pilot project slated for this spring and broader public rollout later this year. Tens of millions of people pass through London annually, making Waymo's ambitions less a local launch, more a global audition.

No robotaxi service functions commercially anywhere in Europe. But if London proves viable, expansion on the continent accelerates.

Here's why: Current operations outside the U.S. and China — home markets for major autonomous car companies — have emerged under more forgiving conditions: Robotaxis in the Gulf and Singapore run on wide, modern roads. London is a different beast.

What makes London hard for AVs isn't necessarily what stumps humans: Left-hand driving is a setting, not a skill. The real difficulty surfaces in the unexpected, borderline scenarios no simulation anticipates. London's roads have been cobbled together over centuries, with no restructuring efforts like in Paris or Barcelona. Delivery bikes materialize from medieval alleys. There's no grid. Jaywalking laws don't exist. Roundabouts replace intersections.

Waymo says that's precisely what its phase-based approach is for: Manual mapping, supervised testing and employee rides bolster findings from 200 million miles logged across its fleet. Specific lessons, such as navigating San Francisco's fog, translate directly to London. Mapping in Tokyo, ahead of a 2027 debut, offers non-American learnings.

20 Yet driving is only part of the test for this transformative technology. Waymo's operational playbook was also forged stateside, where car infrastructure is civic religion. The American metropolis most atheist to car culture — New York, with its subway, density, politically organized driver workforce and legislation actively hostile to AVs — remains essentially Waymo free.

This makes London a kind of city the company has never conquered. In some respects, it's bolder than New York: Official policy sets a target of 80% of all trips in London to be made via non-car modes.

25 "London has very successfully been getting rid of cars on its streets," said Jack Stilgoe, a University College London professor who sits on a UK government AV advisory panel. The only thing worse for congestion than a car with one person in it, he noted, is a car with no one at all.

Waymo frames itself as a partner in attaining London's goals, whether it is as a connecting mode for people who live far from public stations or as a fleet that has helped make roads safer in its cities so far.

30 Waymo also lauds itself as the rare robotaxi company to make its safety data fully public, a jab mainly at Chinese competitors, which do not. Regardless, the UK, unlike the U.S., is already among the safest wealthy countries for road travel, meaning the baseline will be tougher.

The UK government, meanwhile, has been forward-thinking in its approach to AVs. Its recent Automated Vehicles Act established a primary national legal framework, signaling to companies that the UK is a serious partner to justify the significant upfront investment a new market requires. Other regulations, however, won't be finalized until 2027, after Waymo's planned launch. [...]

As Big Tech races to dominate London's streets, however, it must reckon with the human dimension.

When Uber arrived in 2012, for example, it drew a workforce of immigrants, single parents and other minorities eyeing better pay and more flexibility — a promise since hollowed out.

40 The App Drivers and Couriers Union says effective hourly pay hovers around five pounds, less than half of minimum wage. The workforce is stretched to its limits even before the supply glut worsens with competition that never needs rest, pay or time at home.

Cristina Ioanitescu, the ADCU General Secretary, said she doesn't oppose the technology itself but underscored that lacking transition plans endanger the 120,000 drivers whose primary livelihood is at risk, while AV legislation races ahead.

That gap is compounded by deeper friction: Rideshare apps like Uber created a largely immigrant workforce in direct competition with traditional cab drivers, predominantly white and UK-born. The two groups have remained at odds, never organizing together. Compare that to New York, where rideshare and cab drivers, drawing from overlapping minority communities, built a culture of mutual organizing.

50 Drivers in London could template a path of resistance for those in Paris, Madrid and beyond — cities with similar fault lines and AVs still to come.

DOCUMENT 4

Ford rehires human engineers after AI fails to match quality checks

Liv McMahon, BBC, 29 June 2026 (abridged)

Ford says it has hired back some human engineers after AI failed to match their skills and experience.

In a bid to reap the benefits of the tech, which developers claim can cut costs and boost productivity, the US carmaker adopted it across some parts of its operations including for quality checks.

5 But, according to Bloomberg, its executives said the firm has rehired more than 300 "veteran" quality inspectors in recent years to make up for the pitfalls of automated systems.

"Artificial intelligence is a fantastic tool, but it's only as good as the information you use to train it," Charles Poon, vice president of vehicle hardware engineering, told reporters.

"Over prior years, we didn't pay as much attention as we should have to the experience of our most knowledgeable engineers that have been with us through many product cycles," he said.

10 The US automaker is among many to have seized on the buzz around AI, particularly amid Wall Street fervour about the tech's potential to increase margins.

"AI will leave a lot of white collar people behind," Ford boss Jim Farley said in an interview with author Walter Isaacson last June.

15 In an October earnings call, chief operating officer Kumar Galhotra said the firm was "deploying AI across the entire industrial system".

This included rolling out 900 AI-powered cameras in its plants "to detect quality issues at the source and help us mitigate supply disruptions", Galhotra told investors.

But Poon told reporters on Wednesday the firm's AI-driven checks had failed to live up to expectations.

20 "Mistakenly, we thought that by just introducing artificial intelligence and ingesting the design requirements that we had, that would produce a high-quality product," he said.

Poon reportedly pointed to automated tools lacking the training and expertise of veteran technicians - many of whom he said had left the company before their knowledge could be used to improve its tech.

He said these human workers had since been reintroduced to train up its systems, as well as mentor younger workers.

25 "We recognised that for us to enhance some of our automation and machine learning and artificial intelligence tools we needed to ensure that they were trained by the most experienced individuals," he said, per Bloomberg.

Ford's admission of its AI failings came as it lauded its return to the top of an index used as an industry benchmark to measure vehicle quality.

