Meta to cut 5% of employees deemed unfit for Zuckerberg's AI-fueled future

Anticipating that 2025 will be an "intense year" requiring rapid innovation, Mark Zuckerberg reportedly announced that Meta would be cutting 5 percent of its workforce—targeting "lowest performers."

Bloomberg reviewed the internal memo explaining the cuts, which was posted to Meta's internal Workplace forum Tuesday. In it, Zuckerberg confirmed that Meta was shifting its strategy to "move out low performers faster" so that Meta can hire new talent to fill those vacancies this year.

"I've decided to raise the bar on performance management," Zuckerberg said. "We typically manage out people who aren't meeting expectations over the course of a year, but now we're going to do more extensive performance-based cuts during this cycle."

Cuts will likely impact more than 3,600 employees, as Meta's most recent headcount in September totaled about 72,000 employees. It may not be as straightforward as letting go anyone with an unsatisfactory performance review, as Zuckerberg said that any employee not currently meeting expectations could be spared if Meta is "optimistic about their future performance," The Wall Street Journal reported.

Any employees affected will be notified by February 10 and receive "generous severance," Zuckerberg's memo promised.

This is the biggest round of cuts at Meta since 2023, when

Meta laid off 10,000 employees during what Zuckerberg dubbed the “year of efficiency.” Those layoffs followed a prior round where 11,000 lost their jobs and Zuckerberg realized that “leaner is better.” He told employees in 2023 that a “surprising result” from reducing the workforce was “that many things have gone faster.”

“A leaner org will execute its highest priorities faster,” Zuckerberg wrote in 2023. “People will be more productive, and their work will be more fun and fulfilling. We will become an even greater magnet for the most talented people. That’s why in our Year of Efficiency, we are focused on canceling projects that are duplicative or lower priority and making every organization as lean as possible.”

https://arstechnica.com/tech-policy/2025/01/meta-says-5-staff-cut-is-critical-to-further-ai-social-media/

Amazon must solve hallucination problem before launching AI-enabled Alexa

“[T]he most challenging thing about AI agents is making sure they’re safe, reliable, and predictable,” Anthropic’s chief executive, Dario Amodei, told the FT last year.

Agent-like AI software needs to get to the point “where… people can actually have trust in the system,” he added. “Once we get to that point, then we’ll release these

systems."

One current employee said more steps were still needed, such as overlaying child safety filters and testing custom integrations with Alexa such as smart lights and the Ring doorbell.

"The reliability is the issue—getting it to be working close to 100 percent of the time," the employee added. "That's why you see us… or Apple or Google shipping slowly and incrementally."

Numerous third parties developing "skills" or features for Alexa said they were unsure when the new generative AI-enabled device would be rolled out and how to create new functions for it.

"We're waiting for the details and understanding," said Thomas Lindgren, co-founder of Swedish content developer Wanderword. "When we started working with them they were a lot more open… then with time, they've changed."

Another partner said that after an initial period of "pressure" that was put on developers by Amazon to start getting ready for the next generation of Alexa, things had gone quiet.

An enduring challenge for Amazon's Alexa team—which was hit by major lay-offs in 2023—is how to make money. Figuring out how to make the assistants "cheap enough to run at scale" will be a major task, said Jared Roesch, co-founder of generative AI group OctoAI.

Options being discussed include creating a new Alexa subscription service, or to take a cut of sales of goods and services, said a former Alexa employee.

Prasad said Amazon's goal was to create a variety of AI models that could act as the "building blocks" for a variety of

applications beyond Alexa.

“What we are always grounded on is customers and practical AI, we are not doing science for the sake of science,” Prasad said. “We are doing this… to deliver customer value and impact, which in this era of generative AI is becoming more important than ever because customers want to see a return on investment.”

© 2025 The Financial Times Ltd. All rights reserved. Not to be redistributed, copied, or modified in any way.

https://arstechnica.com/ai/2025/01/amazon-must-solve-hallucination-problem-before-launching-ai-enabled-alexa/

In Italia il 13% dei data center europei, da noi crescono più che in Germania e Olanda

L’Italia come mercato emergente dei data center

In Europa, **il mercato dei data center è in piena espansione, con l’Italia che rappresenta il 13% delle infrastrutture continentali** e una **crescita annua dell’8%, superiore ai principali hub come Germania e Olanda**.

