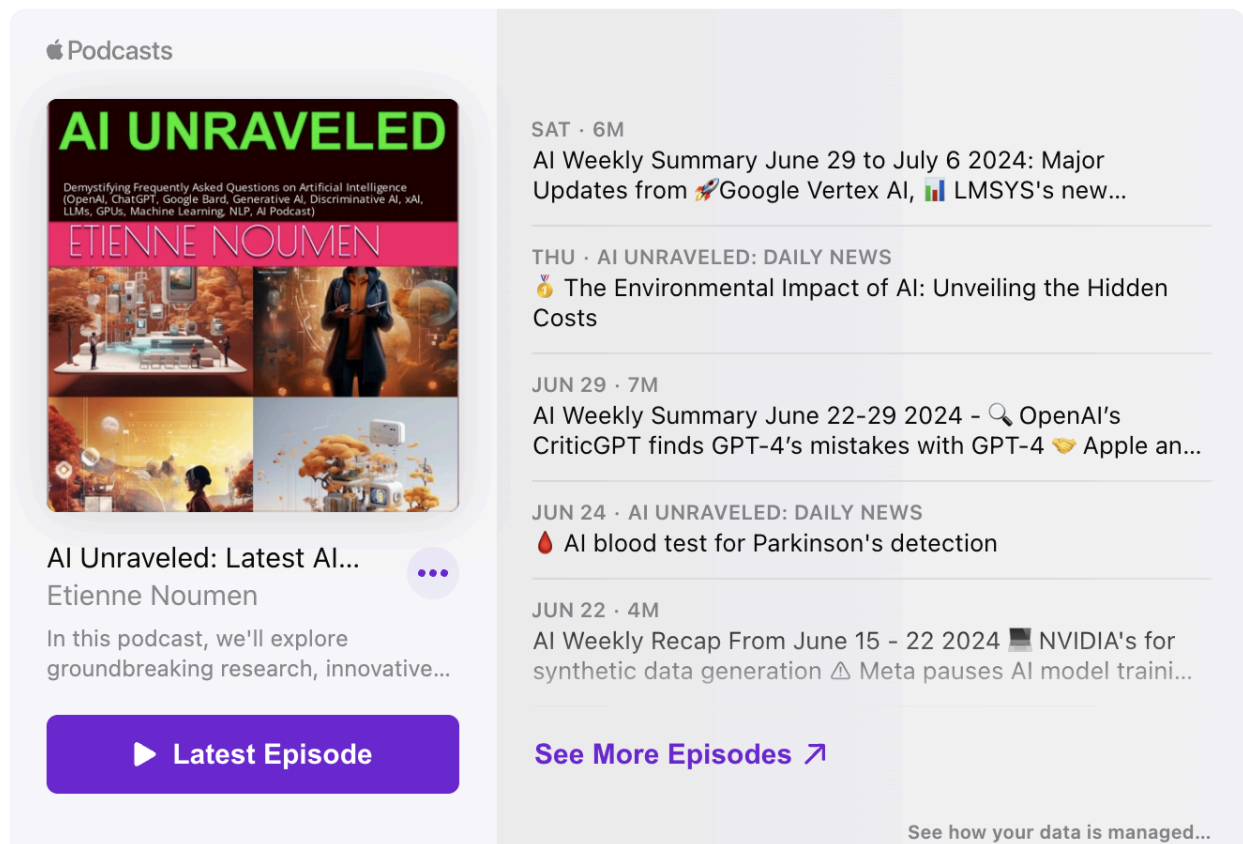


Daily AI Innovation in August 2024

[Become a Subscriber](#) - [Donate to Support us](#)

[Listen to our podcast here](#)



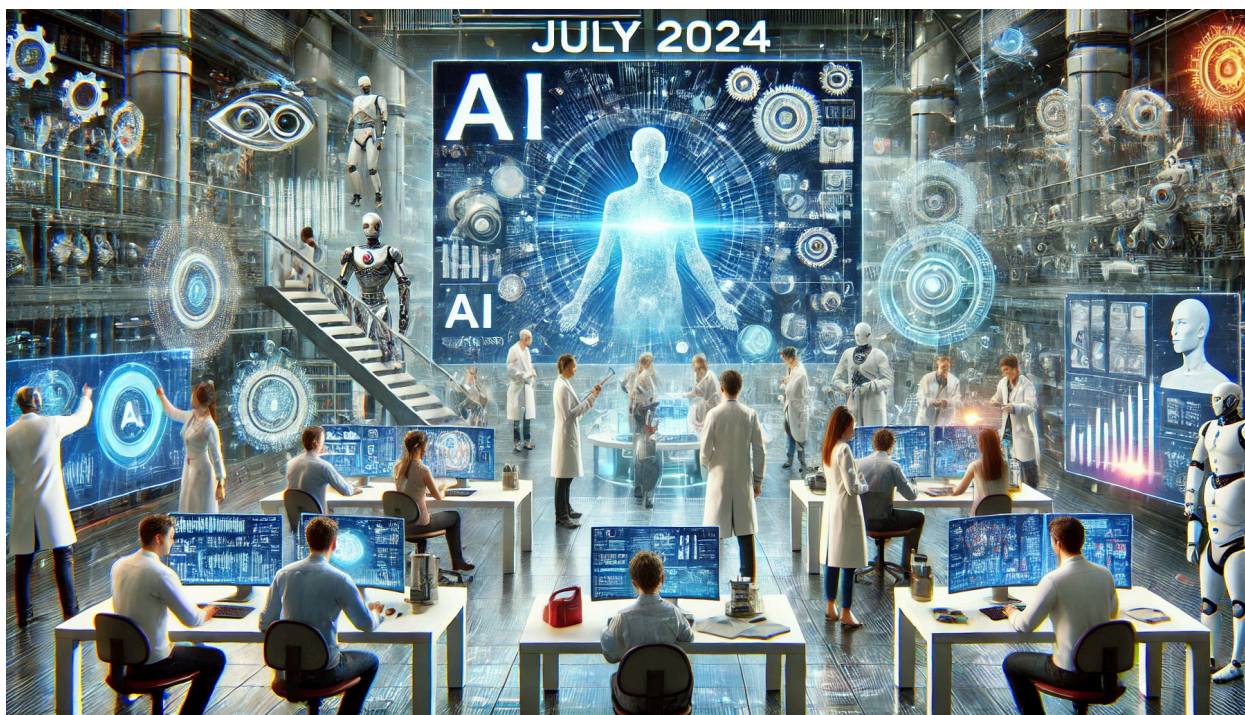
The screenshot shows the Apple Podcasts interface for the 'AI UNRAVELED' podcast. The main card features a cover image with the title 'AI UNRAVELED' in green and 'ETIENNE NOUMEN' in pink. Below the cover, it says 'AI Unraveled: Latest AI...' and 'Etienne Noumen'. A purple button labeled 'Latest Episode' is at the bottom. To the right, a list of recent episodes is shown, including 'AI Weekly Summary June 29 to July 6 2024: Major Updates from Google Vertex AI, LMSYS's new...', 'The Environmental Impact of AI: Unveiling the Hidden Costs', 'AI Weekly Summary June 22-29 2024 - OpenAI's CriticGPT finds GPT-4's mistakes with GPT-4 Apple an...', 'AI blood test for Parkinson's detection', and 'AI Weekly Recap From June 15 - 22 2024 NVIDIA's for synthetic data generation Meta pauses AI model traini...'. A 'See More Episodes' link is at the bottom of the list. A small link 'See how your data is managed...' is at the bottom right of the screenshot.

Welcome to our blog series "AI Innovations in July 2024"! As we continue to ride the wave of extraordinary developments from June, the momentum in artificial intelligence shows no signs of slowing down. Last month, we witnessed groundbreaking achievements such as the unveiling of the first quantum AI chip, the successful deployment of autonomous medical drones in remote areas, and significant advancements in natural language understanding that have set new benchmarks for AI-human interaction.

July promises to be just as exhilarating, with researchers, engineers, and visionaries pushing the boundaries of what's possible even further. In this evolving article, updated daily throughout the month, we'll dive deep into the latest AI breakthroughs, advancements, and milestones shaping the future.

From revolutionary AI-powered technologies and cutting-edge research to the societal and ethical implications of these innovations, we provide you with a comprehensive and insightful look at the rapidly evolving world of artificial intelligence. Whether you're an AI enthusiast, a tech-savvy professional, or simply someone curious about the future, this blog will keep you informed, inspired, and engaged.

Join us on this journey of discovery as we explore the frontiers of AI, uncovering the innovations that are transforming industries, enhancing our lives, and shaping our future. Stay tuned for daily updates, and get ready to be amazed by the incredible advancements happening in the world of AI!



Enjoying these daily updates, subscribe and listen to our podcast FREE at [AI Unraveled: Latest AI News & Trends, GPT, Gemini, Generative AI, LLMs, Prompt, AI Bedtime Stories on Apple Podcasts](#)

Increase your salary by Acing the AWS Certified Data Engineer Exam (DEA-C01) with this eBook below

Ace the AWS Certified Data Engineer - Associate Exam



Quizzes, Practice Exams, Tutorials,
Testimonials, and more

DJAMGATECH

AWS

DEA

[Ace the AWS Certified Data Enginner Associate Exam Book at Apple here](#)

[Ace the AWS Certified Data Enginner Associate Exam Book at Google here](#)

[Ace the AWS Certified Data Enginner Associate Exam Book at Etsy here](#)

[You can also support us by buying the
Wonderland Bedtime Adventure Book for your
kids.](#)

[Wonderland Bedtime Adventure print Book at Barnes & Nobles](#)

[Wonderland Bedtime Adventures eBook - Google](#)

[Wonderland Bedtime Adventures eBook - Etsy](#)

[Wonderland Bedtime Adventure eBook at Apple](#)

[A Daily chronicle of AI Innovations July 24th 2024:](#)

[A Daily chronicle of AI Innovations July 23rd 2024:](#)

[A Daily chronicle of AI Innovations July 18th 2024:](#)

[A Daily chronicle of AI Innovations July 17th 2024:](#)

[A Daily chronicle of AI Innovations July 16th 2024:](#)

[A Daily chronicle of AI Innovations July 15th 2024:](#)

[A Daily chronicle of AI Innovations July 12th 2024:](#)

[A Daily chronicle of AI Innovations July 11th 2024:](#)

[A Daily chronicle of AI Innovations July 10th 2024:](#)

[A Daily chronicle of AI Innovations July 09th 2024:](#)

[A Daily chronicle of AI Innovations July 08th 2024:](#)

[A Daily chronicle of AI Innovations July 05th 2024:](#)

[A Daily chronicle of AI Innovations July 04th 2024:](#)

[A Daily chronicle of AI Innovations July 03rd 2024:](#)

[A Daily chronicle of AI Innovations July 02nd 2024:](#)

[A Daily chronicle of AI Innovations July 01st 2024:](#)


A Daily chronicle of AI Innovations July 24th 2024:



