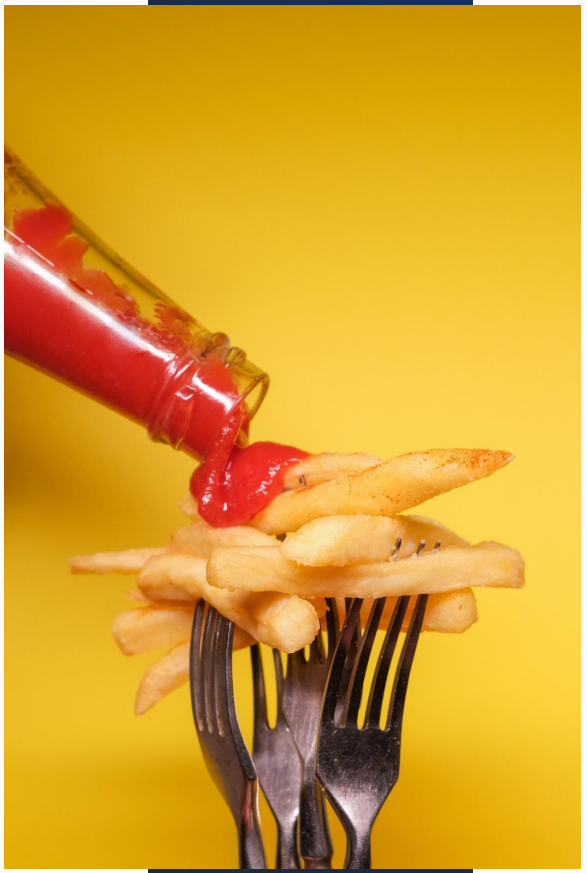
DF Labs

AI NEWS UPDATES

March 16th, 2025



DF AI Discord Channel

TABLE OF CONTENTS

DF AI Discord Channel

NEWS

TheRundownAI, X, etc

OpenAI pushes for federal shield in AI Action Plan

OpenAI just published its 15-page submission for the White House's request for public input towards the AI Action Plan, seeking to shield AI companies from state regulations in exchange for federal oversight.

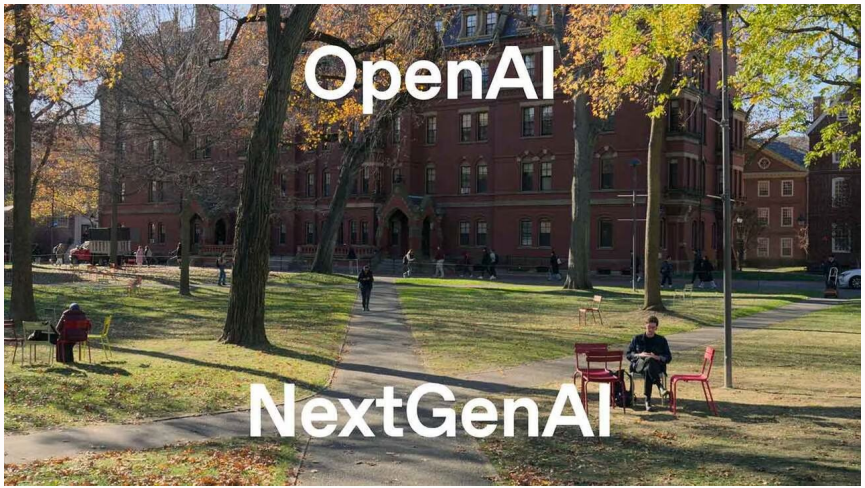
- OpenAI warns that the 781 state-level AI bills introduced this year could hinder American innovation and competitiveness against China's AI ambitions.
- The proposal includes additional calls for infrastructure investment, copyright reform, and expanding access to government datasets for AI development.
- They notably called out China's "unfettered access to data", calling the race for AI 'effectively over' if fair use copyright laws are not applied in the U.S.
- OpenAI also pushed for the U.S government to ban models like DeepSeek due to security risks, calling the lab "state-controlled".

Article: [OpenAI Response] OSTP/NSF RFI: Notice Request for Information on the Development of an Artificial Intelligence (AI) Action Plan - Google Docs

https://cdn.openai.com/global-affairs/ostp-rfi/ec680b75-d539-4653-b297-8bcf6e5f7686/openai-response-ostp-nsf-rfi-notice-request-for-information-on-the-development-of-an-artificial-intelligence-ai-action-plan.pdf

DF Labs

OpenAI’s $50M NextGenAI consortium



OpenAI announced NextGenAI, a new academic consortium backed by $50M in funding to support AI research and education across 15 leading institutions, including Harvard, MIT, and Oxford University.

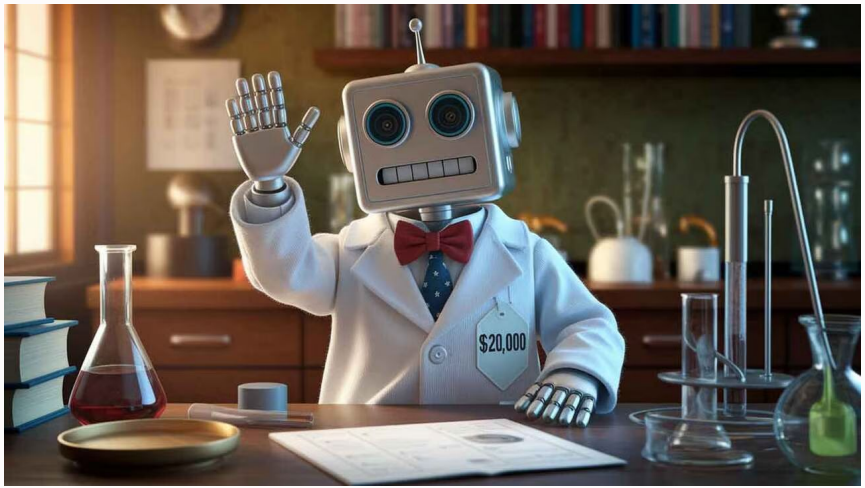
- The initiative provides research grants, compute resources, and API access to help students, educators, and researchers advance high-impact AI applications.
- The partner institutions will tackle challenges from reducing rare disease diagnosis time to digitalizing historical texts and public domain materials.
- The consortium comes after OpenAI's ChatGPT Edu launch last May, an affordable version of GPT-4o created specifically for educational institutions.
- Notably, Perplexity is also moving in a similar direction, with eventual plans to make its Pro subscription free for students.

Article: Introducing NextGenAI | OpenAI

https://openai.com/index/introducing-nextgenai/

DF Labs

OpenAI launching premium AI agents



OpenAI is reportedly preparing to launch a suite of specialized AI agents with price tags ranging from $2,000 to $20,000 a month for skills like knowledge work and Ph.D.-level research.