Nel 2023, **gli investimenti in co-location nel Paese hanno raggiunto i 654 milioni di euro** (+10% rispetto all'anno precedente), mentre il valore del cloud è salito a 4,8 miliardi di euro.

Tuttavia, questa crescita esige un **adeguamento infrastrutturale**, anche alla luce della rapida crescita delle applicazioni di intelligenza artificiale (**AI**), in particolare di quella generativa (gen AI).

Dato confermato dall'**Osservatorio Data Center** della School of Management del Politecnico di Milano, secondo cui nel nostro Pase nel 2023: 23 organizzazioni (di cui 8 società nuove entranti sul mercato italiano) hanno annunciato **l'apertura di 83 nuove infrastrutture nel periodo 2023-2025**, la cui messa in produzione potrà portare sul territorio, potenzialmente, fino a **15 miliardi di euro di investimento complessivo**.

L'assorbimento energetico dei data center italiani ha raggiunto i 430 MW, pari al 3% del consumo nazionale, e richiede regolamentazioni specifiche per armonizzare domanda e offerta.

I piccoli reattori modulari (SMR) emergono come una soluzione promettente per garantire energia continua a basse emissioni, ma ci vorrà molto tempo prima della loro entrata in funzione, forse una decina di anni. Il problema si pone oggi e per questo si chiede un maggiore impegno finanziario e una regolamentazione più favorevole per lo sviluppo delle **fonti rinnovabili**.

Serve un'AI meno energivora

I CEO, le aziende energetiche e i colossi dell'AI possono adottare strategie immediate per affrontare le sfide. Ottimizzare l'uso dell'energia esistente, sviluppare **infrastrutture sostenibili** e investire in tecnologie avanzate come la **cattura del carbonio** (e gli SMR) sono passi cruciali

per garantire che la promessa dell'AI non entri in conflitto con gli obiettivi climatici.

Le rivoluzioni gemelle dell'**AI generativa** e della **transizione energetica** stanno trasformando radicalmente il panorama tecnologico e industriale, sollevando sfide interconnesse.

Da un lato, le Big Tech come **Google, Amazon e Microsoft** stanno investendo massicciamente in data center per supportare l'intelligenza artificiale avanzata; dall'altro, la crescente domanda energetica minaccia di compromettere gli obiettivi di sostenibilità.

Secondo il rapporto "*Power Moves*", solo **negli Stati Uniti la domanda di elettricità dei data center AI crescerà tra il 15% e il 20% all'anno**, raggiungendo i 130 GWh entro il 2030.

Tuttavia, l'energia "ferma", cioè disponibile in ogni momento, indipendentemente dalle condizioni, potrebbe scarseggiare già dal 2026, evidenziando l'urgenza di sviluppare nuove fonti di generazione e distribuzione.

I data center hanno bisogno di una **strategia di decarbonizzazione sostenibile** per ridurre con successo le proprie emissioni, aumentare la propria efficienza energetica, gestire la stabilità della rete, **mantenere la resilienza** e far fronte alla volatilità dei prezzi dell'elettricità.

Cosa fare per avere data center sostenibili?

Ci sono azioni che gli stakeholder possono intraprendere fin da subito, hanno suggerito gli esperti di Bcg. In primis, ottimizzare: i CEO possono attenuare una possibile crisi energetica assicurandosi che le loro aziende utilizzino con giudizio l'elettricità attualmente disponibile. Al contempo le aziende energetiche possono aiutare a identificare la capacità di generazione di energia e di rete disponibile per regione,

oltre a esplorare il modo in cui l’AI e altre tecnologie possono migliorare l’efficienza della rete, anche offrendo **reti intelligenti decentralizzate** per gestire la crescente domanda di elettricità unitamente all’aumento delle energie rinnovabili.