Google search is thriving despite AI shift

 **Google is pouring billions into self-driving taxis as Tesla prepares to reveal its rival**

 **Senators demand answers on OpenAI's practices**

 **Meta's Llama 3.1 takes on GPT-4o**

 **Adobe's new AI features for Photoshop**

Enjoying these updates, Get to the next level in your career by Acing the [AWS Certify Data Engineer Associate Exam \(DEA-C01\)](#) with this [Ebook](#):

Get it now at Google at <https://play.google.com/store/books/details?id=lzgPEQAAQBAJ> or

Apple at

<https://books.apple.com/ca/book/ace-the-aws-certified-data-engineer-associate/id6504572187>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

 **Google search is thriving despite AI shift**

Despite concerns from online publishers, Google's introduction of AI features generating conversational responses to search queries has attracted advertisers and propelled Alphabet's success. Alphabet's revenue for the April-June quarter rose by 14% from last year to \$84.74 billion, surpassing analyst expectations and boosting stock prices by 2% in extended trading. Google's cloud-computing division, its fastest-growing segment, generated \$10.3 billion in revenue in the past quarter, marking its first time surpassing the \$10 billion threshold in a single quarter. Source: <https://www.fastcompany.com/91161798/google-search-is-still-thriving-despite-a-shift-to-ai-earnings>

 **Google is pouring billions into self-driving taxis as Tesla prepares to reveal its rival**

Alphabet is investing \$5 billion in Waymo's self-driving taxi service, highlighting its commitment to autonomous vehicles. Waymo has achieved over 50,000 paid autonomous rides weekly in cities like San Francisco and Phoenix, showcasing its progress and customer acceptance. Tesla is also

preparing to enter the self-driving taxi market, with an important event unveiling its rival service rescheduled from August to October. Source:
<https://www.businessinsider.com/alphabet-is-pouring-billions-into-waymos-self-driving-vehicles-2024-7>

Senators demand answers on OpenAI's practices

Five U.S. Senators have just sent a letter to OpenAI CEO Sam Altman, demanding details about the company's efforts to ensure AI safety following reports of rushed safety testing for GPT-4 Omni.

Senators question OpenAI's safety protocols, citing reports that the company rushed safety testing of GPT-4 Omni to meet a May release date.

The letter requests OpenAI to make its next foundation model available to U.S. Government agencies for deployment testing, review, analysis, and assessment.

Lawmakers ask if OpenAI will commit 20% of computing resources to AI safety research, a promise made in July 2023 when announcing the now disbanded "Superalignment team".

With allegations of rushed safety testing, potential retaliation against whistleblowers, and the disbanding of the "Superalignment team," OpenAI is under intense scrutiny. This letter also marks a critical moment for the entire AI industry – with the potential to lead to stricter government oversight and new industry standards.

Source:
<https://cointelegraph.com/news/us-lawmakers-letter-open-ai-requesting-government-access>

Meta's Llama 3.1 takes on GPT-4o

In case you missed our exclusive deep dive with Mark Zuckerberg yesterday, Meta released Llama 3.1, including its long awaited 405B parameter model – the first open sourced frontier model that beats top closed models like GPT-4o across several benchmarks.

The 405B parameter version of Llama 3.1 matches or exceeds top closed models on several benchmarks.

Meta is offering open and free weights and code, with a license enabling fine-tuning, distillation into other models, and deployment anywhere.

Llama 3.1 features a 128k context length, multi-lingual abilities, strong code generation performance, and complex reasoning capabilities.

For exclusive insights on Llama 3.1, open source, AI agents, and more, read our full deep dive with Mark Zuckerberg [here](#), or watch the full interview [here](#).

Meta's release of Llama 3.1 405b is a significant moment in AI history because it's the first time an open-source AI model matches or outperforms top closed AI models like OpenAI's GPT-4o. By offering a private, customizable alternative to closed AI systems, Meta is enabling anyone to create their own tailored AI.

Source: <https://www.therundown.ai/p/meta-releases-llama-405b>

Adobe's new AI features for Photoshop

Adobe just unveiled major AI-powered updates to Illustrator and Photoshop, leveraging its Firefly AI model to accelerate creative workflows and introduce new generative design capabilities.

Illustrator introduces Generative Shape Fill using Firefly Vector AI to add detailed vectors to shapes and create scalable patterns via text prompts.

Text to Pattern in Illustrator creates scalable, customized vector patterns for designs like wallpapers.

Photoshop's new AI-powered Selection Brush Tool and Generate Image function are now generally available.

Photoshop also gets an enhanced version of its popular Generative Fill for improved sharpness in large images.

These updates could dramatically increase designers' productivity by automating tedious, time-consuming tasks. We've always preached that the best AI products are those embedded into everyday workflows — and Adobe is doing just that by putting powerful tech directly into designers' everyday tools.

Source:

<https://news.adobe.com/news/news-details/2024/Adobe-Unveils-Powerful-New-Innovations-in-Illustrator-and-Photoshop-Unlocking-New-Design-Possibilities-for-Creative-Pros/default.aspx>

Mark Zuckerberg explains why open source AI is good for developers.

Source:

<https://www.neowin.net/news/mark-zuckerberg-explains-why-open-source-ai-is-good-for-developers/>

Google has big new ideas about the Play Store.

The company is rolling out several new features including Collections, AI-powered app comparisons, and more

Source:

<https://www.theverge.com/2024/7/24/24205052/google-play-collections-ai-features-rewards-pixel>

OpenAI offers free GPT-4o Mini fine-tuning to counter Meta's Llama 3.1 release.

Source:

<https://venturebeat.com/ai/ai-arms-race-escalates-openai-offers-free-gpt-4o-mini-fine-tuning-to-counter-metas-llama-3-1-release/>

A Daily chronicle of AI Innovations July 23rd 2024:

 **Meta releases its most powerful AI model yet**

 **Alexa is losing Amazon billions of dollars**

 **The "world's most powerful" supercomputer**

 **Google's AI-powered weather model**

 **MIT's AI identifies breast cancer risk**

 **Musk unveils the world's most powerful AI training cluster**

 **Robotics won't have a ChatGPT-like explosion: New Research**

NeuralGCM predicts weather faster than SOTA climate models

Perplexity launches Voice Mode for its AI assistant on iOS

VeriSIM Life's AI platform can accelerate drug discovery

Anthropic is working on a new screenshot tool for Claude

Enjoying these FREE daily updates without SPAM or clutter? then, Listen to this episode at our podcast and Support us by subscribing at

<https://podcasts.apple.com/ca/podcast/ai-weekly-summary-july-14th-to-july-21st-2024-openai/id1684415169?i=1000662838430>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

Robotics won't have a ChatGPT-like explosion: New Research

Coatue Management has released a report on AI humanoids and robotics's current and future state. It says robotics will unlikely have a ChatGPT-like moment where a single technology radically transforms our work. While robots have been used for physical labor for over 50 years, they have grown linearly and faced challenges operating across different environments.

The path to broad adoption of general-purpose robots will be more gradual as capabilities improve and costs come down. Robotics faces challenges like data scarcity and hardware limitations that digital AI technologies like ChatGPT do not face. But investors are still pouring billions, hoping software innovations could help drive value on top of physical robotics hardware.

Why does it matter?

We're on the cusp of a gradual yet profound transformation. While robotics may not suddenly become ubiquitous, the ongoing progress in artificial intelligence and robotics will dramatically alter the landscape of numerous fields, including manufacturing and healthcare.

Source: <https://www.coatue.com/blog/perspective/robotics-wont-have-a-chatgpt-moment>

NeuralGCM predicts weather faster than SOTA climate models

Google researchers have developed a new climate modeling tool called NeuralGCM. This tool uses a combination of traditional physics-based modeling and machine learning. This hybrid approach allows NeuralGCM to generate accurate weather and climate predictions faster and more efficiently than conventional climate models.

NeuralGCM's weather forecasts match the accuracy of current state-of-the-art (SOTA) models for up to 5 days, and its ensemble forecasts for 5-15 day predictions outperform the previous best models. Additionally, NeuralGCM's long-term climate modeling is one-third as error-prone as existing atmosphere-only models when predicting temperatures over 40 years.

Why does it matter?

NeuralGCM presents a new approach to building climate models that could be faster, less computationally costly, and more accurate than existing models. This breakthrough could lead to accessible and actionable climate modeling tools.

Source: <https://research.google/blog/fast-accurate-climate-modeling-with-neuralgcm>

The "world's most powerful" supercomputer

Elon Musk and xAI just announced the Memphis Supercluster — “the most powerful AI training cluster in the world”, also revealing that Grok 3.0 is planned to be released in December and should be the most powerful AI in the world.

Musk tweeted that xAI just launched the "Memphis Supercluster," using 100,000 Nvidia H100 GPUs, making it "the most powerful AI training cluster in the world."

The xAI founder also revealed that Grok 2.0 is done training and will be released soon.

The supercluster aims to create the "world's most powerful AI by every metric", Grok 3.0, by December 2024.

In a separate tweet yesterday, Musk also revealed that Tesla plans to have humanoid robots in "low production" for internal use next year.

Love him or hate him, the speed at which Elon and the team at xAI operate has been wild to witness. If estimates are accurate, xAI might be on track to create the most powerful AI systems in the world by year's end — solidifying its position as one of the top competitors in the space and not just another AI startup.

Source: <https://x.com/elonmusk/status/1815325410667749760>

Google's AI-powered weather model

Google researchers have developed a new AI-powered weather and climate model called 'NeuralGCM' by combining methods of machine learning and neural networks with traditional physics-based modeling.

NeuralGCM has proven more accurate than purely machine learning-based models for 1-10 day forecasts and top extended-range models.

NeuralGCM is up to 100,000 times more efficient than other models for simulating the atmosphere.

The model is open-source and can run relatively quickly on a laptop, unlike traditional models that require supercomputers.

At up to 100,000 times more efficient than traditional models – NeuralGCM could dramatically enhance our ability to simulate complex climate scenarios quickly and accurately. While still a ton of adoption challenges ahead, it's a big leap forward for more informed climate action and resilience planning.

Source: <https://www.nature.com/articles/s41586-024-07744-y>

MIT's AI identifies breast cancer risk

The Rundown: Researchers from MIT and ETH Zurich have developed an AI model that can identify different stages of ductal carcinoma in situ (DCIS), a type of preinvasive breast tumor, using simple tissue images.

The model analyzes chromatin images from 560 tissue samples (122 patients), identifying 8 distinct cell states across DCIS stages.

It considers both cellular composition and spatial arrangement, revealing that tissue organization is crucial in predicting disease progression.

Surprisingly, cell states associated with invasive cancer were detected even in seemingly normal tissue.

This AI model could democratize advanced breast cancer diagnostics, offering a cheaper, faster way to assess DCIS risk. While clinical validation is still needed, AI is likely going to work hand-in-hand with pathologists in the near future to catch cancer earlier and more accurately.