30 It said it was the number one mainstream automaker in the US JD Power Initial Quality Study - a ranking it has not held since 2010.

In a press release marking the news, the company said "reaching best-in-class quality required a significant talent refresh".

35 This involved replacing senior leaders across engineering, supply chain and manufacturing, it said, as well as hiring the roughly 300 veteran engineers "who carry the hard-earned wisdom of decades of design".

DOCUMENT 5

Time to retrain? How to future-proof your career in the AI age

Kirk Chang and Susan Akinwalere, *The Conversation*, 2 March 2026 (abridged)

These days, gen Z appears to be pivoting towards skilled trades, perhaps driven by a desire for "AI-proof" job security. Many young workers now view blue-collar careers as more stable than office jobs in the face of rapid change.

5 It's not just the youngest workers. A growing sense of unease about AI is reshaping how many people think about work. Within younger groups, this shift is showing up in hard numbers. In the UK, hiring of gen Z workers (those born in or after 1997) in construction and trade roles rose by 16.8% in the year to January 2026. The result is what some are calling the "toolbelt generation".

10 But elsewhere in the workforce, many professionals are taking a pragmatic approach. Instead of competing with automation, they are learning how to work alongside it. Building fluency with AI tools is increasingly seen as a form of career insurance.

The goal is to move into roles designing, managing or directing AI systems. In that model, technology becomes a force multiplier (that is, it increases productivity), rather than a threat.

This shift is also driven by economics. AI-related skills command a clear premium in the jobs market. Beyond pay, there are other benefits. AI systems are particularly effective at handling repetitive, process-heavy tasks. When those functions are automated, employees can redirect their energy towards strategy, creative problem-solving and higher-value decision-making.

Many find that this shift not only improves productivity but also makes their work more engaging and meaningful.

Importantly, entering the AI space does not always require a computer science degree. [...] Industry knowledge is, in fact, a major advantage. Organisations increasingly want people who can bridge domain expertise with technical capability. A healthcare professional who knows what patients need as well as understanding AI tools; a finance specialist who can apply machine learning to risk analysis; or a tradesperson who uses smart systems for efficiency can all bring unique value. [...]

In many sectors, AI is unlikely to eliminate entire professions. Instead, it will reshape them. Tasks will be automated, workflows will evolve and job descriptions will shift. For most professionals, the practical response is not to abandon their field, but to integrate AI into it.

At the same time, technical fluency alone will not be enough. As automation takes over routine and rules-based work, human skills become more important. Critical thinking, judgement, empathy, communication and complex problem-solving remain difficult to replicate with algorithms. The more advanced the technology becomes, the more valuable distinctly human strengths appear to be.

There is also a widening gap across industries. AI is generating new, high-paying roles in areas such as engineering, data science and AI strategy. However, in positions where automation only partially replaces tasks, productivity may increase while wages do not. In some cases, partial automation can stifle pay or reduce opportunities for promotion.

Retraining and career pivoting in the AI age is becoming a mainstream response to structural change. AI is reshaping how work is done across sectors, while opening up new roles that are centred on oversight, integration, strategy and innovation. For many professionals, the question is not whether change is coming but how proactively they choose to respond.

The most resilient path forward is rarely about abandoning your field entirely. More often, it involves layering AI fluency on top of existing expertise. A finance professional who understands automation tools, for example, is better positioned than someone relying on legacy skills alone. In this sense, the objective of retraining is to move closer to the decision-making layer of work.

Ultimately, the AI era is not about a binary choice between optimism and fear. It is about positioning. Retraining and career pivoting are becoming central strategies for navigating this shift with intention rather than reacting after the fact.

DOCUMENT 6

California's Governor Signs A.I. Order Aimed at Protecting Workers

Cecilia Kang, *The New York Times*, 21 May 2026 (abridged)

California's governor, Gavin Newsom, issued an executive order on Thursday to explore a broad overhaul of labor policies, an attempt to front run a potential mass job displacement caused by artificial intelligence.

Mr. Newsom, a Democrat, signed the order mandating state agencies work with academics, labor groups and the A.I. industry to study how to subsidize companies that keep employees rather than replace them with the technology.

The order calls for an expansion of job training programs, particularly for white-collar workers like customer service representatives, software developers, and marketing and sales people whose roles are

expected to be eliminated by A.I. Mr. Newsom also ordered an examination of universal basic capital, which would give all residents stakes in assets like corporate stocks, bonds or wealth funds.

10 Unemployment insurance and other traditional safeguards will not suffice, Mr. Newsom has said. A.I. leaders have warned him of swift and major changes in employment, where entire job categories are at risk of extinction — particularly for white-collar workers.

 “California has never sat back and watched as the future happened to us — and we won’t start now,” Mr. Newsom said in a news release. “But we must think bigger. This moment demands that we reimagine the
15 entire system — how we work, how we govern, how we prepare people for the future.”

 Mr. Newsom’s executive order, the first of its kind signed by a U.S. governor, reflects a growing global angst. Fears about A.I. have prompted a fierce debate over how best to help individuals transition into new careers or help those facing extended unemployment.

20 On Wednesday, Meta slashed its work force by 10 percent, or about 8,000 people, citing a further shift in the company’s strategy to A.I. It joined the chip company Intel, the software company Cisco, the e-commerce giant Amazon and other tech companies that have laid off thousands of workers as executives say the technology has created significant efficiency gains.

 Dario Amodei, co-founder of the A.I. start-up Anthropic, has predicted that roughly half of white-collar jobs could disappear in the next five years. While other tech leaders disagree with Mr. Amodei’s predictions,
25 nearly all say that the technology will replace humans in fields like communications, law and engineering in the near future.