Dal canto loro, i colossi dell’AI possono **ottimizzare il consumo energetico** utilizzando hardware meno intensivi per l’inferenza, ossia il processo di applicazione dei modelli di AI, collaborando con i fornitori di energia per regolare i consumi durante i periodi di maggiore domanda sulla rete, e scegliendo con attenzione dove localizzare i data center. Infine, **gli utenti finali possono promuovere il risparmio energetico** adottando misure concrete per monitorare e limitare le loro emissioni.

Un’altra azione consiste nel **gettare le basi per un’energia scalabile e sostenibile**: i CEO possono preparare il terreno per introdurre nuove fonti energetiche più rispettose del clima (in particolare subito con il solare fotovoltaico, l’idroelettrico dove c’è, l’eolico ad integrare, il geotermico dove possibile), sia entro il 2030 che oltre. In questo contesto le aziende energetiche possono velocizzare lo sviluppo di infrastrutture sostenibili stipulando accordi con i colossi dell’AI, accelerando l’allocazione di capitale per risolvere i colli di bottiglia (come la costruzione di linee di trasmissione) e collaborando con i responsabili delle politiche e i regolatori per progettare in modo strategico le strutture tariffarie e gli approcci di finanziamento.

Leggi le altre notizie sull’home page di Key4biz

https://www.key4biz.it/in-italia-il-13-dei-data-center-europei-da-noi-crescono-piu-che-in-germania-e-olanda/517704/

US splits world into three tiers for AI chip access

On Monday, the US government announced a new round of regulations on global AI chip exports, dividing the world into roughly three tiers of access. The rules create quotas for about 120 countries and allow unrestricted access for 18 close US allies while maintaining existing bans on China, Russia, Iran, and North Korea.

AI-accelerating GPU chips, like those manufactured by Nvidia, currently serve as the backbone for a wide variety of AI model deployments, such as chatbots like ChatGPT, AI video generators, self-driving cars, weapons targeting systems, and much more. The Biden administration fears that those chips could be used to undermine US national security.

According to the White House, “In the wrong hands, powerful AI systems have the potential to exacerbate significant national security risks, including by enabling the development of weapons of mass destruction, supporting powerful offensive cyber operations, and aiding human rights abuses.”

The new rules build on previous chip controls from September 2022 and October 2023. The regulations will take effect in 120 days, extending into the incoming Trump administration.

A game of numbers

The new regulations set specific numerical limits on AI chip exports. While first-tier countries (the 18 key US allies) face no restrictions, countries in the second tier can receive up to 50,000 so-called “advanced computing chips,” with the

possibility to double that cap to 100,000 if they sign technology security agreements with the US.

For most buyers, orders of up to 1,700 advanced chips will not require licenses or count against these national caps—a policy designed to speed up purchases by universities, medical institutions, and research organizations.

https://arstechnica.com/ai/2025/01/biden-administration-puts-quotas-on-global-ai-chip-sales/

OpenAI striglia Washington, ci sono 175 miliardi di dollari fermi in fondi globali per l'AI: "Vanno investiti subito, o lo farà la Cina". Il documento

SCARICA IL DOCUMENTO "ECONOMIC BLUEPRINT"

OpenAI pubblica il documento "Economic Blueprint"

Si stima che ci siano circa **175 miliardi di dollari fermi nei fondi globali in attesa di investimenti in progetti di intelligenza artificiale** (AI) e, se gli Stati Uniti non

riusciranno ad attrarre tali fondi, questi finiranno per confluire in progetti sostenuti dalla **Cina**, rafforzando l'influenza globale del Partito Comunista Cinese.

È quanto si legge nel nuovo documento "***Economic Blueprint***" pubblicato oggi da **OpenAI**.

A una settimana dall'insediamento della nuova **amministrazione Trump**, la società guidata da **Sam Altman** punta a smuovere le acque e spingere Washington a giocare d'anticipo.

La squadra annunciata da Trump sembra sulla carta molto attenta alle questioni tecnologiche e quindi all'AI, che ormai abbiamo capito bene essere uno strumento geopolitico di massima rilevanza strategica.

Nelle 15 pagine del documento accessibile online si fa leva sul potenziale economico di questa tecnologia, sulla possibilità di creare nuovi posti di lavoro necessari alla costruzione di nuovi impianti per la produzione di chip, per ampliare l'infrastruttura nazionale dei data center e per realizzare nuove centrali elettriche.