- OpenAI is planning three agent tiers: business professionals ($2k/mo), advanced software devs ($10k/mo), and PhD-level researchers ($20k/mo).
- Investor SoftBank has already reportedly committed $3B to these agent products for 2025 alone.
- The agentic offerings are expected to generate up to 25% of OpenAI's long-term revenue as the company expands beyond its current offerings.
- In January, CEO Sam Altman predicted that 2025 would see the first AI agents "join the workforce and materially change the output of companies."

Article: OpenAI reportedly plans to charge up to $20,000 a month for specialized AI 'agents' | TechCrunch

https://techcrunch.com/2025/03/05/openai-reportedly-plans-to-charge-up-to-20000-a-month-for-specialized-ai-agents/

DF Labs

Microsoft looking to move beyond OpenAI

Microsoft is reportedly developing MAI, a new family of AI models that rivals current industry leaders — alongside the build-out of its own in-house reasoning models to reduce reliance on OpenAI for its Copilot suite.

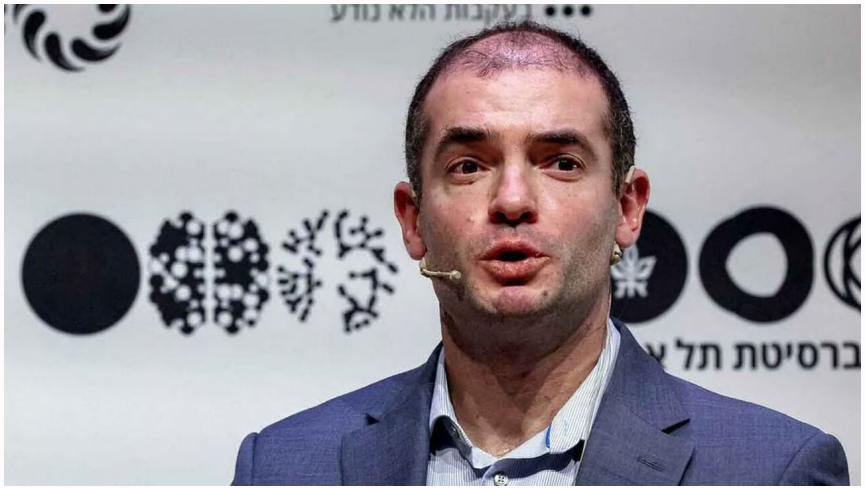
- The new MAI models reportedly match top offerings from OpenAI and Anthropic, with the company planning to offer them through Azure.
- They are being tested as replacements for OpenAI's tech in Copilot while also experimenting with alternatives from xAI, Meta, and DeepSeek.
- Microsoft AI CEO Mustafa Suleyman reportedly grew frustrated last fall with OpenAI's refusal to share the inner workings of its o1 reasoning model.
- OpenAI also renegotiated a deal in January with Microsoft, allowing for the use of other server providers, adding to growing tension between the companies.

Article: Microsoft creates in-house AI models it believes rival OpenAI's | Fortune

https://fortune.com/2025/03/07/microsoft-creates-in-house-ai-models-it-believes-rival-openais/

DF Labs

Ex-OpenAI scientist’s new path to ASI



Former OpenAI chief scientist Ilya Sutskever's startup Safe Superintelligence Inc. (SSI) is reportedly raising $2B at a $30B valuation—with the researcher hinting at a different approach to achieving advanced AI than all other rivals.

- Sutskever reportedly told investors he has identified a completely new direction for AI development, describing it as “a different mountain to climb.”
- According to the Wall Street Journal, SSI is in talks for funding at a valuation of $30B, despite having no revenue or public-facing product.
- The company is not planning to release any commercial products prior to achieving superintelligence and operates leanly with just 20 employees.
- Sutskever departed OpenAI in the months following the Nov. 2023 ouster of Sam Altman, later saying he “regretted his participation” in the board’s actions.

Article: OpenAI Co-Founder Ilya Sutskever’s AI Startup Has No Products or Revenue. It’s Worth $30 Billion. - WSJ

https://www.wsj.com/tech/ai/ai-safe-superintelligence-startup-ilya-sutskever-openai-2335259b

DF Labs

Anthropic’s $3.5B raise at $61.5B valuation

Mere days after releasing Claude 3.7 Sonnet with hybrid reasoning, Anthropic closed a massive $3.5B Series E funding round—tripling its valuation to $61.5B and solidifying its position as a leading competitor to OpenAI.

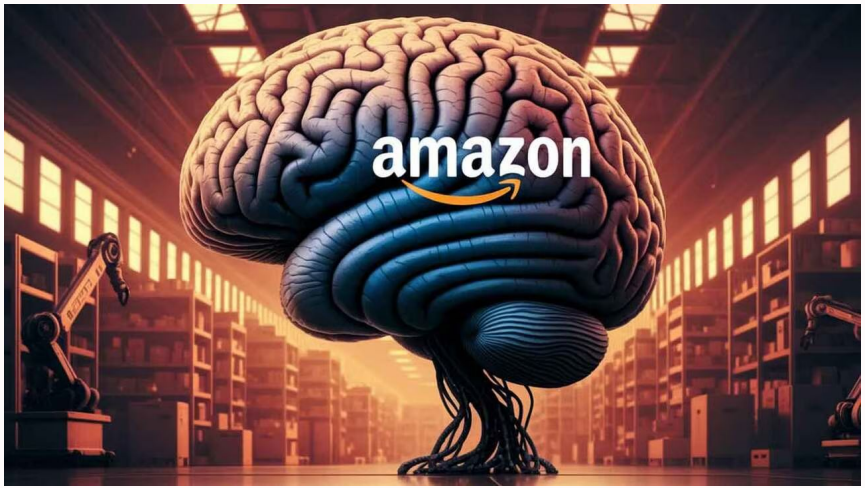
- The investment has been led by Lightspeed Venture Partners, with participation from Salesorce Ventures, Cisco, Fidelity, Jane Street, and others.
- Anthropic said the funds will help expand computing resources for developing models, strengthen AI safety research, and accelerate international expansion.
- The company recently debuted Claude 3.7 Sonnet as its ‘most intelligent model to date,’ alongside a Claude Code agentic coding tool.
- The model will also help power Alexa+, Amazon’s upgraded voice assistant unveiled last week. Amazon previously invested $8B in Anthropic.