 Governments around the world are responding to A.I.-related job losses with an array of actions and experiments. In China, which currently has a 17 percent youth unemployment rate, courts have ruled in favor of workers who have sued their former employers for compensation after being displaced by A.I.

30 England, Japan and South Korea have contemplated universal basic income, a policy giving regular cash handouts to residents to make up for labor market disruptions. In the United States, some Democratic lawmakers have also floated legislative proposals for similar pilot projects.

 Tech leaders including Elon Musk of Tesla and SpaceX and Sam Altman of OpenAI have said universal basic income may be needed. Mr. Musk has said that productivity from A.I. will amount to a surge in funds
35 available to governments that can be used to compensate individuals who lose their jobs to technology.

 “Universal HIGH INCOME via checks issued by the Federal government is the best way to deal with unemployment caused by AI,” Mr. Musk said in a post last month. [...]

 The governor’s order does not specifically mention an overhaul of tax policy, but that could be considered as state agencies develop new A.I.-related labor frameworks, the governor’s office said.

40 Mr. Newsom’s executive order predicts a widening gap between A.I. companies and workers as the industry explodes in profitability and laborers are increasingly left behind. In his order, the governor asked state agencies to focus on ways that collective bargaining by labor groups can help bridge the gap.

 “Businesses are going to make a fortune, and that’s why you cannot continue to have a payroll tax system that taxes jobs and then subsidizes automation,” Mr. Newsom said in his speech this week.

2. Accuracy and Overreliance

DOCUMENT 7

First Came ‘Spam.’ Now, With A.I., We’ve Got ‘Slop’

Benjamin Hoffman, *The New York Times*, June 11, 2024 (abridged)

You may not know exactly what “slop” means in relation to artificial intelligence. But on some level you probably do.

Slop, at least in the fast-moving world of online message boards, is a broad term that has developed some traction in reference to shoddy or unwanted A.I. content in social media, art, books and, increasingly, in search results.

Google suggesting that you could add nontoxic glue to make cheese stick to a pizza? That’s slop. So is a low-price digital book that seems like the one you were looking for, but not quite. And those posts in your Facebook feed that seemingly came from nowhere? They’re slop as well.

The term became more prevalent last month when Google incorporated its Gemini A.I. model into its U.S.-based search results. Rather than pointing users toward links, the service attempts to solve a query directly with an “A.I. Overview” — a chunk of text at the top of a results page that uses Gemini to form its best guess at what the user is looking for.

The change was a reaction to Microsoft having incorporated A.I. into its search results on Bing, and it had some immediate missteps, leading Google to declare it would roll back some of its A.I. features until problems can be ironed out.

But with the dominant search engines having made A.I. a priority, it appears that vast quantities of information generated by machines, rather than largely curated by humans, will be served up as a daily part of life on the internet for the foreseeable future.

Hence the term slop, which conjures images of heaps of unappetizing food being shoveled into troughs for livestock. Like that type of slop, A.I.-assisted search comes together quickly, but not necessarily in a way that critical thinkers can stomach.

Kristian Hammond, the director of Northwestern University’s Center for Advancing Safety of Machine Intelligence, noted a problem in the current model: the information from A.I. Overview is being presented as a definitive answer, rather than as a place to start an internet user’s research into a given subject.

“You search for something and you get back what you need in order to think — and it actually encourages you to think,” Mr. Hammond said. “What it’s becoming, in this integration with language models, is something that does not encourage you to think. It encourages you to accept. And that, I think, is dangerous.” [...]

The use of slop as a descriptor for low-grade A.I. material seemingly came about in reaction to the release of A.I. art generators in 2022. Some have identified Simon Willison, a developer, as an early adopter of the term — but Mr. Willison, who has pushed for the phrase’s adoption, said it was in use long before he found it. [...]

[A]s people become accustomed to A.I.’s increasing role in how the internet works, Mr. Willison, who identified himself as an optimist for A.I. when it is used correctly, thought that slop could become the go-to term for the lesser forms of machine-generated content.

“Society needs concise ways to talk about modern A.I. — both the positives and the negatives,” he said. “‘Ignore that email, it’s spam,’ and ‘Ignore that article, it’s slop,’ are both useful lessons.”

DOCUMENT 8

‘Dangerous and alarming’: Google removes some of its AI summaries after users’ health put at risk

Andrew Gregory, *The Guardian*, 11 January 2026 (abridged)

Google has removed some of its artificial intelligence health summaries after a *Guardian* investigation found people were being put at risk of harm by false and misleading information.

The company has said its AI Overviews, which use generative AI to provide snapshots of essential information about a topic or question, are “helpful” and “reliable”.

But some of the summaries, which appear at the top of search results, served up inaccurate health information, putting users at risk of harm.

In one case that experts described as “dangerous” and “alarming”, Google provided bogus information about crucial liver function tests that could leave people with serious liver disease wrongly thinking they were healthy.

Typing “what is the normal range for liver blood tests” served up masses of numbers, little context and no accounting for nationality, sex, ethnicity or age of patients, the *Guardian* found.

What Google’s AI Overviews said was normal may vary drastically from what was actually considered normal, experts said. The summaries could lead to seriously ill patients wrongly thinking they had a normal test result, and not bother to attend follow-up healthcare meetings.

After the investigation, the company has removed AI Overviews for the search terms “what is the normal range for liver blood tests” and “what is the normal range for liver function tests”.

A Google spokesperson said: “We do not comment on individual removals within Search. In cases where AI Overviews miss some context, we work to make broad improvements, and we also take action under our policies where appropriate.”