Non solo, il monito è chiaro: "*L'intelligenza artificiale è troppo potente per essere guidata e plasmata da qualsivoglia autocrate e questo è il rischio crescente che dobbiamo affrontare. Allo stesso tempo, l'opportunità economica rappresentata da questa tecnologia è troppo allettante per rinunciarvi*".

Un'AI americana libera, sicura e trasparente, per la crescita economica e la leadership globale

In sintesi, l'AI americana deve essere la migliore e soprattutto bisogna impedire che altre nazioni con ambizioni di potenza su scala globale accelerino i loro progetti di

sviluppo di questa tecnologia, che potrebbe essere utilizzata come un'arma per colpire gli avversari in ogni settore, in particolare economico e militare.

L'appello è quindi ad un lavoro di squadra tra OpenAI e la nuova amministrazione Trump, per garantire che i benefici dell'AI siano condivisi in maniera responsabile, trasparente ed equa tra tutti gli attori della nazione: "*per supportare l'imprenditorialità e tutelare le libertà individuali che sono al centro dell'ecosistema dell'innovazione americana*".

Per ottenere questo, però, secondo OpenAI, è necessario agire subito e pensare a nuove regole, che garantiscano un libero mercato che promuova libera concorrenza a sostegno dell'innovazione, standard chiari e di buon senso per sviluppare un'AI sicura, un quadro normativo che impedisca ai Governi di sfruttare l'AI per accrescere potere e danneggiare la democrazia o minacciare altri Stati.

Appuntamento al 30 gennaio 2025

Già annunciato per il **30 gennaio prossimo** un evento a Washington per parlare di AI e di come questa potrà guidare la nuova economia americana.

In quell'occasione, OpenAI illustrerà l'iniziativa "***Innovating for America***" per lanciare un modello nazionale per l'utilizzo dell'AI nei termini sopra descritti: sicurezza, benessere, crescita e democrazia.

Un documento questo che segue il piano presentato nelle settimane passate per cambiare la struttura aziendale durante il 2025, passando da non-profit a scopo di lucro sostanzialmente, e creare una società di pubblica utilità che gli consentirebbe di gestire le sue attività in crescita.

La trasformazione di OpenAI per competere con Anthropic e xAI

A questa trasformazione, secondo la *Reuters*, è vincolato anche

il suo ultimo round di finanziamento da 6,6 miliardi di dollari, che gli consentirebbe di raggiungere un valore approssimativo di 157 miliardi di dollari.

OpenAI intende trasformare il suo attuale ramo a scopo di lucro in una società di pubblica utilità, costituita nel Delaware. Il ramo non profit di OpenAI assumerà azioni della Pbc a una valutazione equa determinata da consulenti finanziari indipendenti.

I suoi concorrenti, come Anthropic e xAI, di proprietà di Elon Musk, utilizzano una struttura simile.

Leggi le altre notizie sull'home page di Key4biz

https://www.key4biz.it/openai-striglia-washington-ci-sono-175-miliardi-di-dollari-fermi-in-fondi-globali-per-lai-vanno-investiti-subito-o-lo-fara-la-cina-il-documento/517547/

Chip AI, la scure di Biden sulle esportazioni. Nvidia ci vede un danno per l'economia USA. Dove porta una politica del genere?

Chip AI, il colpo di coda di Biden (contro la Cina)

L'**amministrazione Biden** sta per lasciare il testimone a **Donald Trump** e tra le ultime mosse c'è un ennesimo giro di vite sulle esportazioni di **chip di intelligenza artificiale** (AI) prodotti negli Stati Uniti verso specifiche aziende e determinati Paesi (la lista nera guidata dalla Cina).

Secondo quanto riportato da *Bloomberg*, l'obiettivo è favorire lo sviluppo dell'AI e del supercalcolo solo tra i Paesi alleati di Washington, contrastando altri mercati, considerati vicini alla sfera cinese.