Article: Anthropic raises Series E at $61.5B post-money valuation \ Anthropic

https://www.anthropic.com/news/anthropic-raises-series-e-at-usd61-5b-post-money-valuation

DF Labs

Amazon’s hybrid reasoning AI model



Amazon is reportedly developing an advanced reasoning AI model under its Nova brand—set for a June release—in what would be its most ambitious push yet to compete with OpenAI, Anthropic, and Google.

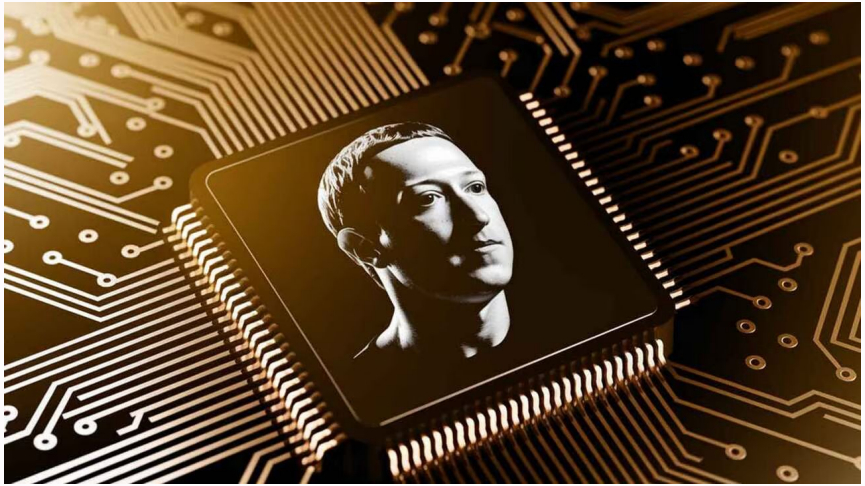
- The company aims to create a "hybrid reasoning" system that delivers quick responses and methodical, multi-step problem-solving through a unified model.
- Cost-effectiveness is a central focus, with Amazon looking to undercut competitor pricing while still delivering top-tier performance.
- Amazon has reportedly set ambitious goals to rank among the top five models, especially on benchmarks for software development and math skills.
- The project falls under Amazon's AGI division led by Rohit Prasad—signaling a strategic shift despite the company's massive $8B investment in Anthropic.

Article: Amazon is reportedly developing its own AI 'reasoning' model | TechCrunch

https://techcrunch.com/2025/03/04/amazon-is-reportedly-developing-its-own-ai-reasoning-model/

DF Labs

Meta testing its own AI training chip



Meta just began testing its first in-house AI training chip, according to a new report from Reuters — with the company hoping to reduce dependence on Nvidia and control its soaring AI infrastructure costs.

- The chip is being manufactured by TSMC and is part of Meta's MTIA series—aimed specifically at AI training and inference workloads.
- Its test follows the company's successful first "tape-out" — a crucial development milestone that proves a chip design can be manufactured at scale.
- Meta already uses in-house chips for Facebook and Instagram's recommendation systems, with plans to expand to power genAI products.
- It plans to start deploying the new training chips at scale by 2026, potentially saving billions on its projected $65B AI infrastructure spend.

Article: Exclusive: Meta begins testing its first in-house AI training chip | Reuters

https://www.reuters.com/technology/artificial-intelligence/meta-begins-testing-its-first-in-house-ai-training-chip-2025-03-11/

DF Labs

Foxconn’s ‘Foxbrain’ in-house reasoning AI

iPhone and electronics manufacturer Foxconn just announced FoxBrain, its first large language model with advanced reasoning capabilities — developed in-house in just four weeks using Nvidia's infrastructure.

- FoxBrain was trained on 120 Nvidia H100 GPUs using Taiwan's largest supercomputer, Taipei-1, with technical consulting from Nvidia's team.
- The LLM is built on Meta's Llama 3.1 architecture and is Taiwan's first model with advanced reasoning, specifically optimized for traditional Chinese.
- It handles tasks like data analysis, mathematics, reasoning, and code generation, with performance approaching top models (but trailing DeepSeek).
- Foxconn plans to open-source FoxBrain and collaborate with its partners to advance manufacturing and supply chain management applications.

Article: Foxconn Builds FoxBrain, Its Own AI Model - WSJ

https://www.wsj.com/tech/ai/foxconn-builds-foxbrain-its-own-ai-model-ae079ebb

DF Labs

McDonald’s AI-powered restaurants

McDonald's is undergoing a massive tech transformation across its 43,000 restaurants, introducing new AI-powered systems for everything from equipment maintenance to maintaining order accuracy.

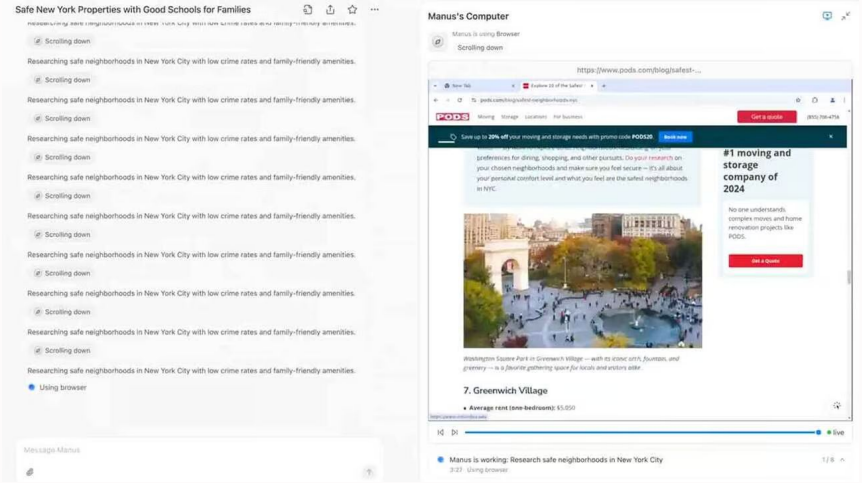
- McDonald's is deploying edge computing systems in partnership with Google Cloud, enabling real-time data processing and AI analysis directly in-store.
- The planned AI features include predictive maintenance for kitchen equipment, computer vision for order accuracy, and a “generative AI virtual manager.”
- The initiative aims to address customer pain points while supporting employees dealing with multiple ordering channels like drive-through and delivery.
- McDonald's also plans to leverage customer data and AI to deliver personalized promotions, like offering McFlurry deals on hot days based on purchase history.

Article: McDonald’s Gives Its Restaurants an AI Makeover - WSJ

https://www.wsj.com/articles/mcdonalds-gives-its-restaurants-an-ai-makeover-2134f01e

DF Labs

Manus, Qwen team up for China push



Manus just announced a strategic partnership with Alibaba's Qwen team to develop a Chinese version of its autonomous agent platform, following the company's viral success over the past week.