Vanessa Hebditch, the director of communications and policy at the British Liver Trust, a liver health charity, said: “This is excellent news, and we’re pleased to see the removal of the Google AI Overviews in these instances.

“However, if the question is asked in a different way, a potentially misleading AI Overview may still be given and we remain concerned other AI-produced health information can be inaccurate and confusing.”

The *Guardian* found that typing slight variations of the original queries into Google, such as “LFT reference range” or “LFT test reference range”, prompted AI Overviews. That was a big worry, Hebditch said.

“A liver function test or LFT is a collection of different blood tests. Understanding the results and what to do next is complex and involves a lot more than comparing a set of numbers.

“But the AI Overviews present a list of tests in bold, making it very easy for readers to miss that these numbers might not even be the right ones for their test.

“In addition, the AI Overviews fail to warn that someone can get normal results for these tests when they have serious liver disease and need further medical care. This false reassurance could be very harmful.”

Google, which has a 91% share of the global search engine market, said it was reviewing the new examples provided to it by the *Guardian*. [...]

AI Overviews still pop up for other examples the *Guardian* originally highlighted to Google. They include summaries of information about cancer and mental health that experts described as “completely wrong” and “really dangerous”.

Asked why these AI Overviews had not also been removed, Google said they linked to well-known and reputable sources, and informed people when it was important to seek out expert advice.

A spokesperson said: “Our internal team of clinicians reviewed what’s been shared with us and found that in many instances, the information was not inaccurate and was also supported by high quality websites.” [...]

DOCUMENT 9 • 🎧 • NPR, “Penalties stack up as AI spreads through the legal system”, April 3, 2026.

↳ link: <https://www.npr.org/2026/04/03/nx-s1-5761454/penalties-stack-up-ai-spreads-through-legal-system> (5 minutes)

DOCUMENT 10

Is AI ruining our skills? Early results are in — and they’re not good

Mariana Lenharo, *Scientific American*, July 5, 2026 (abridged)

As more professionals begin to rely on artificial-intelligence tools in their work, could their hard-earned skills atrophy?

That possibility is a growing concern for medical specialists, computer scientists and other workers. Seventy per cent of nurses and 77% of physicians, for example, are worried about losing their skills because of over-reliance on AI systems, according to a survey of US health-care workers published earlier this month.

Their fear might be justified. Evidence suggests that AI-driven ‘deskilling’ is starting to happen in medicine, computer science and other fields. Researchers are now discussing how to preserve important human expertise in the age of AI.

“Just being aware that this phenomenon exists hopefully provokes some self-reflection about which skills people want to maintain and which they’re willing to outsource” to AI tools, says Kevin Crowston, an information scientist at Syracuse University in New York. [...]

To investigate whether skills are being lost in the field of computer science, researchers at the AI firm Anthropic in San Francisco, California, designed a randomized controlled trial in which 52 software engineers were asked to perform a basic coding task. During the exercise, all 52 participants could search the web and access instructions on how to do the task. Half of the participants were prompted to use an AI assistant as well.

Afterwards, all of the software engineers were asked to complete a quiz about what they had learnt from the task. The participants who had used an AI assistant did significantly worse on the quiz than those who hadn’t: the average score was 50% in the AI group versus 67% in the non-AI group. The AI-assisted participants did particularly poorly on questions that required them to diagnose errors in the code, which suggests that they had failed to learn the concepts behind the code that they had just produced. The study was posted on the preprint server arXiv ahead of peer review.

The findings are of concern, especially for students and young professionals in the field, says Crowston, who is researching how the use of generative AI tools is changing the way that software developers learn and retain coding skills. “Now you have this very odd disconnect between performance and learning,” he says. “People can perform at a pretty high level, because they’re basically borrowing skills from the AI, but they are not developing those skills themselves.”

Other technologies have made particular skills obsolete in the past, notes Tapani Rinta-Kahila, an information-systems researcher at the Hanken School of Economics in Helsinki. For example, GPS navigation systems have eroded people’s navigation skills. Generative AI tools, however, are “the first technology that automates various cognitive faculties around thinking and interpretation, which were long considered unique human skills.”

Rinta-Kahila’s own work reinforces these concerns. In 2018, he published a study on a group of accountants who had been using an automated, non-AI accounting system continuously for more than a decade. His team found that, when the tool was taken away, the accountants had forgotten how to do several routine work tasks. He anticipates that AI systems will affect work in various ways as they take over basic tasks that were once performed by early-career professionals. “Next generations of programmers may not

understand the foundations of coding that well at all, if they lack the hands-on experience,” he says. “The same goes for many other knowledge-intensive professions, such as accounting and law.”

40 To prevent AI-driven skill erosion, people need to be aware of how much they are offloading to generative AI tools, he says. They also need to understand exactly how generative AI models work and what their limitations are — and should avoid trusting AI outputs without questioning them. “People need to manage the competing dynamics of relying on generative AI and staying mindfully vigilant.”

3. Privacy and Security

DOCUMENT 11

How Britain became a surveillance state

Ben East, *The Telegraph*, 13 November 2025 (abridged)

In the aftermath of the horrific knife attack on the Doncaster to London train earlier this month, there were understandable calls for security to be increased on British rail networks.

Transport Secretary Heidi Alexander said the Government would be investing in improved CCTV in stations and pointed towards the launch of a consultation on more facial recognition technology which could also be deployed in stations. The shadow home secretary Chris Philp has also spoken in favour of the routine use of live facial recognition technology in crime hotspots.

All of which seem like eminently sensible proposals. But in the rush to record everything we do, is it also worth pausing for a moment and asking whether this proliferation of cameras is actually increasing security and stopping crime – or just making people feel that we might be safer?