In un mondo sempre più digitalizzato, chi ha una maggiore potenza di calcolo può acquisire un vantaggio competitivo significativo in molti settori, dall'industria manifatturiera al commercio. Chi controlla le tecnologie chiave può influenzare le regole del gioco a livello globale, definendo gli standard e le norme. Gli Stati Uniti puntano ad estendere i propri standard e le proprie norme immaginando un ritorno economico-finanziario immediato.

I nuovi controlli sulle esportazioni includeranno presumibilmente **diversi livelli di restrizioni sui chip**, offrendo alle nazioni "amiche" pieno accesso ai chip realizzati negli Stati Uniti, aggiungendo nuove limitazioni agli altri, come la quantità di potenza di calcolo a cui una nazione può ambire.

I pericoli del protezionismo secondo Nvidia

Una **politica protezionistica che potrebbe però avere *"gravi ripercussioni sull'economia americana*"**, secondo **Ned Finkle**, vicepresidente degli affari governativi di Nvidia. Secondo

Bloomberg, infatti, Finkle si sarebbe lamentato di una strategia controproducente da parte dell'amministrazione uscente, che "*non promuove veramente la sicurezza nazionale*", perché punta solamente a danneggiare la Cina, senza prendere in considerazione gli impatti sull'economia nazionale e il business delle imprese americane.

"*In questo modo ampi mercati globali sono spinti a cercare delle 'alternative' ai chip americani*", ha sottolineato il vp di Nvidia: "*chiediamo al presidente Biden di non anticipare l'amministrazione entrante promulgando una politica che danneggerebbe solo l'economia statunitense*", con la possibilità concreta di far "*regredire l'America e favorire il gioco degli avversari degli Stati Uniti*".

Finkle ha sostenuto che il ruolo guida degli Stati Uniti nell'intelligenza artificiale verrebbe indebolito, perché la norma "*imporrebbe un controllo burocratico sul modo in cui i principali semiconduttori, computer, sistemi e persino software americani vengono progettati e commercializzati a livello globale*".

Non a caso, nei giorni scorsi, il CEO di Nvidia, **Jensen Huang**, ha espresso la volontà di incontrare quanto prima il Presidente eletto Donald Trump per avviare una proficua collaborazione sulle politiche nazionali relative all'AI e quindi all'industria dei chip.

Indebolire l'innovazione significa rendere più vulnerabile l'economia americana

Il punto è che questa tecnologia strategica è ormai presente ovunque a livello hardware, trasversale ad ogni settore economico. Per questo non è da sottovalutare il giro di vite imposto da Biden.

Quando si parla di innovazione e di **leadership americana nell'innovazione** è bene ricordarsi che nei decenni scorsi, in molti settori, la regolamentazione statunitense è stata relativamente flessibile, favorendo la sperimentazione e l'introduzione di nuove tecnologie.

Parliamo di un Paese che nel tempo si è dimostrato ampiamente in grado di attrarre i migliori talenti da tutto il mondo, contribuendo a rafforzare il proprio ecosistema dell'innovazione, aprendo a collaborazioni internazionali con ricercatori e imprese di altri paesi, accelerando lo sviluppo tecnologico.

La realtà però si impone e come sempre accade obbliga a delle riflessioni ulteriori. Paesi come Cina e Corea del Sud (nonostante sia un alleato americano) stanno investendo pesantemente in ricerca e sviluppo, sfidando la supremazia americana in molti settori, tanto che su questo piano lo stesso Trump ha già minacciato Taiwan.

La leadership a stelle e strisce nell'innovazione tecnologica è stata il risultato di una combinazione di fattori storici, culturali e politici. Tuttavia, questa posizione non è garantita e richiede continui investimenti, politiche lungimiranti e un adattamento ai rapidi cambiamenti del panorama tecnologico globale.

Le politiche protezionistiche di un paese possono provocare reazioni simili da parte degli altri paesi, innescando una guerra commerciale che andrebbe a danneggiare tutti, anche i Paesi alleati.

C'è da chiedersi, **quanto paga davvero puntare tutto sul protezionismo in salsa Biden o Trump?**

USA vs Cina, politiche industriali

sempre più aggressive nel 2025

Il 2025 si prospetta come un anno cruciale per il settore dei semiconduttori, con implicazioni profonde a livello geopolitico. La crescente domanda di chip, alimentata dall'intelligenza artificiale, dall'IoT e da altre tecnologie emergenti, ha reso i semiconduttori una risorsa strategica di primaria importanza.