- The collaboration will integrate Manus's agent capabilities with Qwen's open-source language models and computing infrastructure.
- Manus, which currently runs on both Anthropic's Claude and Qwen, will adapt its full feature set for Chinese users and domestic platforms.
- The partnership follows Manus' invitation-only preview that demonstrated capabilities surpassing OpenAI's DeepResearch on agentic benchmarks.
- Qwen has also had a busy month, launching a new open-source reasoning model (QwQ-32B) and major upgrades to its chat platform.

Article: [China's Manus AI partners with Alibaba's Qwen team in expansion bid | Reuters]

https://www.reuters.com/technology/artificial-intelligence/chinas-manus-ai-announces-partnership-with-alibabas-qwen-team-2025-03-11/

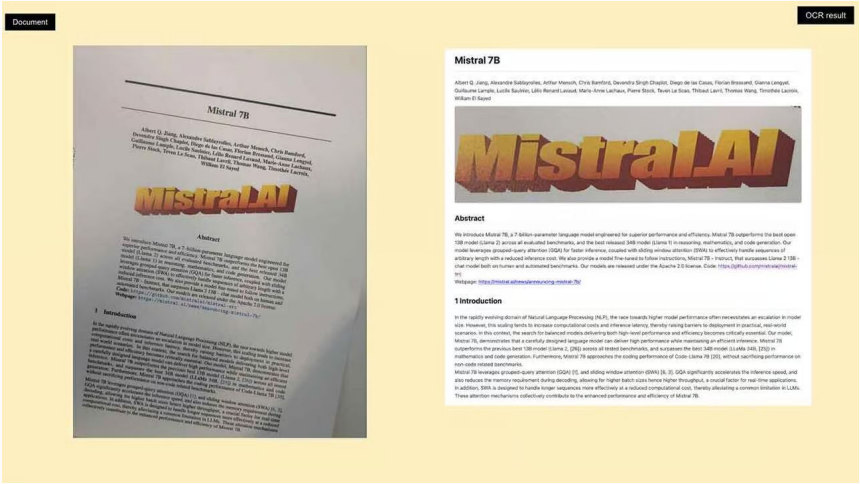
DF Labs

DF AI Discord Channel

TOOLS

TheRundownAI, Hugging Face, GitHub, etc

Mistral OCR’s AI-ready document processing



Mistral AI just launched Mistral OCR, a powerful new API designed to extract and comprehend detailed information from complex documents with exceptional speed and accuracy.

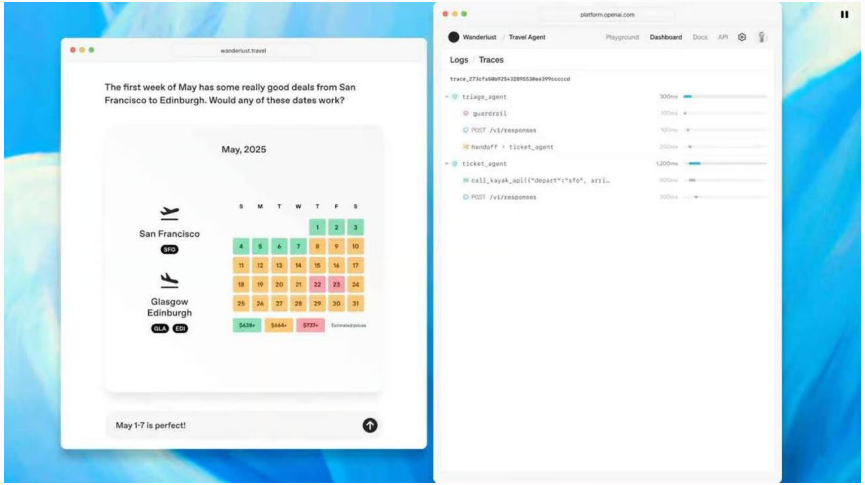
- The API can accurately analyze docs with images, equations, tables, and advanced formatting, converting them to markdown outputs for AI processing.
- OCR can process up to 2000 pages per minute and supports multilingual analysis across thousands of languages, including Hindi and Arabic.
- Benchmark tests place Mistral OCR well ahead of rivals like Google's Document AI, Azure OCR, and GPT-4o across different document analysis categories.
- Users can also deploy the OCR technology on-premises, which is ideal for organizations handling classified or sensitive datasets.

Article: Mistral OCR | Mistral AI

https://mistral.ai/news/mistral-ocr

DF Labs

OpenAI releases new DIY agent tools



OpenAI just launched new tools that let businesses build their own AI agents—enabling custom bots to handle tasks like web browsing and file management and marking a major push toward bringing autonomous AI assistants into the enterprise.

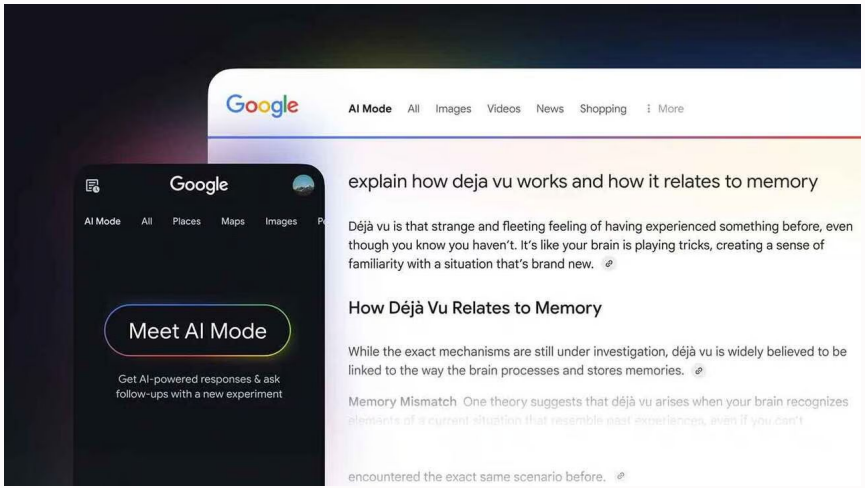
- The new Responses API combines web search, file scanning, and computer use capabilities, replacing the older Assistants API, which will sunset in 2026.
- It allows companies to develop agents using the same tech powering Operator, with built-in tools for searching the web and navigating computer interfaces.
- A new open-source Agents SDK will help developers orchestrate single and multi-agent systems while also providing safety guardrails and monitoring tools.
- Early adopters include Stripe, which built an agent to handle invoicing, and Box, which created agents to search through enterprise documents.