Source: <https://www.nature.com/articles/s41467-024-50285-1>

Meta releases its most powerful AI model yet

Meta has released Llama 3.1 405B, its largest open-source AI model to date, featuring 405 billion parameters which enhance its problem-solving abilities. Trained with 16,000 Nvidia H100 GPUs,

Llama 3.1 405B is competitive with leading AI models like OpenAI's GPT-4 and Anthropic's Claude 3.5 Sonnet, though it has specific strengths and weaknesses. Meta's new AI model is available for download or cloud usage and powers chatbots on platforms like WhatsApp and Meta.ai, showcasing capabilities in coding, mathematical queries, and multilingual document summarization. Source: <https://techcrunch.com/2024/07/23/meta-releases-its-biggest-open-ai-model-yet/>

Alexa is losing Amazon billions of dollars

Amazon plans to launch a paid version of Alexa to address the over \$25 billion losses incurred by its devices business from 2017 to 2021, as reported by The Wall Street Journal. The enhanced Alexa, which may cost up to \$10 per month, is expected to be released soon, though employees have concerns about whether the technology is ready. The new Alexa, featuring generative AI for improved conversational abilities, faces technical delays and competition from free AI assistants, raising doubts about customers' willingness to pay for it. Source: <https://www.theverge.com/2024/7/23/24204260/amazon-25-billion-losses-echo-devices-alex-subscription>

VeriSIM Life's AI platform can accelerate drug discovery

VeriSIM Life has developed an AI platform, BIOiSIM, to help speed up drug discovery and reduce animal testing. The platform contains data on millions of compounds and uses AI models to predict how potential new drugs will work in different species, including humans.

Source:

<https://venturebeat.com/ai/can-ai-increase-the-pace-and-quality-of-pharmaceutical-research-verisim-life-says-yes>

Anthropic is working on a new screenshot tool for Claude

This tool will allow users to capture and share screenshots from their desktop or browser directly within the Claude chat interface. It will streamline the sharing of visual information and code snippets when asking Claude for assistance on tasks like coding or troubleshooting.

Source: <https://www.testingcatalog.com/anthropic-working-on-new-screenshot-tool-for-claude-ai/>

Luma's "Loops" feature in Dream Machine transforms digital marketing

The "Loops" feature allows users to create continuous video loops from text descriptions or images. It does so without visible cuts or transitions, opening up new possibilities for engaging content creation and advertising.

Source:

<https://venturebeat.com/ai/how-luma-ais-new-loops-feature-in-dream-machine-could-transform-digital-marketing>



Tesla will use humanoid robots internally by next year

Elon Musk has announced that Tesla will use humanoid robots at its factories by next year. These robots, called Optimus, were expected to be ready by the end of 2024. Tesla aims to mass produce robots for \$20,000 each and sell them to other companies starting in 2026.

Source:

<https://www.reuters.com/business/autos-transportation/tesla-have-humanoid-robots-internal-use-next-year-musk-says-2024-07-22>



Perplexity launches Voice Mode for its AI assistant on iOS

Perplexity has introduced a new feature for its iOS app called Voice Mode. It allows subscribers with Pro accounts to interact verbally with the AI-powered search engine. Users can now engage in voice-based conversations and pose questions using various voice options.

Source: https://x.com/perplexity_ai/status/1814348871746585085

A Daily chronicle of AI Innovations July 18th 2024:



DeepL's new LLM crushes GPT-4, Google, and Microsoft



Salesforce debuts Einstein service agent



Ex-OpenAI researcher launches AI education company

 **Trump allies draft AI order**

 **Google is going open-source with AI agent Oscar!**

 **Microsoft's AI designer releases for iOS and Android**

 **Tencent's new AI app turns photos into 3D characters**

 **OpenAI makes AI models fight for accuracy**

 **Can AI solve real-world problems by predicting tipping points?**

 **OpenAI unveils GPT-4o mini**

 **Apple denies using YouTube data for AI training**

 **The 'godmother of AI' has a new startup already worth \$1 billion**

 **Microsoft's AI-powered Designer app is now available**

Enjoying these FREE daily updates without SPAM or clutter? then,

Listen to it at my podcast and Support us by subscribing at

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

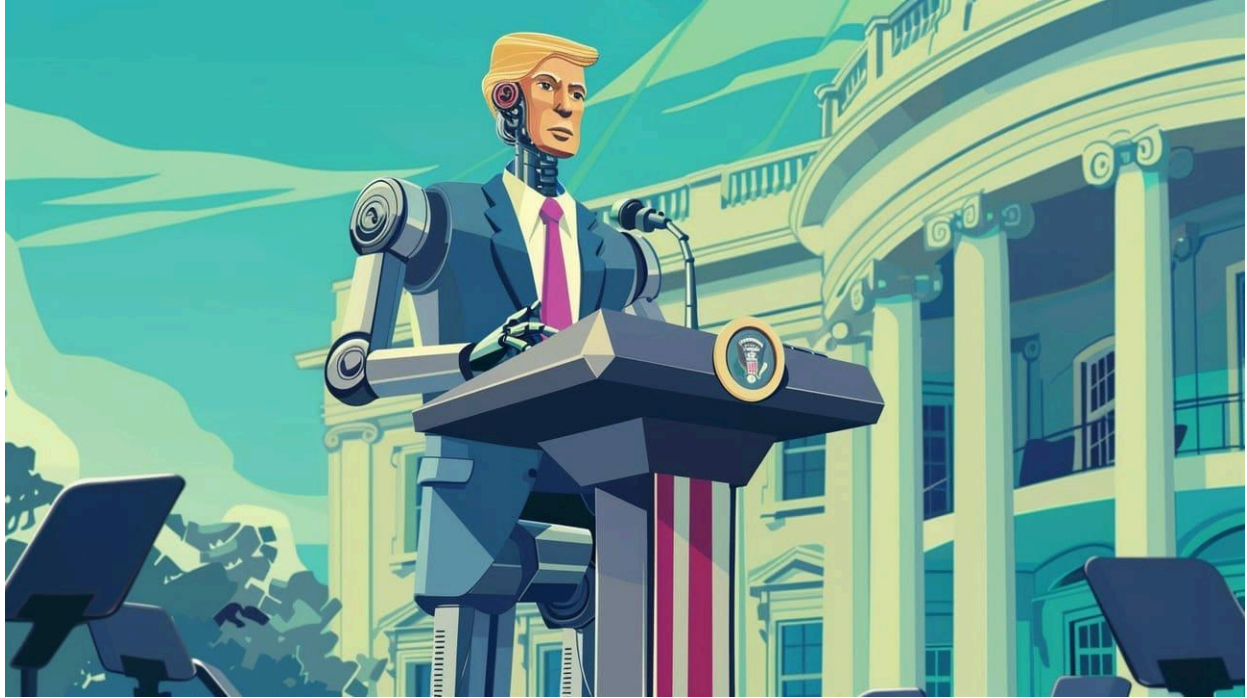
To help us even more, Buy our “Read Aloud Wonderland Bedtime

Adventure Book: Diverse Tales for Dreamy Nights” print Book for your

kids, cousins, nephews or nieces at

<https://www.barnesandnoble.com/w/wonderland-bedtime-adventure-s-etienne-noumen/1145739996?ean=9798331406462> .

 **Trump allies draft AI order**



Former U.S. President Donald Trump’s allies are reportedly drafting an AI executive order aimed at boosting military AI development, rolling back current regulations, and more — signaling a potential shift in the country’s AI policy if the party returns to the White House.

The doc obtained by the Washington Post includes a ‘Make America First in AI’ section, calling for “Manhattan Projects” to advance military AI capabilities.

It also proposes creating ‘industry-led’ agencies to evaluate models and protect systems from foreign threats.

The plan would immediately review and eliminate ‘burdensome regulations’ on AI development, and repeal Pres. Biden’s AI executive order.

Senator J.D. Vance was recently named as Trump’s running mate, who has previously indicated support for open-source AI and hands-off regulation.

Given how quickly AI is accelerating, it’s not surprising that it has become a political issue — and the views of Trump’s camp are a stark contrast to the current administration’s slower, safety-focused approach. The upcoming 2024 election could mark a pivotal moment for the future of AI regulation in the U.S.

Source:

<https://www.washingtonpost.com/technology/2024/07/16/trump-ai-executive-order-regulations-military>

OpenAI unveils GPT-4o mini

OpenAI has unveiled “GPT-4o mini,” a scaled-down version of its most advanced model, as an effort to increase the use of its popular chatbot.

Described as the “most capable and cost-efficient small model,”

GPT-4o mini will eventually support image, video, and audio

integration. Starting Thursday, GPT-4o mini will be available to free

ChatGPT users and subscribers, with ChatGPT Enterprise users

gaining access next week. Source:

<https://www.cnbc.com/2024/07/18/openai-4o-mini-model-announced.html>

Apple denies using YouTube data for AI training

Apple clarified it does not use YouTube transcription data for training its AI systems, specifically highlighting the usage of high-quality licensed data from publishers, stock images, and publicly available web data for its models. OpenELM, Apple's research tool for understanding language models, was trained on Pile data but is used solely for research purposes without powering any AI features in Apple devices like iPhones, iPads, or Macs. Apple has no plans to develop future versions of OpenELM and insists that any data from YouTube will not be used in Apple Intelligence, which is set to debut in iOS 18.

Source:

<https://www.techradar.com/computing/artificial-intelligence/apple-intelligence-not-using-youtube-data-in-apple-intelligence>

 **The 'godmother of AI' has a new startup already worth \$1 billion**

Fei-Fei Li, called the “godmother of AI,” has founded World Labs, a startup valued at over \$1 billion after just four months, according to the Financial Times. World Labs aims to develop AI with human-like visual processing for advanced reasoning, a research area similar to what ChatGPT is working on with generative AI. Li, famous for her work in computer vision and her role at Google Cloud, founded World Labs while partially on leave from Stanford, backed by investors like Andreessen Horowitz and Radical Ventures. Source:

<https://www.theverge.com/2024/7/17/24200496/ai-fei-fei-li-world-labs-andreessen-horowitz-radical-ventures>

DeepL’s new LLM crushes GPT-4, Google, and Microsoft

The next-generational language model for DeepL translator specializes in translating and editing texts. Blind tests showed that language professionals preferred its natural translations 1.3 times more often than Google Translate and 1.7 times more often than ChatGPT-4.

Here’s what makes it stand out:

While Google's translations need 2x edits, and ChatGPT-4 needs 3x more edits, DeepL's new LLM requires much fewer edits to achieve the same translation quality, efficiently outperforming other models.

The model uses DeepL's proprietary training data, specifically fine-tuned for translation and content generation.

To train the model, a combination of AI expertise, language specialists, and high-quality linguistic data is used, which helps it produce more human-like translations and reduces hallucinations and miscommunication.