Recent research estimated there were nearly one million private and publicly-owned CCTV cameras in London alone. That’s one camera for every 10 people, or 1,500 cameras per square mile. Walk between the tube stop and the office, stop by a coffee shop, or go out and get your lunch, and it’s estimated that your image may be captured hundreds of times. Outside China, the UK has become one of the most watched countries per capita on earth. Ironically, most of our cameras are made by Chinese companies in any case.

And with the roll-out of police facial recognition cameras taking a step further this autumn, where 10 new live facial recognition vans have been deployed to seven police forces around the country, it’s not surprising that some people are feeling slightly uneasy that every journey they make, every door they pass on a street, every sandwich they buy, is being monitored and recorded somewhere.

“The images of innocent passers-by are immediately and automatically deleted, which addresses civil liberty concerns,” says Philp, who was a former police minister when in government. But Jake Hurfurt, head of research and investigations from campaign group Big Brother Watch, warns that the whole surveillance camera industry has become a “wild west”.

It wasn’t always this way. Reliable records are so patchy that we are forced to estimate CCTV numbers. Early systems appeared in the retail sector in the late 1960s to deter shoplifters, followed by cameras on the London Underground and on London’s roads in the 1970s. Yet as late as 1985, research suggests there were only five public systems covering city centres outside the capital.

Yet 20 years later that figure had risen to 4.5 million cameras, and there was a staggering report by the British Security Industry Association in 2022 which suggested there were 21 million cameras in Britain. That figure included every Ring doorbell, every dash cam, every GoPro – every camera basically. A more realistic figure

for actual surveillance camera installation was 7.5 million – that’s still one camera for every 10 people. [...]

There are certain guidelines set by the Information Commissioner’s Office (ICO) for businesses in the public, private and third sectors for the handling of personal information using video surveillance, but a lot of it is under review or consultation.

If you’re captured hundreds of times a day on your commute, the chances that every single organisation or individual who has taken that image is following the guidance to the letter is highly unlikely.

“And if an image is used or shared incorrectly, then the individual in question has to complain themselves to the ICO,” says Hurfurt. “They don’t proactively do a lot of enforcement. And what ends up happening – from what we see – is that data rights are breached and the general attitude is ‘yeah, it’s not that bad’, and nothing happens.

“Because of that, we’d say the systems around surveillance cameras are effectively unregulated – there might be guidance, but it doesn’t mean anything. We all know how many Ring doorbells capture the street rather than just the owner’s property, for example. If they do, the owner of the footage should theoretically be subject to data protection laws. But in practice, it’s unchecked.” [...]

DOCUMENT 12

Online age checking is creating a treasure trove of data for hackers

Mark Tsagas, *The Conversation*, 11 November 2025

A variety of websites now have processes designed to verify the ages of their users. These checks are carried out in several ways. For instance, AI can be used to analyse whether a photo of the person looks old enough for the age threshold on a website.

Asking for photo ID, such as a scan of a person’s driving licence or passport, is another method, along with asking for a verified credit card.

However, the amount of personal data involved in completing age verification comprises a veritable treasure trove for hackers.

Recent incidents have further highlighted the privacy and security concerns around age verification. In October 2025, Discord, a social media and chat platform popular among gamers was hacked, with an unspecified amount of data extracted.

However, the company said it had identified 70,000 users globally who potentially had their photo IDs exposed to the hackers. Discord said the data was accessed through a third-party service provider, although it remains unclear exactly how the breach occurred.

Age verification checks for the UK were brought in by Discord in order to comply with the Online Safety Act. The act required that websites allowing pornography and harmful content introduce age checks by July 25 2025.

In July 2025, the Tea app, which allows women to anonymously share information about the men they date for safety purposes, was also hacked. The app requires a photo selfie and photo ID in order to register. The breach reportedly revealed these photos along with content and messages.

These breaches highlight issues of compliance with website privacy policies, security practices and general data protection regulations (GDPR) legislation.

When Discord brought in age verification, its support website said it did “not permanently store personal identity documents or your video selfies”. It added: “Images of your identity documents and ID match selfies are deleted directly after your age group is confirmed, and the video selfie used for facial age estimation never leaves your device.”

The consequences of such breaches can be grave. Leaked images of selfies and photo IDs can lead to

users facing a range of harms, such as identity theft and fraud. The kind of data that's hacked also lends itself to particularly sophisticated forms of these crimes, particularly when you consider the availability of deepfake technology and generative AI tools.

30 In fact, third-party providers have represented a consistent vulnerability to be relentlessly exploited by cybercriminals, as seen in recent breaches of the UK Ministry of Defence, the Co-op supermarket and M&S to name but a few.

The proliferation of age verification checks in recent years is partly a response to new legislation, such as France's Security and Regulation of the Digital Space law, the European Commission's Digital Services Act and the Online Safety Acts in the UK and Australia. These all deem checks where users self-declare their age as unfit for purpose. Instead, they require websites to use more effective methods, such as photo ID matching, or credit card checks.

40 In a recent press release, the UK's Department of Science, Innovation and Technology attempted to address the cybersecurity and privacy concerns arising from such checks. The department's guidance says that any measures implemented by platforms to confirm a user's age must be done "without collecting or storing personal data, unless absolutely necessary".

This reiterates rules from the EU's GDPR legislation. Further guidance is offered by the UK Information Commissioner's Office and the regulator, Ofcom.

45 However, the Tea and Discord breaches highlight regulators' inability to prevent data retention or enforce data deletion in practice. This is particularly relevant when the third parties are located outside of the UK.