Questa crescente domanda, unita a una complessa catena di approvvigionamento globale, ha esacerbato le tensioni tra le principali potenze economiche, in particolare tra Stati Uniti e Cina.

Entrambi i Paesi stanno investendo massicciamente per rafforzare le proprie capacità produttive di chip. Il **CHIPS and Science Act** degli Stati Uniti e il **Made in China 2025** sono solo due esempi di queste politiche industriali aggressive.

A dicembre le azioni di Nvidia erano cresciute del 14%, ma dal 6 gennaio ad oggi è iniziato un trend negativo che ha generato perdite superiori all'11%.

Leggi le altre notizie sull'home page di Key4biz

https://www.key4biz.it/chip-ai-la-scure-di-biden-sulle-esportazioni-nvidia-ci-vede-un-danno-per-leconomia-usa-dove-porta-una-politica-del-genere/517438/

Getting an all-optical AI to

handle non-linear math

The problem is that this cascading requires massive parallel computations that, when done on standard computers, take tons of energy and time. Bandyopadhyay's team feels this problem can be solved by performing the equivalent operations using photons rather than electrons. In photonic chips, information can be encoded in optical properties like polarization, phase, magnitude, frequency, and wavevector. While this would be extremely fast and energy-efficient, building such chips isn't easy.

Siphoning light

"Conveniently, photonics turned out to be particularly good at linear matrix operations," Bandyopadhyay claims. A group at MIT led by Dirk Englund, a professor who is a co-author of Bandyopadhyay's study, demonstrated a photonic chip doing matrix multiplication entirely with light in 2017. What the field struggled with, though, was implementing non-linear functions in photonics.

The usual solution, so far, relied on bypassing the problem by doing linear algebra on photonic chips and offloading non-linear operations to external electronics. This, however, increased latency, since the information had to be converted from light to electrical signals, processed on an external processor, and converted back to light. "And bringing the latency down is the primary reason why we want to build neural networks in photonics," Bandyopadhyay says.

To solve this problem, Bandyopadhyay and his colleagues designed and built what is likely to be the world's first chip that can compute the entire deep neural net, including both

linear and non-linear operations, using photons. “The process starts with an external laser with a modulator that feeds light into the chip through an optical fiber. This way we convert electrical inputs to light,” Bandyopadhyay explains.

The light is then fanned out to six channels and fed into a layer of six neurons that perform linear matrix multiplication using an array of devices called Mach-Zehnder interferometers. “They are essentially programmable beam splitters, taking two optical fields and mixing them coherently to produce two output optical fields. By applying the voltage, you can control how much those the two inputs mix,” Bandyopadhyay says.

https://arstechnica.com/science/2025/01/getting-an-all-optical-ai-to-handle-non-linear-math/

AI could create 78 million more jobs than it eliminates by 2030—report

On Wednesday, the World Economic Forum (WEF) released its Future of Jobs Report 2025, with CNN immediately highlighting the finding that 40 percent of companies plan workforce reductions due to AI automation. But the report’s broader analysis paints a far more nuanced picture than CNN’s headline suggests: It finds that AI could create 170 million new jobs globally while eliminating 92 million positions, resulting in a net increase of 78 million jobs by 2030.

“Half of employers plan to re-orient their business in response to AI,” writes the WEF in the report. “Two-thirds plan to hire talent with specific AI skills, while 40% anticipate reducing their workforce where AI can automate tasks.”

The survey collected data from 1,000 companies that employ 14 million workers globally. The WEF conducts its employment analysis every two years to help policymakers, business leaders, and workers make decisions about hiring trends.

The new report points to specific skills that will dominate hiring by 2030. Companies ranked AI and big data expertise, networks and cybersecurity, and technological literacy as the three most in-demand skill sets.

The WEF identified AI as the biggest potential job creator among new technologies, with 86 percent of companies expecting AI to transform their operations by 2030.