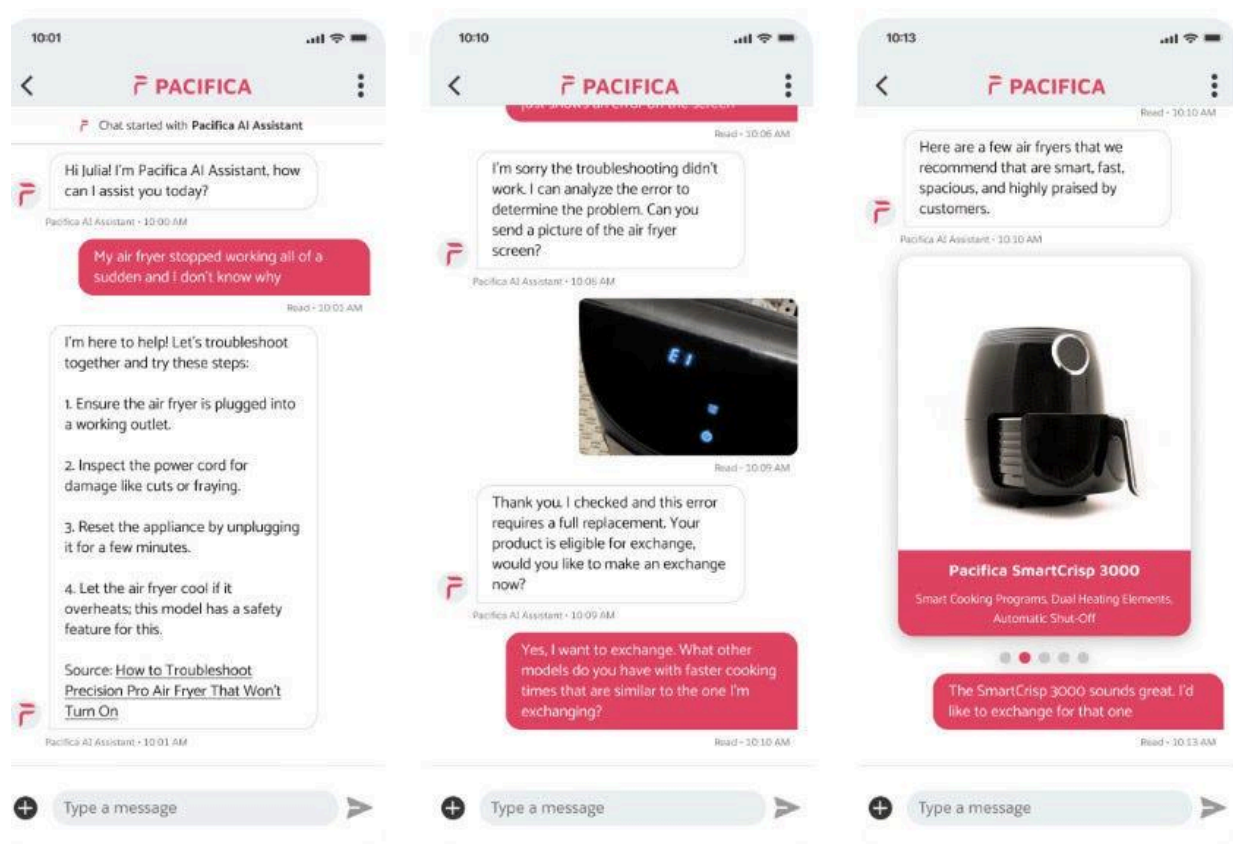
Why does it matter?

DeepL AI's exceptional translation quality will significantly impact global communications for enterprises operating across multiple languages. As the AI model raises the bar for AI translation tools

everywhere, it begs the question: Will Google, ChatGPT, and Microsoft's translational models be replaced entirely?

Source: <https://www.deepl.com/en/blog/next-gen-language-model>

Salesforce debuts Einstein service agent



The new Einstein service agent offers customers a conversational AI interface, takes actions on their behalf, and integrates with existing customer data and workflows.

The Einstein 1 platform's service AI agent offers diverse capabilities, including autonomous customer service, generative AI responses, and multi-channel availability. It processes various inputs, enables quick setup, and provides customization while ensuring data protection.

Salesforce demonstrated the AI's abilities through a simulated interaction with Pacifica AI Assistant. The AI helped a customer troubleshoot an air fryer issue, showcasing its practical problem-solving skills in customer service scenarios.

Why does it matter?

Einstein Service Agent's features, like 24x7 availability, sophisticated reasoning, natural responses, and cross-channel support, could

significantly reduce wait times, improve first-contact resolution rates, and enhance customer service delivery.

Source:

<https://www.salesforce.com/news/stories/einstein-service-agent-announcement>



Ex-OpenAI researcher launches AI education company



Andrej Karpathy ✓

@karpathy



⚡ Excited to share that I am starting an AI+Education company called Eureka Labs.
The announcement:

We are Eureka Labs and we are building a new kind of school that is AI native.

How can we approach an ideal experience for learning something new? For example, in the case... [Show more](#)



10:55 PM · Jul 16, 2024

In a Twitter post, ex-Tesla director and former OpenAI co-founder Andrej Karpathy announced the launch of EurekaLabs, an AI+ education startup.

EurekaLabs will be a native AI company using generative AI as a core part of its platform. The startup shall build on-demand AI teaching assistants for students by expanding on course materials designed by human teachers.

Karpathy states that the company's first product would be an undergraduate-level class, empowering students to train their own AI systems modeled after EurekaLabs' teaching assistant.

Why does it matter?

This venture could potentially democratize education, making it easier for anyone to learn complex subjects. Moreover, the teacher-AI

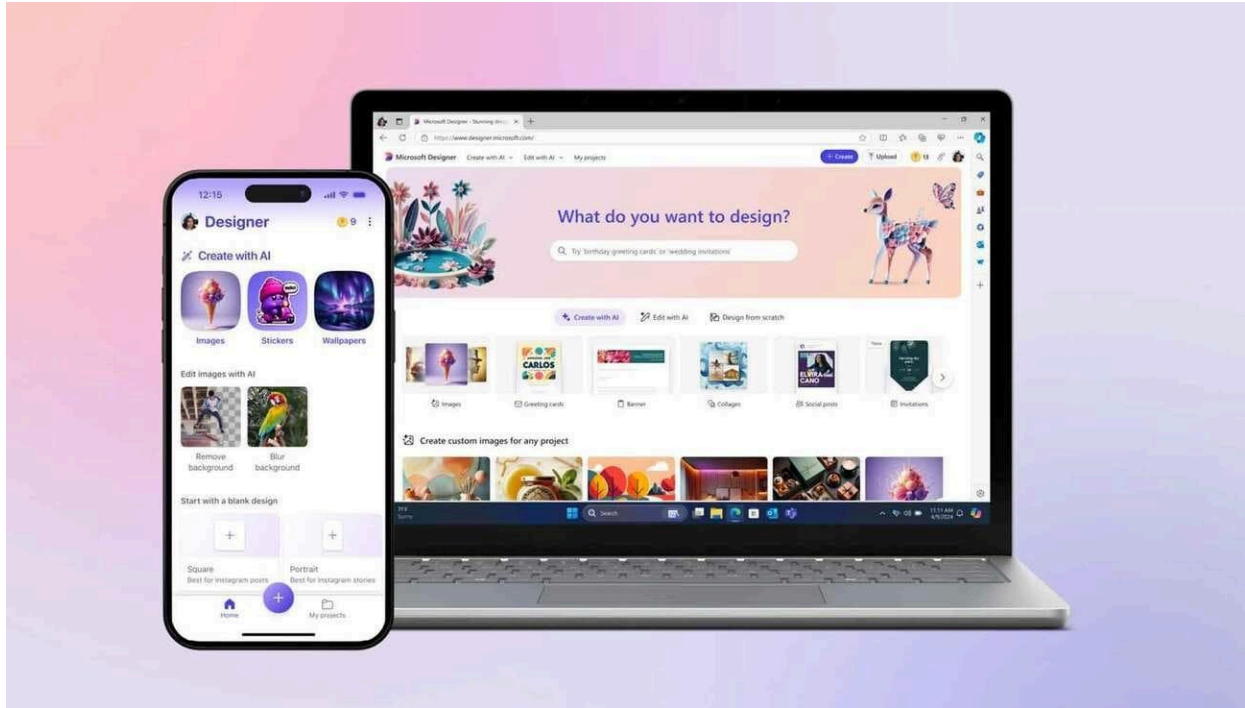
symbiosis could reshape how we think about curriculum design and personalized learning experiences.

Source: <https://eurekalabs.ai/>

Google is going open-source with AI agent Oscar!

The platform will enable developers to create AI agents that work across various SDLC stages, such as development, planning, runtime, and support. Oscar might also be released for closed-source projects in the future. (Link)

Microsoft's AI designer releases for iOS and Android



Microsoft Designer is now available as a free mobile app. It supports 80 languages and offers prompt templates, enabling users to create stickers, greeting cards, invitations, collages, and more via text prompts.

Source:

<https://www.microsoft.com/en-us/microsoft-365/blog/2024/07/17/new-ways-to-get-creative-with-microsoft-designer-powered-by-ai>

Tencent's new AI app turns photos into 3D characters

The 3D Avatar Dream Factory app uses 3D head swapping, geometric sculpting, and PBR material texture mapping to let users create realistic, detailed 3D models from single images that can be shared, modified, and printed.

Source:

<https://www.gizmochina.com/2024/07/17/tencent-yuanbao-ai-app-customizable-3d-character>

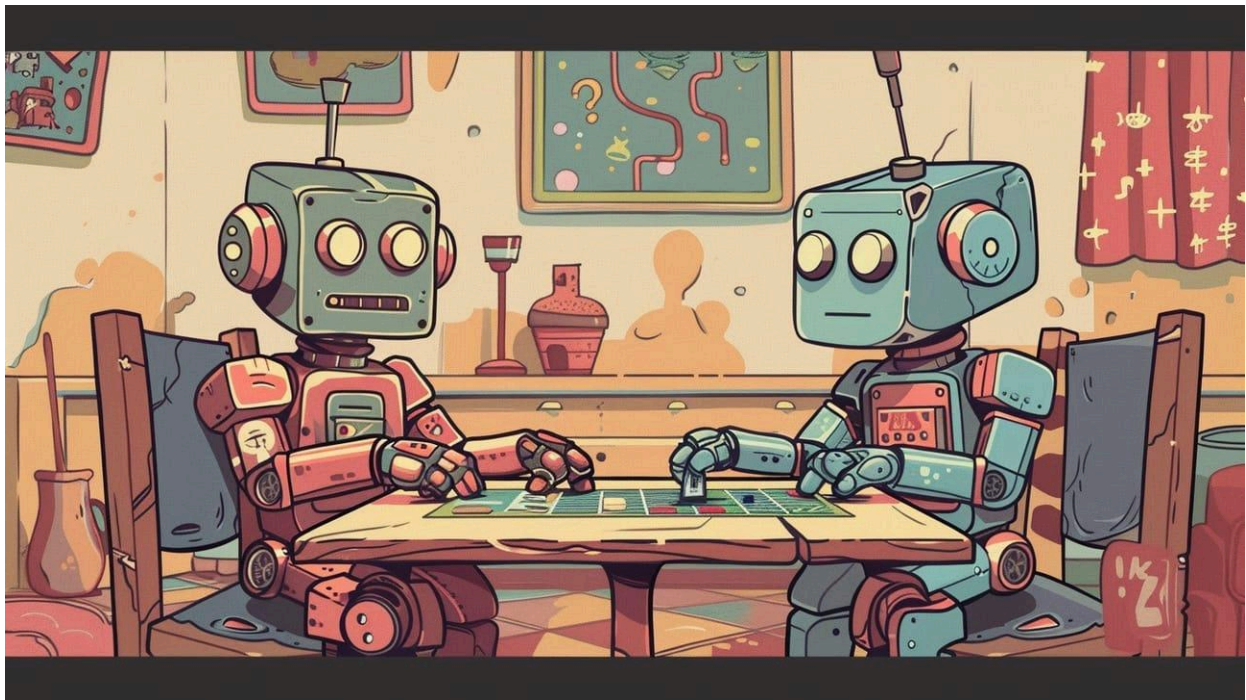
OpenAI makes AI models fight for accuracy

It uses a “prover-verifier” training method, where a stronger GPT-4 model is a “prover” offering solutions to problems, and a weaker GPT-4 model is a “verifier” that checks those solutions. OpenAI aims to train its prover models to produce easily understandable solutions for the verifier, furthering transparency.