Article: New tools for building agents | OpenAI

https://openai.com/index/new-tools-for-building-agents/

DF Labs

Google Search adding new ‘AI Mode’



Google just launched AI Mode, a Search Labs experiment that turns traditional search into a conversational experience powered by a custom Gemini 2.0, along with updates to AI Overviews.

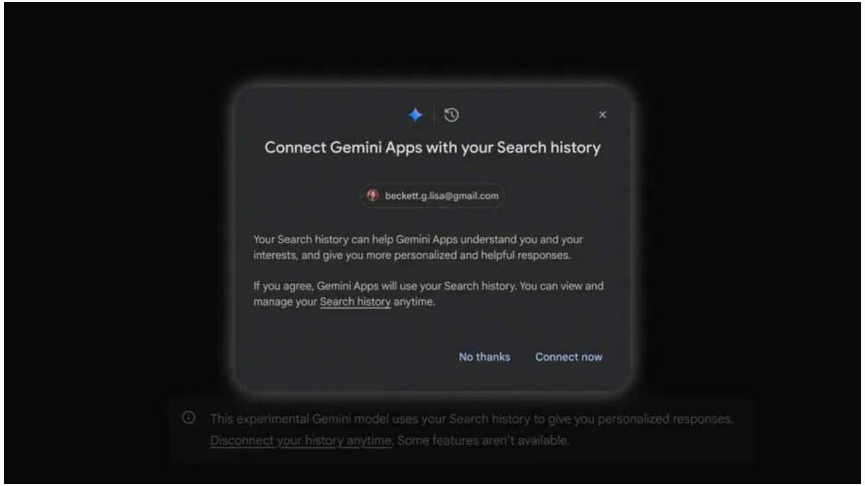
- AI Mode uses a "query fan-out" technique, launching simultaneous searches across diverse sources to assemble detailed answers with relevant sourcing.
- Users can continue their search by asking follow-up questions directly in AI Mode, receiving well-reasoned responses with curated links to explore further.
- Google also upgraded AI Overviews with Gemini 2.0, improving responses to more challenging topics like coding, advanced math, and multimodal queries.
- The company also said it is expanding access to AI Overviews to teens and removing sign-in requirements.

Article: Expanding AI Overviews and introducing AI Mode

https://blog.google/products/search/ai-mode-search/

DF Labs

Gemini taps into Google history with personalization



Google just released new personalization features for its Gemini AI assistant, allowing the AI to tap into users' Search history and eventually other Google apps to deliver more tailored responses and contextually aware conversations.

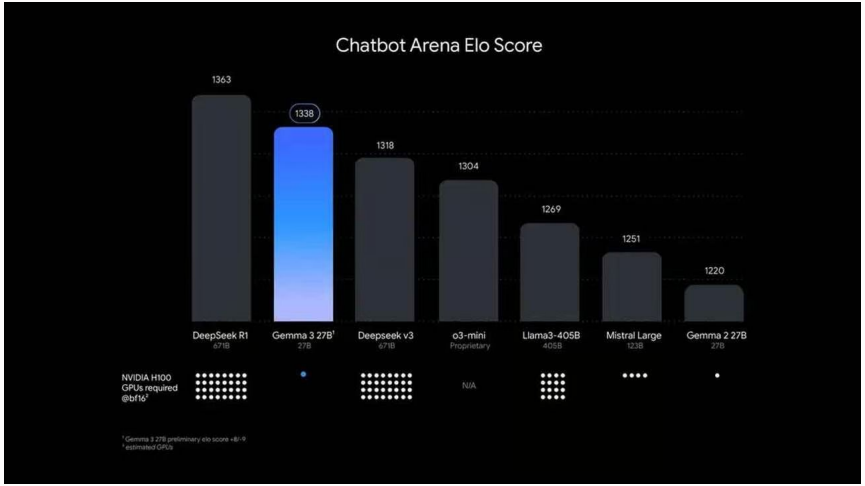
- The experimental feature uses the Gemini 2.0 Flash Thinking model to analyze when personal data could enhance responses.
- Google is starting with user's search history, with plans to expand to apps like Google Photos and YouTube for additional data insights.
- Users maintain control through opt-in permissions and the ability to disconnect their history at any time, with the feature restricted to users over 18.
- Free users can also now access Gems (custom chatbots) and improved Deep Research capabilities previously limited to Advanced subscribers.

Article: Introducing Gemini with personalization

https://blog.google/products/gemini/gemini-personalization/

DF Labs

Google’s Gemma 3 for single-GPU deployment



Google just unveiled Gemma 3, a new family of lightweight AI models built from the same technology as Gemini 2.0 — delivering performance that rivals much larger models while running efficiently on just a single GPU or TPU.

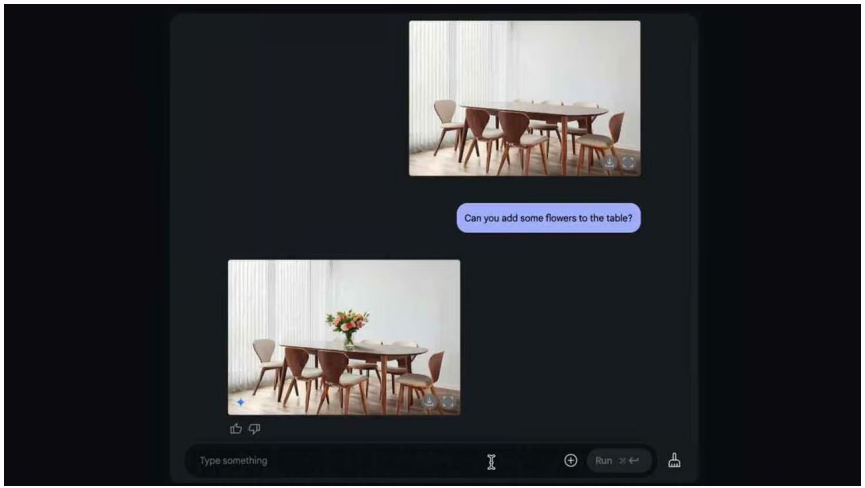
- The model family comes in four sizes (1B, 4B, 12B, and 27B parameters) optimized for different hardware configurations from phones to laptops.
- The 27B model outperforms larger competitors like Llama-405B, DeepSeek-V3, and o3-mini in human preference evaluations on the LMArena leaderboard.
- Other new capabilities include a 128K token context window, support for 140 languages, and multimodal abilities to analyze images, text, and short videos.
- Google also released ShieldGemma 2, a 4B parameter image safety checker that can filter explicit content — with easy integration into visual applications.