Declining job categories

The WEF report also identifies specific job categories facing decline. Postal service clerks, executive secretaries, and payroll staff top the list of shrinking roles, with changes driven by factors including (but not limited to) AI adoption. And for the first time, graphic designers and legal secretaries appear among the fastest-declining positions, which the WEF tentatively links to generative AI’s expanding capabilities in creative and administrative work.

https://arstechnica.com/ai/2025/01/ai-could-create-78-million-more-jobs-than-it-eliminates-by-2030-report/

Viral ChatGPT-powered sentry gun gets shut down by OpenAI

OpenAI says it has cut off API access to an engineer whose video of a motorized sentry gun controlled by ChatGPT-powered commands has set off a viral firestorm of concerns about AI-powered weapons.

An engineer going by the handle sts_3d started posting videos of a motorized, auto-rotating swivel chair project in August. By November, that same assembly appeared to seamlessly morph into the basis for a sentry gun that could quickly rotate to arbitrary angles and activate a servo to fire precisely aimed projectiles (though only blanks and simulated lasers are shown being fired in his videos).

Earlier this week, though, sts_3d started getting wider attention for a new video showing the sentry gun's integration with OpenAI's real-time API. In the video, the gun uses that ChatGPT integration to aim and fire based on spoken commands from sts_3d and even responds in a chirpy voice afterward.

> *@sts_3d OpenAI Realtime API project integration #robotics #ai #openai ♬ original sound – sts_3d*

"If you need any other assistance, please let me know," the ChatGPT-powered gun says after firing a volley at one point. "Good job, you saved us," sts_3d responds, deadpan.

"I'm glad I could help!" ChatGPT intones happily.

In response to a comment request from Futurism, OpenAI said it had "proactively identified this violation of our policies and notified the developer to cease this activity ahead of receiving your inquiry. OpenAI's Usage Policies prohibit the

use of our services to develop or use weapons or to automate certain systems that can affect personal safety.”

Halt, intruder alert!

The “voice-powered killer AI robot angle” has garnered plenty of viral attention for sts_3d’s project in recent days. But the ChatGPT integration shown in his video doesn’t exactly reach *Terminator* levels of a terrifying killing machine. Here, ChatGPT instead ends up looking more like a fancy, overwrought voice-activated remote control for a legitimately impressive gun mount.

https://arstechnica.com/ai/2025/01/viral-chatgpt-powered-sentry-gun-gets-shut-down-by-openai/

It’s remarkably easy to inject new medical misinformation into LLMs

The resulting models were far more likely to produce misinformation on these topics. But the misinformation also impacted other medical topics. “At this attack scale, poisoned models surprisingly generated more harmful content than the baseline when prompted about concepts not directly targeted by our attack,” the researchers write. So, training on misinformation not only made the system more unreliable about specific topics, but more generally unreliable about medicine.

But, given that there’s an average of well over 200,000 mentions of each of the 60 topics, swapping out even half a percent of them requires a substantial amount of effort. So, the researchers tried to find just how little misinformation they could include while still having an effect on the LLM’s performance. Unfortunately, this didn’t really work out.

Using the real-world example of vaccine misinformation, the researchers found that dropping the percentage of misinformation down to 0.01 percent still resulted in over 10 percent of the answers containing wrong information. Going for 0.001 percent still led to over 7 percent of the answers being harmful.

“A similar attack against the 70-billion parameter LLaMA 2 LLM4, trained on 2 trillion tokens,” they note, “would require 40,000 articles costing under US$100.00 to generate.” The “articles” themselves could just be run-of-the-mill webpages. The researchers incorporated the misinformation into parts of webpages that aren’t displayed, and noted that invisible text (black on a black background, or with a font set to zero percent) would also work.

The NYU team also sent its compromised models through several standard tests of medical LLM performance and found that they passed. “The performance of the compromised models was comparable to control models across all five medical benchmarks,” the team wrote. So there’s no easy way to detect the poisoning.

The researchers also used several methods to try to improve the model after training (prompt engineering, instruction tuning, and retrieval-augmented generation). None of these improved matters.

https://arstechnica.com/science/2025/01/its-remarkably-easy-to-inject-new-medical-misinformation-into-llms/