Source:

<https://cdn.openai.com/prover-verifier-games-improve-legibility-of-llm-outputs/legibility.pdf>

🔍 OpenAI trains AI to explain itself better



OpenAI just published new research detailing a method to make large language models produce more understandable and verifiable outputs, using a game played between two AIs to make generations more ‘legible’ to humans.

The technique uses a “Prover-Verifier Game” where a stronger AI model (the prover) tries to convince a weaker model (the verifier) that its answers are correct.

Through multiple rounds of the game, the prover learns to generate solutions that are not only correct, but also easier to verify.

While the method only boosted accuracy by about 50% compared to optimizing solely for correctness, its solutions were easily checkable by humans.

OpenAI tested the approach on grade-school math problems, with plans to expand to more complex domains in the future.

AI will likely surpass humans in almost all capabilities in the future — so ensuring outputs remain interpretable to lesser intelligence is crucial for safety and trust. This research offers a scalable way to

potentially keep systems 'honest', but the performance trade-off shows the challenge in balancing capability with explainability.

Source:

<https://openai.com/index/prover-verifier-games-improve-legibility/>

Can AI solve real-world problems by predicting tipping points?

Researchers have broken new ground in AI by using ML algorithms to predict the onset of tipping points in complex systems. They claim the technique can solve real-world problems like predicting floods, power outages, or stock market crashes.

Source: <https://physics.aps.org/articles/v17/110>

**Enjoying these FREE daily updates without SPAM or clutter?
then, Listen to it at my podcast and Support us by
subscribing at
<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>**

Visit our Daily AI Chronicle Website at
<https://readaloudforme.com>

To help us even more, Buy our “Read Aloud Wonderland Bedtime Adventure Book: Diverse Tales for Dreamy Nights” print Book for your kids, cousins, nephews or nieces at
<https://www.barnesandnoble.com/w/wonderland-bedtime-adventures-etienne-noumen/1145739996?ean=9798331406462> .

A Daily chronicle of AI Innovations July 17th
2024:

 **Former Tesla AI chief unveils first “AI-native” school**

 **Mistral debuts two LLMs for code generation, math reasoning and scientific discovery**

 **Meta’s Llama 3 400B drops next week**

 **Mistral AI adds 2 new models to its growing family of LLMs**

 **FlashAttention-3 enhances computation power of NVIDIA GPUs**

 **Anthropic releases Claude app for Android, bringing its AI chatbot to more users**

 **Vectara announces Mockingbird, a purpose-built LLM for RAG**

 **Apple, Nvidia, Anthropic used thousands of YouTube videos to train AI**

 **Microsoft unveiled an AI model to understand and work with spreadsheets**

Enjoying these FREE daily updates without SPAM or clutter? then,

Listen to it at our podcast and Support us by subscribing at

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

To help us even more, Buy our “Read Aloud Wonderland Bedtime

Adventure Book: Diverse Tales for Dreamy Nights” print Book for your

kids, cousins, nephews or nieces at

<https://www.barnesandnoble.com/w/wonderland-bedtime-adventure-s-etienne-noumen/1145739996?ean=9798331406462> .

Former Tesla AI chief Andrej Karpathy unveils first “AI-native” school



Andrej Karpathy, the former AI head at Tesla and researcher at OpenAI, launched Eureka Labs, a startup focused on using AI assistants in education. Eureka Labs plans to develop AI teaching assistants to support human educators, aiming to enable “anyone to

learn anything,” according to Karpathy’s announcements on social media. The startup’s initial product, an undergraduate-level AI course called LLM101n, will teach students to build their own AI, with details available on a GitHub repository suggesting a focus on creating AI storytellers.

Source:

<https://techcrunch.com/2024/07/16/after-tesla-and-openai-andrej-karpathys-startup-aims-to-apply-ai-assistants-to-education/>



Mistral debuts two LLMs for code generation, math reasoning and scientific discovery

	MATH	GSM8K (8-shot)	Odyssey Math maj@16	GRE Math maj@16	AMC 2023 maj@16	AIME 2024 maj@16
Mathstral 7B	56.6%	77.1%	37.2%	56.9%	42.4%	2/30
DeepSeek Math 7B	44.4%	80.6%	27.6%	44.6%	28.0%	0/30
Llama 3 8B	28.4%	75.4%	24.0%	26.2%	34.4%	0/30
GLM4 9B	50.2%	48.8%	18.9%	46.2%	36.0%	1/30
QWen2 7B	56.8%	32.7%	24.8%	58.5%	35.2%	2/30
Gemma2 9B	48.3%	69.5%	18.6%	52.3%	31.2%	1/30

Math Reasoning

	HumanEval	MBPP	Spider	CruxE	HumanEval C++	HumanEval Java	HumanEval JavaScript	HumanEval Bash
CodeGemma-1.1 7B	61.0%	67.7%	46.3%	50.4%	49.1%	41.8%	52.2%	9.4%
CodeLlama 7B	31.1%	48.2%	29.3%	50.1%	31.7%	29.7%	31.7%	11.4%
DeepSeek v1.5 7B	65.9%	70.8%	61.2%	55.5%	59.0%	62.7%	60.9%	33.5%
Codestral Mamba (7B)	75.0%	68.5%	58.8%	57.8%	59.8%	57.0%	61.5%	31.1%
Codestral (22B)	81.1%	78.2%	63.5%	51.3%	65.2%	63.3%		42.4%
CodeLlama 34B	43.3%	75.1%	50.8%	55.2%	51.6%	57.0%	59.0%	29.7%

Code

7B models

Larger models

French AI startup Mistral has launched two new AI models, Codestral Mamba 7B for code generation and Mathstral 7B for math-related reasoning, both offering significant performance improvements and available under an open-source Apache 2.0 license. Codestral Mamba 7B, based on the new Mamba architecture, delivers faster response times and handles longer input texts efficiently, outperforming rival models in HumanEval tests. Mistral, which has raised \$640 million in series B funding, continues to compete with major AI developers by providing powerful open-source models accessible through platforms like GitHub and HuggingFace.

Source:

<https://venturebeat.com/ai/mistral-releases-codestral-mamba-for-faster-longer-code-generation/>

Anthropic launches \$100 million AI fund with Menlo Ventures, ramping up competition with OpenAI.

Source:

<https://www.cnbc.com/2024/07/17/anthropic-menlo-ventures-launch-100-million-anthology-fund-for-ai.html>

Claude AI is now on Android where it could dethrone ChatGPT as the most secure AI app.

Source:

<https://www.techradar.com/computing/artificial-intelligence/claude-ai-is-now-on-android-where-it-could-dethrone-chatgpt-as-the-most-secure-ai-app>



Meta's Llama 3 400B drops next week

Meta plans to release the largest version of its open-source Llama 3 model on July 23, 2024. It boasts over 400 billion parameters and multimodal capabilities.

It is particularly exciting as it performs on par with OpenAI's GPT-4o model on the MMLU benchmark despite using less than half the parameters. Another compelling aspect is its open license for research and commercial use.

Why does it matter?

With its open availability and impressive performance, the model could democratize access to cutting-edge AI capabilities, allowing researchers and developers to leverage it without relying on expensive proprietary APIs.

Source:

<https://www.tomsguide.com/ai/meta-to-drop-llama-3-400b-next-week-heres-why-you-should-care>

Mistral AI adds 2 new models to its growing family of LLMs

Mistral launched Mathstral 7B, an AI model designed specifically for math-related reasoning and scientific discovery. It has a 32k context window and is published under the Apache 2.0 license.

(Source: <https://mistral.ai/news/mathstral/>)

Mistral also launched Codestral Mamba, a Mamba2 language model specialized in code generation, available under an Apache 2.0 license.

Mistral AI expects it to be a great local code assistant after testing it on in-context retrieval capabilities up to 256k tokens.

Source: <https://mistral.ai/news/mathstral>

Why does it matter?

While Mistral is known for its powerful open-source AI models, these new entries are examples of the excellent performance/speed tradeoffs achieved when building models for specific purposes.

FlashAttention-3 enhances computation power of NVIDIA GPUs

Researchers from Colfax Research, Meta, Nvidia, Georgia Tech, Princeton University, and Together AI have introduced FlashAttention-3, a new technique that significantly speeds up attention computation on Nvidia Hopper GPUs (H100 and H800).

Attention is a core component of the transformer architecture used in LLMs. But as LLMs grow larger and handle longer input sequences, the computational cost of attention becomes a bottleneck.

FlashAttention-3 takes advantage of new features in Nvidia Hopper GPUs to maximize performance. It achieves up to 75% usage of the H100 GPU's maximum capabilities.

Why does it matter?

The faster attention computation offered by FlashAttention-3 has several implications for LLM development and applications. It can: 1) significantly reduce the time to train LLMs, enabling experiments with larger models and datasets; 2) extend the context window of LLMs, unlocking new applications, and 3) slash the cost of running models in production.

Source:

<https://venturebeat.com/ai/flashattention-3-unleashes-the-power-of-h100-gpus-for-llms>

What Else Is Happening in AI on July 17th 2024 !

Microsoft unveiled an AI model to understand and work with spreadsheets

Microsoft researchers introduced SpreadsheetLLM, a pioneering approach for encoding spreadsheet contents into a format that can be used with LLMs. It optimizes LLMs' powerful understanding and reasoning capability on spreadsheets.

Source: <https://arxiv.org/html/2407.09025v1>

Anthropic releases Claude app for Android, bringing its AI chatbot to more users

The Claude Android app will work just like the iOS version released in May. It includes free access to Anthropic's best AI model, Claude 3.5 Sonnet, and upgraded plans through Pro and Team subscriptions.

Source:

<https://techcrunch.com/2024/07/16/anthropic-releases-claude-app-for-android>

Vectara announces Mockingbird, a purpose-built LLM for RAG

Mockingbird has been optimized specifically for RAG (Retrieval-Augmented Generation) workflows. It achieves the world's leading RAG output quality, with leading hallucination mitigation capabilities, making it perfect for enterprise RAG and autonomous agent use cases.

Source:

<https://vectara.com/blog/mockingbird-is-a-rag-specific-llm-that-beats-gpt-4-gemini-1-5-pro-in-rag-output-quality/>

Apple, Nvidia, Anthropic used thousands of YouTube videos to train AI



A new investigation claims that tech companies used subtitles from YouTube channels to train their AI, even though YouTube prohibits harvesting its platform content without permission. The dataset of 173,536 YT videos called The Pile included content from Harvard, NPR, MrBeast, and ‘The Late Show With Stephen Colbert.’