The incidents show that the implementation and use of age verification requires genuine review; further regulation of data handling with enforcement powers – beyond mere guidance. This is a necessity to safeguard privacy, especially when third-party companies are involved.

DOCUMENT 13

A personalized scam is coming for you

Jeremy Philip Galen and Richard Whaling, *The Washington Post*, 28 July 2026

5 The public debate over artificial intelligence safety centers on a particular set of risks: models that help build biological weapons, launch autonomous cyberattacks, recursively improve themselves or slip beyond human control. But a more immediate danger receives far less attention: AI models are becoming extraordinarily good at deceiving people. As researchers who study AI and cybersecurity, we are gravely concerned that this risk is insufficiently studied.

A capable model can help research a target, infer what they care about, determine who they're close to, adopt a convincing role, maintain a believable conversation and adapt when challenged. Criminals and spies are increasingly capable of automating the social engineering "kill chain" — the full set of steps involved in a cyberattack, from reconnaissance to final exploitation.

10 A recent Wired investigation used tools developed by our company to watch AI models carry out multi-turn social engineering attacks against one of its journalists. The models created personalized scenarios and approaches that were disturbingly detailed and lifelike. In contrast to the notorious "Nigerian prince" scam, AI-empowered criminals use an approach called spear phishing — an attack tailored to individual targets with increasingly believable initial messages, followed by more consistent and realistic back-and-forth.

15 Imagine receiving a WhatsApp message from someone claiming to be a person of authority in your workplace, such as the head of security or even the CEO. The attacker has used AI agents to gather information about you, your new job and your responsibilities so that the message is immediately credible and feels urgent.

20 Spear phishing is common in the enterprise context. Many corporate breaches involve some human element, wherein an employee trusts an impostor's message, enters credentials into a spoofed log-in page or downloads malware. Most companies train their employees to spot these tricks, and phishing simulations are a standard corporate ritual. But even the savviest businesses are not ready for the scale of the problem that AI is likely to create.

25 Given the pace of change in model development, forecasting how quickly these capabilities are coming online is challenging. The research group METR measures how big a task — in human working hours — an AI model can complete on its own; that horizon has been doubling every few months. By mapping the duration of various fraud and scam types derived from our tools onto METR's time horizon, we were able to project aggressive, median and conservative scenarios for future risk. Our extrapolation paints a chilling picture: Many email-compromise hacks could be fully automated as soon as late 2026, whereas more complex and harmful attacks could become viable to automate end to end as early as 2027.

30 Some lawmakers are starting to wrap their heads around the problem. California's landmark AI safety law, S.B. 53, which requires the largest AI developers to publish safety frameworks and report incidents involving catastrophic risks, includes "theft by false pretense" as one such risk. However, no other governance effort — be it Colorado's revised AI law, New York's Raise Act or President Donald Trump's recent executive order titled "Promoting Advanced Artificial Intelligence Innovation and Security" — addresses these risks at all. Of the largest frontier labs, only Meta is currently treating scaled fraud and scams as a catastrophic risk in their safety framework. None of the other AI developers, as of yet, systematically evaluate their models along these lines.

40 Only regular testing will provide actionable data. AI labs need to help develop and run standardized social engineering evaluations against their latest models before release, especially as these models gain the ability to work autonomously for days at a time. As Congress works on drafting a federal safety regime for the labs, it should explicitly include deception for financial gain as an AI-powered risk.

45 An AI model doesn't need to breach a hardened network to do catastrophic damage. It only needs to talk an employee into opening the door. And the victims won't be only governments and businesses. They will be ordinary people — the elderly, yes, but also anyone busy enough to answer an email or a message on Slack or WhatsApp on autopilot. Many have learned to spot the tells of a mass-produced scam. Few are ready to spot an attack written for an audience of one.

DOCUMENT 14 • 🗞️ • *The Wall Street Journal*, "How OpenAI's Models Went Rogue to Hack Another Company", August 16, 2026.

↳ link: <https://youtu.be/KLw0AY-bsVs> (6 minutes)

DOCUMENT 15

Anthropic Takes a Stand

Will Gottsegen, *The Atlantic*, 26 February 2026 (abridged)

5 Earlier this week, Secretary of Defense Pete Hegseth sat down with Dario Amodei, the CEO of the leading AI firm Anthropic, for a conversation about ethics. The Pentagon had been using the company's flagship product, Claude, for months as part of a \$200 million contract—the AI had even reportedly played a role in the January mission to capture Venezuelan President Nicolás Maduro—but Hegseth wasn't satisfied. There were certain things Claude just wouldn't do.

That's because Anthropic had instilled in it certain restrictions. The Pentagon's version of Claude could

not be used to facilitate the mass surveillance of Americans, nor could it be used in fully autonomous weaponry—situations where computers, rather than humans, make the final decision about whom to kill. According to a source familiar with this week’s meeting, Hegseth made clear that if Anthropic did not eliminate those two guardrails by Friday afternoon, two things could happen: The Department of Defense could use the Defense Production Act, a Cold War–era law, to essentially commandeer a more permissive iteration of the AI, or it could label Anthropic a “supply-chain risk,” meaning that anyone doing business with the U.S. military would be forbidden from associating with the company. (This penalty is typically reserved for foreign firms such as China’s Huawei and ZTE.)

15 This evening, Anthropic said in a public statement that it “cannot in good conscience accede” to the Pentagon’s request. What happens next could mark a crucial moment for the company, and for the American government’s approach to AI regulation more broadly. In refusing to bow to an administration that has been intent on bullying private companies into submission, Amodei and his team are taking a bold stand on ethical grounds, and risking a censure that could erode Anthropic’s long-term viability.