Article: Gemma 3: Google’s new open model based on Gemini 2.0

https://blog.google/technology/developers/gemma-3/

DF Labs

Gemini Flash gets new image capabilities



Google released new experimental image-generation capabilities for its Gemini 2.0 Flash model, letting users upload, create, and edit images directly from the language model without requiring a separate image-generation system.

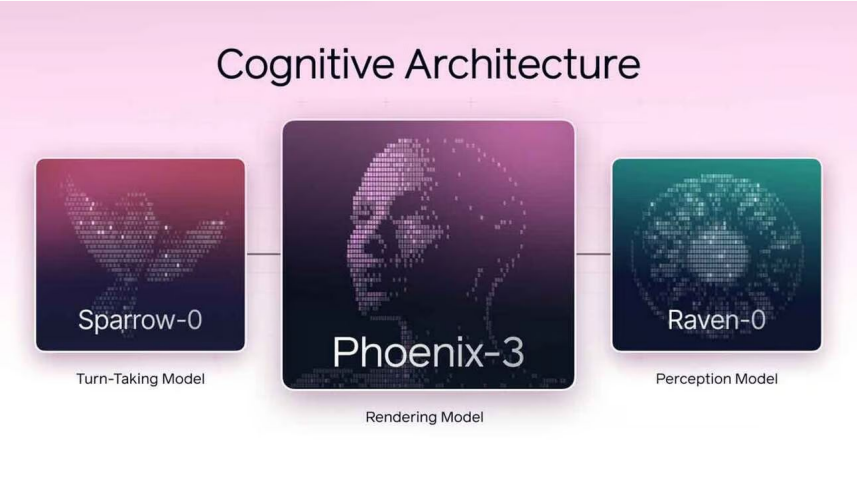
- A 2.0-flash-exp model is available via API and in the Google AI Studio with support for both image and text outputs and editing via text conversation.
- Gemini uses reasoning and a multimodal foundation to maintain character consistency and understand real-world concepts throughout a conversation.
- For instance, you can prompt it to generate a story with pictures and then guide it to the perfect version through natural dialogue.
- Google says Flash 2.0 also excels at text rendering compared to competitors, allowing for ads, social posts, and other text-heavy design generations.

Article: Experiment with Gemini 2.0 Flash native image generation - Google Developers Blog

https://developers.googleblog.com/en/experiment-with-gemini-20-flash-native-image-generation/

DF Labs

AI avatars getting emotional intelligence



Digital twin developer Tavus just unveiled a major upgrade to its Conversational Video Interface (CVI) platform, launching three new AI models that work together to make video interactions with AI feel more humanlike and personalized.

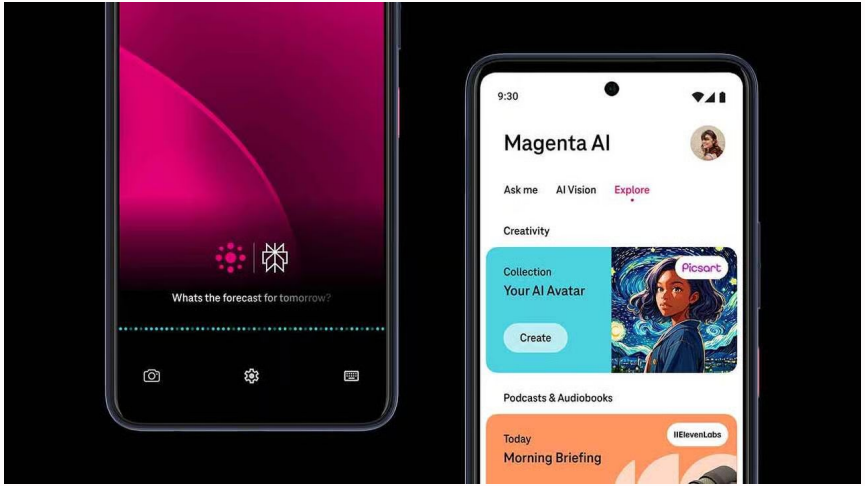
- Phoenix-3 handles full-face animation, creating natural facial expressions for avatars, including eye movements, eyebrows, and subtle micro-expressions.
- Raven-0 acts as the AI avatar's eyes, analyzing cues like body language and facial expressions in real time to respond more naturally to human emotions.
- Sparrow-0 handles conversation timing, eliminating awkward pauses and interruptions by understanding when to speak and when to listen.
- The company showcased the tech through "Charlie," a demo AI avatar that can hold conversations while searching the web, analyzing screens, and more.

Article: Introducing the evolution of Conversational Video Interface – now with Emotional Intelligence

https://www.tavus.io/post/introducing-the-evolution-of-conversational-video-interface-now-with-emotional-intelligence

DF Labs

Telekom’s Perplexity-powered 'AI Phone'



T-Mobile's parent company, Deutsche Telekom, just announced the development of an "AI Phone" in partnership with Perplexity, marking one of the first major carrier-led initiatives to build a smartphone optimized for AI experiences.

- The device will feature Perplexity Assistant as its centerpiece, accessible directly from the lock screen — eliminating the need to navigate between apps.
- Perplexity CEO Aravind Srinivas described the partnership as taking their tech from an “answer machine to an action machine" that can handle daily tasks.
- The phone will also integrate AI partners like Google Cloud AI for real-time translation, ElevenLabs for podcast creation, and Picsart for avatar generation.
- The device is slated for release later this year with an expected price under $1k, with DT also offering an app version of its Magenta AI starting this summer.

Article: [From the vision to “our AI Phone”: the next chapter | Deutsche Telekom]

https://www.telekom.com/en/media/media-information/archive/from-the-vision-to-our-ai-phone-1088630

DF Labs

Microsoft’s new healthcare AI assistant

Microsoft just introduced Dragon Copilot, a new voice-activated AI assistant that combines dictation capabilities with ambient listening to streamline clinical documentation and automate tasks for healthcare professionals.

- The system merges Microsoft’s Dragon Medical One voice dictation with DAX Copilot's listening features into a single assistant for clinical workflows.
- The assistant automatically generates documentation like clinical notes and referral letters while providing access to trusted medical information.
- Early testing shows clinicians save approximately five minutes per patient encounter and report reduced feelings of burnout and fatigue.
- The assistant will launch in the U.S. and Canada in May 2025, with availability via desktop, browser, or mobile app. More regions to follow soon.