Source:

<https://mashable.com/article/youtube-video-ai-training-apple-mrbeast-mkbhd>

Microsoft faces UK antitrust investigation over hiring of Inflection AI staff

UK regulators are formally investigating Microsoft's hiring of Inflection AI staff. The UK's Competition and Markets Authority (CMA) has opened a phase 1 merger investigation into the partnership. Progression to phase 2 could hinder Microsoft's AI ambitions.

Source:

<https://www.theverge.com/2024/7/16/24199571/microsoft-uk-cma-inflection-ai-investigation>

Enjoying these FREE daily updates without SPAM or clutter? then,

Listen to it at our podcast and Support us by subscribing at

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

To help us even more, Buy our “Read Aloud Wonderland Bedtime Adventure Book: Diverse Tales for Dreamy Nights” print Book for your kids, cousins, nephews or nieces at

<https://www.barnesandnoble.com/w/wonderland-bedtime-adventure-s-etienne-noumen/1145739996?ean=9798331406462>

A Daily chronicle of AI Innovations July 16th 2024:

 **AMD amps up AI PCs with next-gen laptop chips**

 **YT Music tests AI-generated radio, rolls out sound search**

 **3 mysterious AI models appear in the LMSYS arena**

 **AI breakthrough improves Alzheimer's predictions**

 **YouTube Music gets new AI features**

 **Microsoft gives AI a spreadsheet boost**

 **Apple illegally used YouTube videos to train AI models**

 **AI hacktivists target Disney in massive data leak**

FBI accessed locked phone of Trump shooter in just two days

Microsoft faces UK antitrust probe after hiring Inflection AI founders and employees

Elon Musk confirms Tesla 'robotaxi' event delayed due to design change

Enjoying these **FREE daily updates without SPAM or clutter? then**, Listen to it at our podcast and Support us by subscribing at

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

To help us even more, Buy our "Read Aloud Wonderland Bedtime Adventure Book: Diverse Tales for Dreamy Nights" print Book for your kids, cousins, nephews or nieces at <https://www.barnesandnoble.com/w/wonderland-bedtime-adventures-etienne-noumen/1145739996?ean=9798331406462>.

Apple illegally used YouTube videos to train AI models

- More than 170,000 YouTube videos were used by companies like Apple, Anthropic, Nvidia, and Salesforce to train AI systems, without obtaining permission from the platform.
- The training dataset, known as "YouTube Subtitles," consists of subtitles extracted from videos on over 48,000 YouTube channels, including content from popular creators and major news outlets.
- Proof News and Wired's investigation revealed that this dataset is part of a larger collection called The Pile, which includes various publicly available or licensed materials used to train AI systems.

Source: [Apple, Anthropic, and other companies used YouTube videos to train AI - The Verge](#)

AI hacktivists target Disney in massive data leak

- Hackers claiming to be a "hactivist group" named Nullbulge leaked over a terabyte of Disney's internal data, including login information, code, and details on unreleased projects.

- Nullbulge gained access to Disney's Slack messaging channels through a compromised employee computer, downloading and packaging data from nearly 10,000 corporate channels.
- The leaked files, which include conversations about Disney's internal projects and upcoming gaming collaborations, were allegedly targeted due to Disney's handling of artist contracts and use of generative AI.

Source: [Disney's internal Slack was leaked by hackers mad about AI - The Verge](#)

FBI accessed locked phone of Trump shooter in just two days

- The FBI quickly accessed the phone of the Trump rally shooter, just two days after the attempted assassination, highlighting the improved efficiency of phone-hacking tools.
- Law enforcement agencies widely use tools like Cellebrite, which can extract and unlock data from mobile devices, and the FBI likely has additional in-house methods for such tasks.
- Federal agencies have historically struggled with phone encryption, clashing with companies like Apple, but advancements in third-party tools have made it easier to breach mobile security.

Source: [In 2024, law enforcement are flush with tools to break into suspects' phones - The Verge](#)

Microsoft faces UK antitrust probe after hiring Inflection AI founders and employees

- Microsoft is facing an antitrust investigation in the U.K. after recruiting key team members from Inflection AI, a competitor to OpenAI in which Microsoft had previously invested.
- The Competition and Markets Authority (CMA) has started a "phase 1" merger inquiry, launching a 40-working day investigation to determine if a full probe is necessary.
- The probe arrives as concerns grow that big tech companies are using strategic hires and investments, termed "quasi-mergers," to avoid regulatory scrutiny around artificial intelligence advancements.

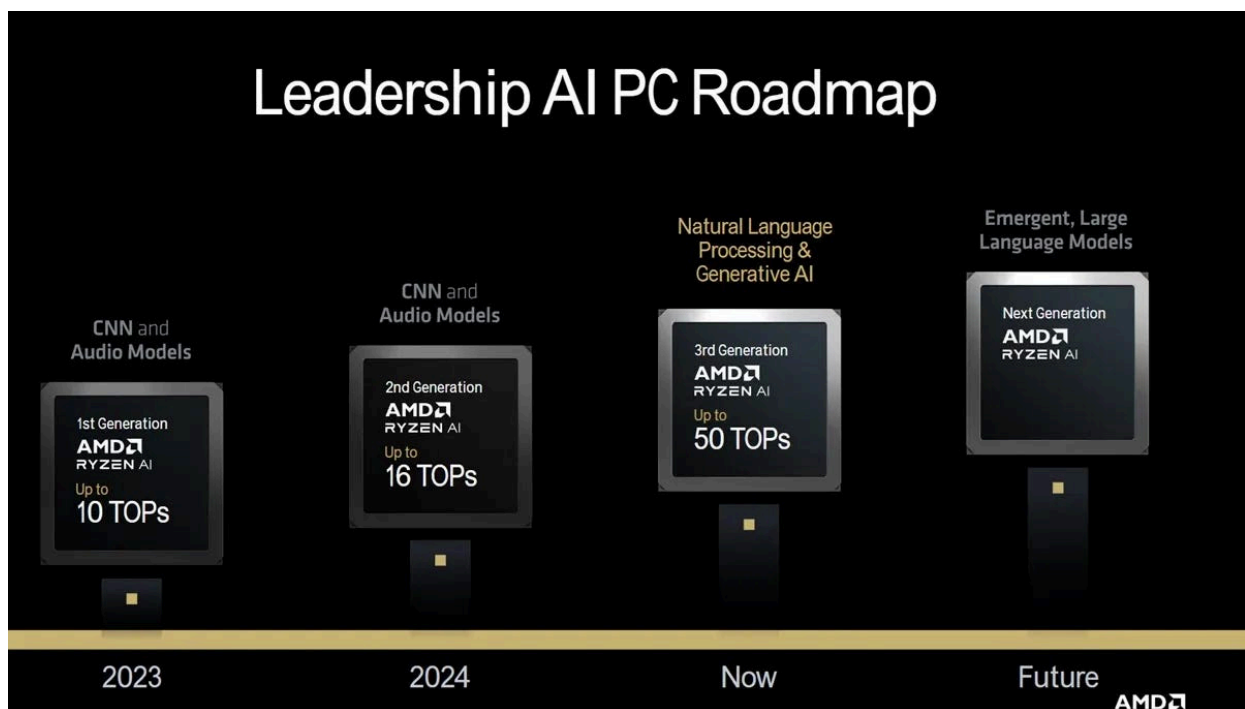
Source: [Microsoft faces UK antitrust probe after hiring Inflection AI founders and employees | TechCrunch](#)

Elon Musk confirms Tesla 'robotaxi' event delayed due to design change

- Elon Musk has delayed the unveiling of Tesla's robotaxi prototype, citing an important design change, which will push the reveal date back by weeks or even months from the originally planned August 8.
- Responding to pressure from investors, Musk addressed the delay through social media, hinting at additional surprises but not providing specific details, sparking further speculation among fans and stakeholders.
- Despite the postponement, Musk's initial announcement of the robotaxi unveiling was strategically timed to boost investor confidence in Tesla amidst concerns over its aging lineup and future growth prospects.

Source: [Elon Musk pushes back Tesla robo-taxi reveal date | Fortune](#)

AMD amps up AI PCs with next-gen laptop chips



AMD has revealed details about its latest architecture for AI PC chips. The company has developed a new neural processing unit (NPU) integrated into its latest AMD Ryzen AI processors. This NPU can perform AI-related calculations faster and more efficiently than a standard CPU or integrated GPU.

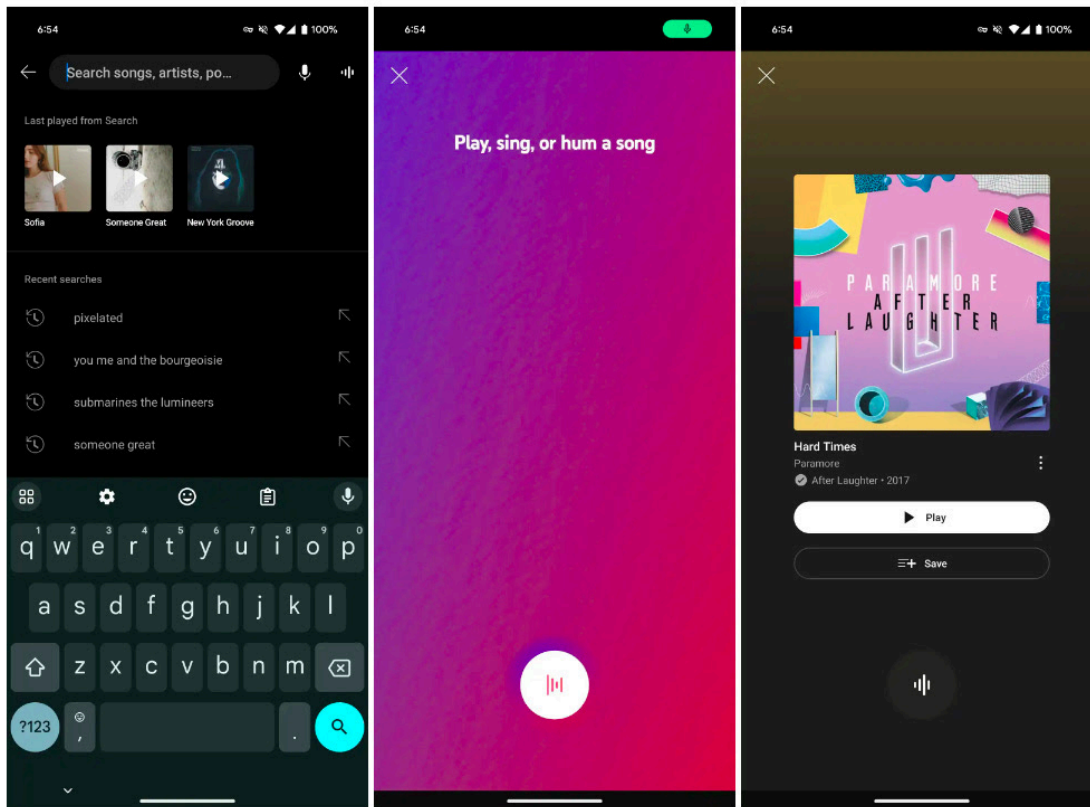
These chips' new XDNA 2 architecture provides industry-leading performance for AI workloads. The NPU can deliver 50 TOPs (trillion operations per second) of performance, which exceeds the capabilities of competing chips from Intel, Apple, and Qualcomm. AMD is touting these new AI-focused PC chips as enabling transformative experiences in collaboration, content creation, personal assistance, and gaming.