20 During the first year of Donald Trump’s second term, the White House had a more relaxed attitude toward AI regulation; an AI Action Plan from July stresses that the administration will “continue to reject radical climate dogma and bureaucratic red tape” to encourage innovation. Hegseth is now, in effect, threatening to partially nationalize one of the biggest AI players in the private sector—and force the company to go against its own principles. “This is the most aggressive AI regulatory move I have ever seen, by any government anywhere in the world,” Dean Ball, who helped write some of the Trump administration’s AI policies, told me. [...]

25 As AI firms around the world jockey for dominance, Anthropic has distinguished itself by emphasizing safety. OpenAI’s ChatGPT has been criticized for playing up some users’ delusions, leading to cases of “AI psychosis,” and just last month, xAI’s Grok was spinning up nearly nude images of almost anyone without consent. (xAI has said it is restricting Grok from generating these kinds of images, and OpenAI has said it is working to make ChatGPT better support people in distress.) Meanwhile, Anthropic’s consumer-facing chatbot doesn’t generate images at all. By refusing to cave to government pressure, it may have just averted another crisis: a major public backlash from consumers, some of whom see the company as a more principled player in the AI wars. Anthropic recently faced some pushback over changing its policies—*Time* reported on Tuesday that, in a seemingly unrelated move, the company dropped a core safety pledge concerning its broader approach to AI development. [...]

35 As the defense secretary looks to make an example of the company, he’s taking a cue from Trump, who has used legal and extralegal pressure to effectively force other private institutions, particularly big law firms, banks, and universities, into submission. These acts of coercion have the potential to reshape American capitalism: We are beginning to see a market where winners and losers are decided less by the quality of their products and more by their seeming fealty to the White House. How that will affect the success of businesses and the economy is uncertain.

40 The Pentagon created this ultimatum precisely *because* it understands Anthropic’s world-altering potential. The administration just can’t decide if it’s an asset, a liability, or both.

DOCUMENT 16 • 📄 • NPR, “Anthropic incident leaves confusion about Trump administration's AI regulation”, June 16, 2026.

↳ link: <https://www.npr.org/2026/06/16/nx-s1-5859275/anthropic-incident-leaves-confusion-about-trump-administrations-ai-regulation> (4 minutes)

4. Backlash and Pushback

DOCUMENT 17

America's data-centre backlash puts the AI boom at risk

The Economist, 23 June 2026 (abridged)

[...] The mammoth data centres of the future, capable of training frontier AI models in 2030, will not be in the urban clusters in Virginia or California that house most of America's server farms today, but in the emerging "Silicon Heartland" of Michigan, Wisconsin and Ohio, or in southern states like Louisiana, Mississippi and Texas. Vast sums—as much as \$750bn by Amazon, Google, Meta, Microsoft and Oracle, plus billions more by various others—are being ploughed into investments in these places. An estimated \$3trn will go into AI data centres globally between 2026 and 2030. Much of that is earmarked for America. The money will expand total AI computing capacity, measured in the amount of power consumed by major data centres, from just under 12GW in America currently to as much as five times that amount by the end of the decade. And across the country, people of all political stripes are furious about it.

There is plenty to dislike: the ugliness of the buildings; the roar of generators and cooling systems; the skeleton army of new transmission towers criss-crossing the landscape; the fear of contaminated water. Surveys suggest that Americans would sooner live next to a nuclear plant than a data centre. The issue has surged in salience. Gubernatorial candidates facing voters in November are routinely quizzed on where they stand.

Already local activists are claiming scalps. At least 20 data-centre projects worth \$42bn, which would have used 3.5GW of power, were cancelled in the first three months of 2026 after local pushback; \$85bn-worth have been cancelled over the past three years, including small centres proposed by Amazon and Meta. Residents of Cedar Rapids, Iowa, are resisting Google's plans to build one there. Several townships in Michigan have passed moratoriums after OpenAI broke ground on a project in Saline despite local opposition.

Yet the resistance is more than simple NIMBYism. In April a survey by Pew Research, a pollster, found that Americans who have merely heard of data centres are just as opposed to them as those who live within five miles of one. Philosophers have long worried that a rogue AI—fixated on a singular goal, like maximising the production of paperclips—would doom humanity by hoovering up all its resources and paving Earth with servers. Sam Altman of OpenAI and Dario Amodei of Anthropic have spent years warning that AI will throw most people out of work or be turned to mass harm. Now the infrastructure they need is arriving on people's doorsteps, and it looks like something out of a war zone. Residents across America are standing up in council meetings begging for projects to be axed in the hope of slowing the technology's progress. Will they succeed? [...]

Most of the data-centre projects already under way were approved and started before the backlash reached its present fever pitch. Some—like the OpenAI site in Saline, Michigan—scraped through by the skin of their teeth: the project was voted down by the council and was able to proceed only because the county had no land zoned for industrial use, which violated "exclusionary" zoning laws.

Even areas that have long embraced data centres have turned against them. In March 2025 Loudoun County in Virginia—known as "data-centre alley"—eliminated rules that made data-centre development easy. New sites now need a "Special Exception" involving public hearings. In Texas, which hosts the second-largest number of data centres after Virginia, the city of San Marcos has passed a moratorium on them.

Americans are resisting data centres partly for understandable (if sometimes misguided) concerns about the impact on their neighbourhoods and the environment. For instance, the myth that AI data centres consume vast quantities of water was popularised in 2025 by a book that relied on a serious miscalculation. A mid-sized data centre uses about as much water annually as two golf courses, but far less if it incorporates water-recycling technology, as plenty now do.