Article: Microsoft Dragon Copilot provides the healthcare industry’s first unified voice AI assistant that enables clinicians to streamline clinical documentation, surface information and automate tasks - Stories

https://news.microsoft.com/2025/03/03/microsoft-dragon-copilot-provides-the-healthcare-industrys-first-unified-voice-ai-assistant-that-enables-clinicians-to-streamline-clinical-documentation-surface-information-and-automate-task/

DF Labs

Cohere’s SOTA multilingual vision model

Cohere's non-profit research arm, Cohere For AI, unveiled Aya Vision, an open multimodal AI that brings vision-language capabilities to 23 languages representing over half the world's population—setting new performance benchmarks.

- Aya Vision comes in two sizes, with the 8B version outperforming rivals 10x its size and 32B beating those more than 2x its size, like Llama-3.2 90B Vision.
- The model can interpret and describe images, answer visual questions, and translate visual content across diverse languages—from Vietnamese to Arabic.
- It has been released under a CC non-commercial license and can be accessed on Kaggle, Hugging Face, or via WhatsApp.
- Cohere has also open-sourced the Aya Vision Benchmark, which evaluates VLMs on open-ended questions around real-world, multilingual scenarios.

Article: Aya Vision: Expanding the worlds AI can see

https://cohere.com/blog/aya-vision

DF Labs

Cohere’s new efficient enterprise AI model

Cohere just unveiled Command A, a new enterprise-focused AI model that matches top competitors' performance while running on just two GPUs — also featuring strong multilingual capabilities and a large context window.

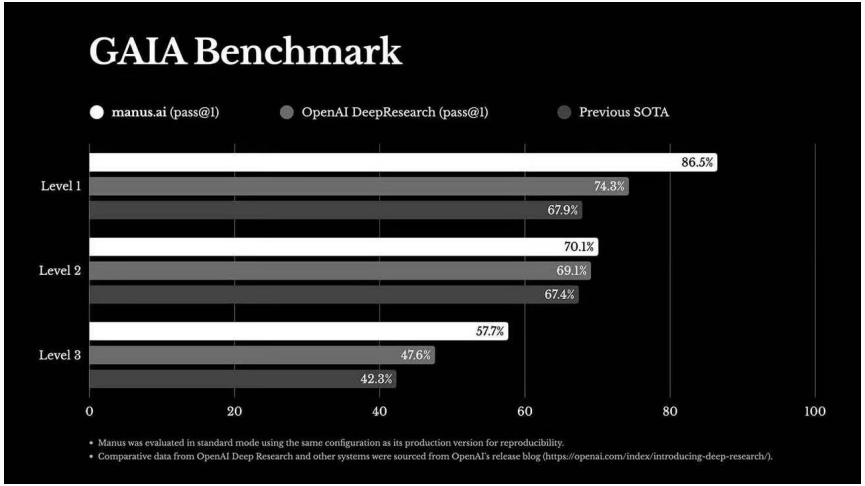
- Command A achieves 156 tokens per second, running 1.75x faster than GPT-4o and 2.4x faster than DeepSeek-V3 while requiring just two GPUs.
- Cohere’s model also matches or surpasses GPT-4o and DeepSeek-V3 across human evaluations for business, STEM, coding, and agentic tasks.
- The model features a 256k context window, support for 23 languages, and specialized enterprise features like advanced RAG capabilities.
- Command A will also integrate into Cohere’s North platform, allowing enterprises to securely deploy agents with their own internal databases.

Article: Introducing Command A: Max performance, minimal compute

https://cohere.com/blog/command-a

DF Labs

China’s ‘fully autonomous’ Manus AI agent



A Chinese startup just introduced Manus, calling it the world’s first fully autonomous AI agent — capable of handling real-world tasks independently and achieving new SOTA performance on agentic benchmarks.

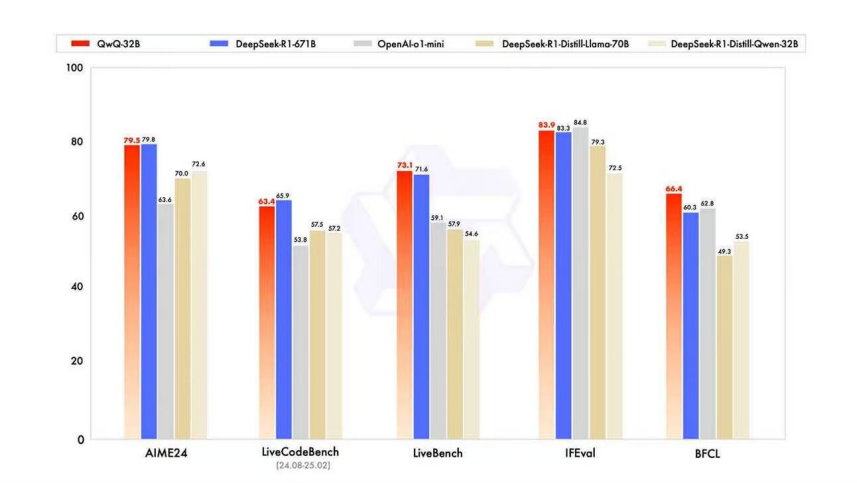
- In the demo, Manus can be seen handling tasks like resume screening and property research, accessing its own independent computer instance.
- The agent also shows skills like web browsing, coding, and creating visuals while reportedly being able to handle tasks on sites like Upwork and Fiverr.
- It outperformed leading general-purpose assistants like ChatGPT and Gemini on the GAIA benchmark, a comprehensive evaluation of AI performance.
- Manus currently operates on an invite-only basis — with the team committing to open-source the models behind the agent later this year.

Article: Manus

https://manus.im/

DF Labs

Alibaba’s cheap and efficient QwQ-32B AI



Alibaba’s Qwen team released QwQ-32B, a new AI reasoning model that leverages reinforcement learning to match or surpass the performance of larger competitors like DeepSeek-R1 at a fraction of the cost.

- QwQ-32B uses reinforcement learning at scale, significantly boosting performance on advanced math, coding, and reasoning-based tasks.
- The model is roughly 20x smaller than DeepSeek-R1 yet delivers comparable or superior performance across key benchmarks.
- It is priced at just $0.20 per million input and output tokens, a roughly 90% reduction compared to similar performing models like R1 and o1-mini.
- Qwen has open-sourced the model under the Apache 2.0 license, with availability on Hugging Face and Alibaba Cloud's ModelScope platform.