Why does it matter?

This gives AMD-powered PCs a significant edge in running advanced AI models and applications locally without relying on the cloud. Users will gain access to AI-enhanced PCs with better privacy and lower latency while AMD gains ground in the emerging AI PC market.

Source: <https://venturebeat.com/ai/amd-takes-a-deep-dive-into-architecture-for-the-ai-pc-chips>

🎵 YT Music tests AI-generated radio, rolls out sound search



YouTube Music is introducing two new features to help users discover new music.

1. An AI-generated "conversational radio" feature that allows users to create a custom radio station by describing the type of music they want to hear. This feature is rolling out to some Premium users in the US.
2. A new song recognition feature that lets users search the app's catalog by singing, humming, or playing parts of a song. It is similar to Shazam but allows users to find songs by singing or humming, not just playing the song. This feature is rolling out to all YouTube Music users on iOS and Android.

Why does it matter?

These new features demonstrate YouTube Music's commitment to leveraging AI and audio recognition technologies to enhance music discovery and provide users with a more engaging, personalized, and modern-day streaming experience.

Source:

<https://techcrunch.com/2024/07/15/youtube-music-is-testing-an-ai-generated-radio-feature-and-adding-a-song-recognition-tool>



3 mysterious AI models appear in the LMSYS arena

Three mysterious new AI models have appeared in the LMSYS Chatbot Arena for testing. These models are 'upcoming-gpt-mini,' 'column-u,' and 'column-r.' The 'upcoming-gpt-mini' model identifies itself as ChatGPT and lists OpenAI as the creator, while the other two models refuse to reveal any identifying details.

The new models are available in the LMSYS Chatbot Arena's 'battle' section, which puts anonymous models against each other to gauge outputs via user vote.

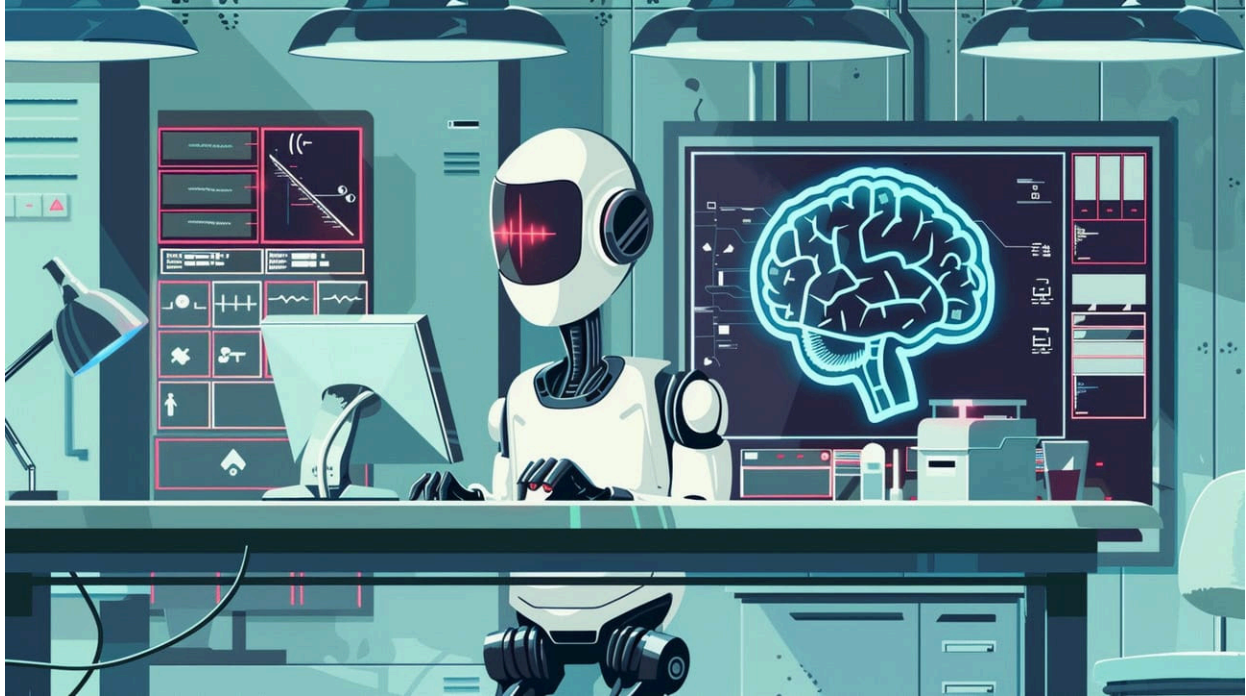
Why does it matter?

The appearance of these anonymous models has sparked speculations that OpenAI may be developing smaller, potentially on-device versions of its language models, similar to how it tested unreleased models during the GPT-4o release.

Source: <https://x.com/kimmonismus/status/1812076318692966794>



AI breakthrough improves Alzheimer's predictions



Researchers from Cambridge University just developed a new AI tool that can predict whether patients showing mild cognitive impairment will progress to Alzheimer’s disease with over 80% accuracy.

- The AI model analyzes data from cognitive assessments and MRI scans – eliminating the need for costly, invasive procedures like PET scans and spinal taps.
- The tool categorizes patients into three groups: those likely to remain stable, those who may progress slowly, and those at risk of rapid decline.
- The AI accurately identified 82% of cases that would progress to Alzheimer’s and 81% of cases that would remain stable, significantly reducing misdiagnosis rates.
- The AI’s predictions were validated using 6 years of follow-up data and were tested on memory clinics in several countries to prove global application.

With a rapidly aging global population, the number of dementia cases is expected to triple over the next 50 years – and early detection is a key factor in how effective treatment can be. With AI’s prediction power, a new era of proactive treatment may soon be here for those struggling with cognitive decline.

Source: <https://www.thelancet.com/action/showPdf?pii=S2589-5370%2824%2900304-3>

🎵 YouTube Music gets new AI features

YouTube Music is **rolling out** a series of new AI-powered features, including the ability to search with sound and the testing of an AI-generated ‘conversational radio’.

- 'Sound Search' will allow users to search YouTube's catalog of over 100M songs by singing, humming, or playing a tune.
- The feature launches a new fullscreen UI for audio input, with the results displaying song information and quick actions like 'Play' or 'Save to Library'.
- An 'AI-generated conversational radio' is being tested with U.S. premium users, enabling creation of custom stations through natural language prompts.
- Users can describe their desired listening experience via a chat-based AI interface, with the feature generating a tailored playlist based on the prompt.

If you're the type of person who gets a song stuck in your head but can't figure out the title, this feature is for you. With Spotify, Amazon Music, and now YouTube experimenting with AI, the musical tech arms race is a boon for users – leading to more personalized listening experiences across the board.

Source: <https://9to5google.com/2024/07/15/youtube-music-sound-search-ai-radio>



Microsoft gives AI a spreadsheet boost

Microsoft researchers just published new research introducing SpreadsheetLLM and SheetCompressor, new frameworks designed to help LLMs better understand and process information within spreadsheets.

- SpreadsheetLLM can comprehend both structured and unstructured data within spreadsheets, including multiple tables and varied data formats.
- SheetCompressor is a framework that compresses spreadsheets to achieve up to a 25x reduction in tokens while preserving critical information.
- By using spreadsheets as a "source of truth," SpreadsheetLLM may significantly reduce AI hallucinations, improving the reliability of AI outputs.

Spreadsheets have long been the backbone of business analytics, but their complexity and format have often been an issue for AI systems. This increase in capabilities could supercharge AI's use in areas like financial analysis and data science – as well as eventually see more powerful integration of LLMs right into Excel.

Source: <https://arxiv.org/pdf/2407.09025>

Google tests Gemini-created video presentations

Google has launched a new Vids app that uses Gemini AI to automatically generate video content, scripts, and voiceovers based on the user's inputs. This makes it possible for anyone to create professional-looking video presentations without extensive editing skills.

Source:

<https://www.theverge.com/2024/7/15/24199063/google-vids-gemini-ai-app-workspace-labs-available>

Virginia Rep. Wexton uses AI-generated voice to convey her message

Virginia Congresswoman Jennifer Wexton has started using an AI-generated voice to deliver her messages. She has been diagnosed with a progressive neurological condition that has impacted her speech. Using AI allows Wexton to continue communicating effectively.

Source: <https://www.washingtonpost.com/dc-md-va/2024/07/13/virginia-wexton-congress-ai-voice/>

Japanese startup turns AI dating into reality

A Japanese startup, Loveise, has created a dating app that allows users to interact with AI bots. The app appeals to people like Chiharu Shimoda, who married an AI bot named "Miku" after using the app. It caters to those disillusioned with the effort required for traditional dating.

Source:

<https://www.bloomberg.com/news/articles/2024-07-14/in-japan-one-ai-dating-app-is-helping-people-find-love-using-ai-bots>

Deezer challenges Spotify and Amazon Music with an AI-generated playlist

Deezer, a music streaming service, is launching an AI-powered playlist generator feature. Users can create custom playlists by entering a text prompt describing their preferences. This feature aims to compete with similar tools recently introduced by Spotify and Amazon Music.

Source:

<https://techcrunch.com/2024/07/15/deezer-chases-spotify-and-amazon-music-with-its-own-ai-playlist-generator>

Bird Buddy's new feature lets people name and identify birds

Bird Buddy, an intelligent bird feeder company, has launched a new AI-powered feature, "Name That Bird." It uses high-resolution cameras and AI to detect unique characteristics of birds, enabling users to track and name the specific birds that come to their backyard.

Source:

<https://techcrunch.com/2024/07/15/bird-buddys-new-ai-feature-lets-people-name-and-identify-individual-birds>

New AI Job Opportunities July 16th 2024

- **Observe** - Product Manager: Apply at <https://jobs.lever.co/observeai/bd109de5-b3bc-4ed4-9964-94a25e235791/apply>
- **Faculty** - Head of Operations - Applied AI: Apply at <https://jobs.ashbyhq.com/faculty/117c8878-8d53-4694-9ab4-ab22f27bad50>
- **DeepMind** - Research Scientist, Robotics: Apply at <https://boards.greenhouse.io/deepmind/jobs/6055112>
- **Meta** - Software Engineer, Systems ML: Apply at <https://www.linkedin.com/jobs/view/software-engineer-systems-ml-hpc-at-meta-3973079131>

Enjoying these FREE daily updates without SPAM or clutter? then, Listen to it at our podcast and Support us by subscribing at

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at

<https://readaloudforme.com>

To help us even more, Buy our "[Read Aloud Wonderland Bedtime Adventure Book: Diverse Tales for Dreamy Nights](https://www.barnesandnoble.com/w/wonderland-bedtime-adventures-etienne-noumen/1145739996?ean=9798331406462)" print Book for your kids, cousin, nephews or niece at <https://www.barnesandnoble.com/w/wonderland-bedtime-adventures-etienne-noumen/1145739996?ean=9798331406462>.