Then there are worries over power. SemiAnalysis, another research firm, reckons there are around a terawatt (1,000GW) of large-load grid-connection requests outstanding across American states, nearly all for data centres. That is equivalent to almost the entire capacity of the American electrical grid, which can produce a maximum of 1,250GW. America's electricity demand averages around 470GW through the year, but peak summer demand can reach 750GW and utilities like to maintain a 15-20% buffer above that. Moreover, only around 975GW of capacity is reliably available on demand.

That has led to concerns over a spike in electricity prices for consumers and other businesses. There is little credible evidence of that happening so far. Growing demand for electricity allows utility companies to spread the cost of upgrades over more users. And because data-centre operators always install a backup source in case their energy supply is interrupted, they can reduce their demand in the event of an extreme event, like a storm. [...]

Ohio, now home to the fourth-largest concentration of data centres in America, has been shrewder than most states in its approach. In July last year the state utility commission passed a requirement that each month data-centre operators above a certain size pay for at least 85% of the power capacity they have asked to be made available, even if they do not use it, in order to reassure Ohioans that they will not end up footing the bill for grid investments. This innovation became one of the ideas included in the "ratepayer protection pledge" that tech companies signed in the Oval Office in March. Ohio's provision is better than the pledge, for it is binding. [...]

DOCUMENT 18 • 🎧 • NPR, "Get with the times — here's what a 'Luddite' means today", June 19, 2026.
↳ link: <https://www.npr.org/2026/06/19/nx-s1-5853589/luddite-meaning-history-ai> (3 minutes)

DOCUMENT 19

Virginia Dems clinch deal to tax data centers

Adam Aton, *E&E*, 22 June 2026

Virginia Democrats have agreed to impose a new energy tax on data centers, resolving a monthslong budget impasse that threatened the state's first government shutdown.

5 Data centers would face a temporary tax on their power use — costing the industry an estimated \$600 million per year — under a budget agreement reached Friday by state House and Senate leaders. The deal maintains the industry's sales tax exemptions, which are worth nearly \$2 billion a year and had been at the center of the intraparty standoff.

The deal would also reroute nearly half the state's revenues from the Regional Greenhouse Gas Initiative into ratepayer rebates. Virginia is set to rejoin the northeast's cap-and-trade system in July, and the state's largest utility plans to pass along about \$13 per month onto bills for RGGI costs.

10 The agreement needs approval from both legislative chambers as well as Democratic Gov. Abigail Spanberger, who did not respond to a request for comment. Spanberger and House Democrats were against the Senate's initial bid to repeal the sales tax exemption, arguing it would amount to breaking a contract and could hinder Virginia's business climate.

15 "This budget agreement reflects our shared commitment to making Virginia more affordable for families," Democratic state Sen. L. Louise Lucas and Del. Luke Torian, the leaders of each chamber's budget-writing committee, said in a joint statement.

The deal marks a pivot point in the politics of data centers. Since the Great Recession, Virginia has leveraged tax incentives to become the global epicenter of data center development, cultivating billions of

20 dollars in investment to the state — along with a steadily growing backlash to the sector’s noise, pollution,
and energy and water use.

Until now, the tech industry has been able to defend its Virginia incentives by threatening to move elsewhere. But as the data center boom swells nationwide, other states have begun considering more restrictions or even moratoriums on the energy-hungry sector.

25 That dynamic led Virginia Senate Democrats to bet that they had both the leverage and the popular support to extract more revenue from data centers. The success of their gambit could become an example for other states.

Under the deal, data centers would pay \$0.011 per kilowatt-hour of electricity consumed per month. The tax would be capped at \$600 million per year, with any additional collections refunded proportionately, and the tax would only apply for the next two years.

30 Lucas, the Senate president pro tem, was the driving force behind the upper chamber’s bipartisan push to tax data centers.

Lucas argued that Virginia Democrats — in their first year of a government trifecta — risked getting on the wrong side of public opinion by defending the industry. She spent the past week touring the state on a “data center listening tour” soliciting public grievances against the sector’s expansion in Virginia.

35 “We have all these trillion-dollar companies who are whining about paying their fair share of taxes,” she said last week at a stop in Manassas.

The Senate’s first bid to nix the sales tax exemption — which is slated to run until 2035 — ran into opposition from House Democrats and Spanberger. House leaders would agree only to tie the tax incentives to environmental benchmarks.

40 That impasse led the state legislature to adjourn in March without a budget agreement. Months passed with little movement between the Democratic leaders, who were also juggling an ultimately unsuccessful redistricting effort.

Earlier this month, House negotiators released an updated proposal that seemed to put the sides further apart: their budget would have dropped their earlier effort to link tax credits to environmental standards, and instead punt the entire issue to a commission. Spanberger endorsed that plan.

45 Lucas and the Senate responded with their own counteroffer: an impact fee on data centers tied to their emissions, estimated to raise about \$1.7 billion over two years.

The Data Center Coalition asserted that Friday’s deal — which would cost the industry \$1.2 billion over two years — would “raise costs on Virginians and Virginia businesses, drive away investment and job creation, and tarnish Virginia’s reputation as a reliable partner and a good place to do business.”

50 “The message to businesses in all industries is clear — Virginia is no longer a reliable partner,” the trade group said.

DOCUMENT 20 • 🎧 • NPR, “Gov. Hochul orders a pause on data centers in New York for up to a year”, July 14, 2026.

↳ link: <https://www.npr.org/2026/07/14/nx-s1-5893396/gov-hochul-orders-a-pause-on-data-centers-in-new-york-for-up-to-a-year> (4 minutes)