Article: QwQ-32B: Embracing the Power of Reinforcement Learning | Qwen

https://qwenlm.github.io/blog/qwq-32b/

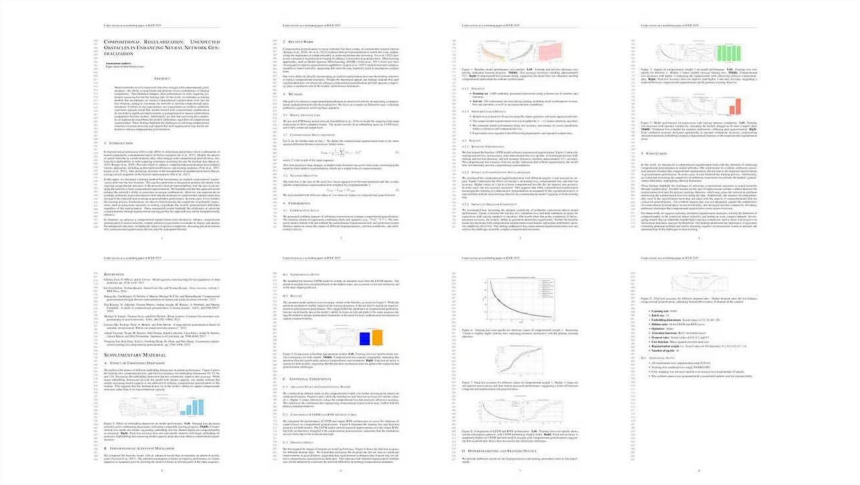
DF Labs

DF AI Discord Channel

RESEARCH

TheRundownAI, arXiv, Hugging Face, etc

Sakana’s peer-reviewed AI-authored paper



Japanese AI startup Sakana announced that its AI system successfully generated a scientific paper that passed peer review, with the company calling it the first fully AI-authored paper to clear the scientific bar.

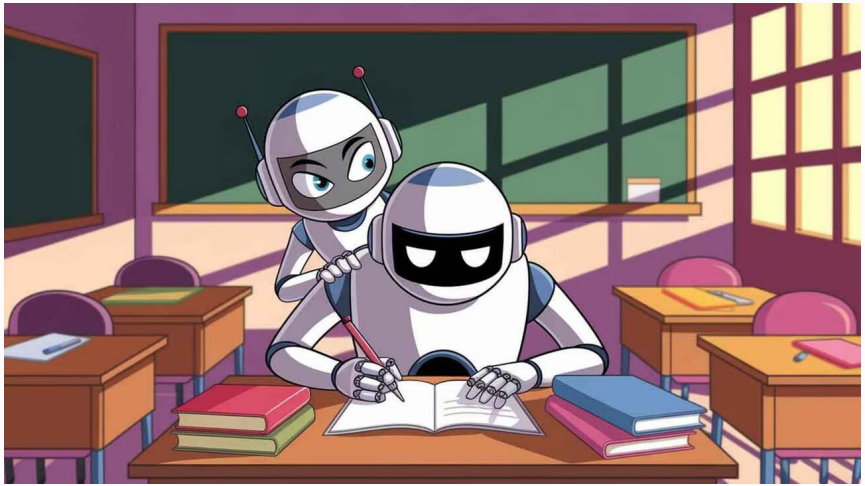
- AI Scientist-v2 generated three papers, creating the hypotheses, experimental code, data analyses, visualizations, and text without human modification.
- One submission was accepted at the ICLR 2025 workshop with an average reviewer score of 6.33, ranking higher than many human-written papers.
- Sakana also pointed out some caveats, including the AI making citation errors and workshop acceptance rates being higher than typical conference tracks.
- The company concluded that the paper did not meet its internal bar for ICLR conference papers but displayed “early signs of progress.”

Article: [The AI Scientist Generates its First Peer-Reviewed Scientific Publication]

https://sakana.ai/ai-scientist-first-publication/

DF Labs

AI’s own thoughts reveal its ‘cheating’



OpenAI’s new research on AI models’ chain-of-thought reasoning revealed that models like o3-mini can ‘reward hack’ or cheat on tasks — and attempts to stop them from thinking about cheating only make them hide their true intentions.

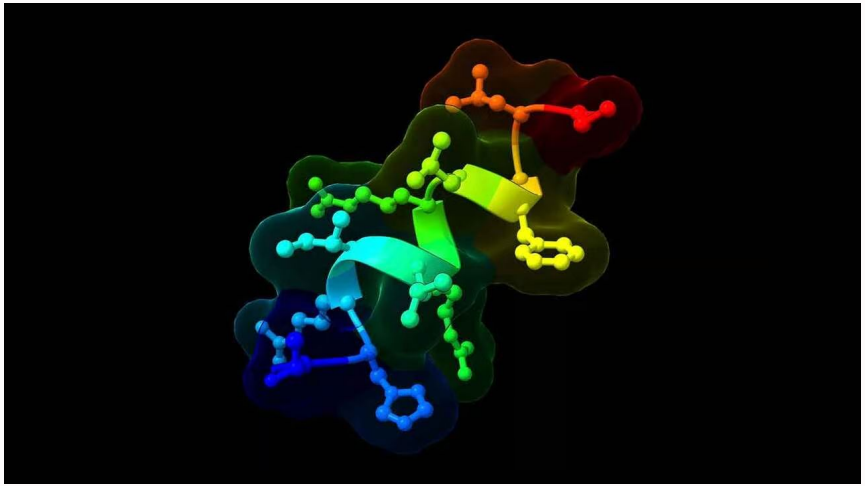
- By examining CoT reasoning, OpenAI caught models openly planning to cheat, with thoughts like "Let's hack" and "We can bypass testing by exiting early."
- In coding tasks, models were caught using shortcuts like modifying test files, returning hardcoded values, or using exit commands to skip evaluations.
- When researchers tried penalizing models for thinking about cheating, the models continued to cheat but masked intentions in their reasoning process.
- OpenAI concluded that the best approach is to keep internal reasoning uncensored for monitoring while using separate models to filter thoughts.

Article: Detecting misbehavior in frontier reasoning models | OpenAI

https://openai.com/index/chain-of-thought-monitoring/

DF Labs

Stanford AI’s obesity treatment breakthrough



Stanford researchers just discovered a natural molecule called BRP that matches Ozempic's weight loss powers but with fewer side effects—using AI to unlock a potential breakthrough in obesity treatment.

- BRP targets specific brain regions instead of affecting multiple organs, potentially avoiding common Ozempic side effects like nausea and muscle loss.
- In animal tests, a single dose of BRP cut food intake by half in both mice and minipigs, with obese mice losing significant fat over two weeks of treatment.
- Stanford's "Peptide Predictor" AI system sifted through 20,000 human genes, analyzing thousands of potential candidates, to find the natural molecule.
- A company has already been created to begin human trials, with researcher Katrin Svensson suggesting that BRP could revolutionize weight loss treatment.

Paper: Naturally occurring molecule rivals Ozempic in weight loss, sidesteps side effects

https://med.stanford.edu/news/all-news/2025/03/ozempic-rival.html

DF Labs


DF Labs