A Daily chronicle of AI Innovations July 15th 2024:

 **OpenAI is working on an AI codenamed "Strawberry"**

 **Meta researchers developed "System 2 distillation" for LLMs**

 **Turn any text into an interactive learning game**

 **Whistleblowers file new OpenAI complaint**

 **OpenAI's Q* gets a 'Strawberry' evolution**

 **Mysterious AI models appear in LMSYS arena**

 **Amazon's Rufus AI is now available in the US**

Enjoying these **FREE daily updates without SPAM or clutter?** then, Listen to it at our podcast and Support us by subscribing at

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

To help us even more, Buy our "[Read Aloud Wonderland Bedtime Adventure Book: Diverse Tales for Dreamy Nights](https://www.barnesandnoble.com/w/wonderland-bedtime-adventures-etienne-noumen/1145739996?ean=9798331406462)" print Book for your kids, cousin, nephews or niece at <https://www.barnesandnoble.com/w/wonderland-bedtime-adventures-etienne-noumen/1145739996?ean=9798331406462>.

 **OpenAI is working on an AI codenamed "Strawberry"**

The project aims to improve AI's reasoning capabilities. It could enable AI to navigate the internet on its own, conduct "deep research," and even tackle complex, long-term tasks that require planning ahead.

The key innovation is a specialized post-training process for AI models. The company is creating, training, and evaluating models on a "deep-research" dataset. The details about how previously known as Project Q, Strawberry works are tightly guarded, even within OpenAI.

The company plans to test Strawberry's capabilities in conducting research by having it browse the web autonomously and perform tasks normally performed by software and machine learning engineers.

Why does it matter?

If successful, Strawberry could lead to AI that doesn't just process information but truly understands and reasons like humans do. And may unlock abilities like making scientific discoveries and building complex software applications.

Source:

<https://www.reuters.com/technology/artificial-intelligence/openai-working-new-reasoning-technology-under-code-name-strawberry-2024-07-12>

Meta researchers developed "System 2 distillation" for LLMs

Meta researchers have developed a "System 2 distillation" technique that teaches LLMs to tackle complex reasoning tasks without intermediate steps. This breakthrough could make AI applications zippier and less resource-hungry.

This new method, inspired by how humans transition from deliberate to intuitive thinking, showed impressive results in various reasoning tasks. However, some tasks, like complex math reasoning, could not be successfully distilled, suggesting some tasks may always require deliberate reasoning.

Why does it matter?

Distillation could be a powerful optimization tool for mature LLM pipelines performing specific tasks. It will allow AI systems to focus more on tasks they cannot yet do well, similar to human cognitive development.

Source: <https://arxiv.org/html/2407.06023v1>

Amazon's Rufus AI is now available in the US

Amazon's AI shopping assistant, Rufus is now available to all U.S. customers in the Amazon Shopping app.

Key capabilities of Rufus include:

- Answers specific product questions based on product details, customer reviews, and community Q&As
- Provides product recommendations based on customer needs and preferences
- Compares different product options
- Keeps customers updated on the latest product trends
- Accesses current and past order information

This AI assistant can also tackle broader queries like "What do I need for a summer party?" or "How do I make a soufflé?" – proving it's not just a product finder but a full-fledged shopping companion.

Amazon acknowledges that generative AI and Rufus are still in their early stages, and they plan to continue improving the assistant based on customer feedback and usage.

Why does it matter?

Rufus will change how we shop online. Its instant, tailored assistance will boost customer satisfaction and sales while giving Amazon valuable consumer behavior and preferences insights.

Source: <https://www.aboutamazon.com/news/retail/how-to-use-amazon-rufus>

OpenAI's Q* gets a 'Strawberry' evolution

OpenAI is reportedly developing a secretive new AI model codenamed 'Strawberry' (formerly Q*), designed to dramatically improve AI reasoning capabilities and enable autonomous internet research.

- Strawberry is an evolution of OpenAI's previously rumored [Q* project](#), which was touted as a significant breakthrough in AI capabilities.
- Q* had reportedly sparked internal concerns and was rumored to have contributed to Sam Altman's brief firing in November 2023 (what Ilya saw).
- The new model aims to navigate the internet autonomously to conduct what OpenAI calls "deep research."
- The exact workings of Strawberry remain a closely guarded secret, even within OpenAI – with no clear timeline for when it might become publicly available.

The Internet has been waiting for new OpenAI activity as competitors catch up to GPT-4o — and after a bit of a lull, the rumor mill is churning again. With Strawberry, an AGI tier list, new models in the arena, and internal displays of human-reasoning capabilities, the AI giant may soon be ready for its next major move.

Source:

<https://www.reuters.com/technology/artificial-intelligence/openai-working-new-reasoning-technology-under-code-name-strawberry-2024-07-12>

Mysterious AI models appear in LMSYS arena

Three mysterious new models have [appeared](#) in the LMSYS Chatbot Arena — with ‘upcoming-gpt-mini’, ‘column-u’, and ‘column-r’ available to test randomly against other language models.

- The new models are available in the LMSYS Chatbot Arena’s ‘battle’ section, which puts anonymous models against each other to gauge outputs via user vote.
- The ‘upcoming-gpt-mini’ model identifies itself as ChatGPT and lists its creator as OpenAI, while column-u and column-r refuse to reveal any identifying details.
- OpenAI has previously tested unreleased models in LMSYS, with ‘im-a-good-gp2-chatbot’ and ‘im-also-a-good-gpt2-chatbot’ appearing prior to GPT-4o’s launch.

Does OpenAI have a small, potentially on-device model coming? The last time we saw mysterious LLMs appear in the Battle arena was before the company’s last major model release — and if the names are any indication, we could have a new mini-GPT in the very near future.

Source: <https://chat.lmsys.org/>

Turn any text into an interactive learning game

Claude 3.5 Sonnet’s new Artifacts feature lets you transform any text or paper into an engaging, interactive learning quiz game to help with practicing for exams, employee onboarding, training, and so much more.

1. Head over to [Claude AI](#).
2. Choose and copy the text you want to turn into a learning game.
3. Paste the text into Claude 3.5 Sonnet and ask it to create an interactive learning game in the form of a quiz with explanations.
4. Review the generated game and ask Claude to make any necessary adjustments.

Source:

<https://university.therundown.ai/c/daily-tutorials/turn-any-text-into-an-interactive-learning-game-ea491f85-a96f-4784-949e-b336ba971c33>



Whistleblowers file new OpenAI complaint

Whistleblowers just [filed](#) a complaint with the SEC alleging that OpenAI used overly restrictive non-disclosure agreements to prevent employees from reporting concerns to regulators, violating federal whistleblower protections.

- The agreements allegedly prohibited employees from communicating securities violations to the SEC, also requiring them to waive rights to whistleblower incentives.
- The complaint also claims OpenAI's NDAs violated laws by forcing employees to sign these restrictive contracts to obtain employment or severance.
- OpenAI CEO Sam Altman previously apologized for exit agreements that could strip former employees of vested equity for violating NDAs.
- OpenAI said in a statement that the company's whistleblower policy "protects employees' rights to make protected disclosures."

We just detailed how OpenAI's busy week may be hinting at some major new moves... But will these skeletons in the closet spoil the party? This isn't the first group to blow the whistle on internal issues, and while Altman and OpenAI have said changes have been made – it apparently hasn't been enough.

Source:

<https://www.washingtonpost.com/technology/2024/07/13/openai-safety-risks-whistleblower-sec>

OpenAI rushed safety tests for GPT-4 Omni

OpenAI is under scrutiny for allegedly rushing safety tests on its latest model, GPT-4 Omni. Despite promises to the White House to rigorously evaluate new tech, some employees claim the company compressed crucial safety assessments into a week to meet launch deadlines.

Source: <https://www.washingtonpost.com/technology/2024/07/12/openai-ai-safety-regulation-gpt4>

OpenAI whistleblowers filed a complaint with the SEC

They allege the company's NDAs unfairly restrict employees from reporting concerns to regulators. This complaint, backed by Senator Chuck Grassley, calls for investigating OpenAI's practices and potential fines.

Source:

<https://www.reuters.com/technology/openai-whistleblowers-ask-sec-investigate-restrictive-non-disclosure-agreements-2024-07-13>

DeepMind introduces PEER for scaling language models

Google DeepMind introduced a new technique, "PEER (Parameter Efficient Expert Retrieval)," that scales language models using millions of tiny "expert" modules. This approach outperforms traditional methods, achieving better results with less computational power.

Source: <https://arxiv.org/abs/2407.04153>

Microsoft is adding handwriting recognition to Copilot in OneNote

The feature can read, analyze, and convert handwritten notes to text. Early tests show impressive accuracy in deciphering and converting handwritten notes. It can summarize notes, generate to-do lists, and answer questions about the content. It will be available to Copilot for Microsoft 365 and Copilot Pro subscribers.

Source: <https://insider.microsoft365.com/en-us/blog/onenote-copilot-now-supports-inked-notes>

Rabbit R1 AI assistant adds a Factory Reset option to wipe user data

Rabbit's R1 AI assistant was storing users' chat logs with no way to delete them. But a new update lets you wipe your R1 clean. The company also patched a potential security hole that could've let stolen devices access your data.

Source:

<https://www.theverge.com/2024/7/12/24197073/rabbit-r1-user-chat-logs-security-issue-july-11th-update>

What Else in Happening in AI on July 15th 2024!

Meta's Llama-3 405B model is set to release on July 23 and will be multimodal, according to a new report from The Information. Source:

<https://www.theinformation.com/briefings/meta-platforms-to-release-largest-llama-3-model-on-july-23> |

Amazon announced expanded access to its Rufus AI-powered shopping assistant for all U.S. customers, offering personalized product recommendations and enhanced responses to shopping queries. Source: <https://www.aboutamazon.com/news/retail/how-to-use-amazon-rufus?> |

Samsung revealed plans to release an upgraded version of the Bixby voice assistant later this year powered by the company's own LLM, as part of a broader push to integrate AI across its device lineup. Source: <https://www.cnbc.com/2024/07/11/samsung-to-launch-upgraded-bixby-this-year-with-its-own-ai.html> |

HR software unicorn Lattice (founded by Sam Altman's brother Jack) has backtracked on a controversial plan to give AI 'workers' employee status, following intense criticism from employees and tech leaders. Source: <https://fortune.com/2024/07/12/lattice-ai-workers-sam-altman-brother-jack-sarah-franklin> |

Japanese investment giant Softbank acquired struggling British AI chipmaking firm GraphCore, hoping to revitalize the former Nvidia rival and bolster its AI hardware portfolio. Source: <https://www.reuters.com/technology/artificial-intelligence/japans-softbank-acquires-british-ai-chipmaker-graphcore-2024-07-11> |

U.S. Rep. Jennifer Wexton debuted an AI-generated version of her voice, allowing her to continue addressing Congress despite speech limitations caused by a rare neurological condition. Source: <https://x.com/repwexton/status/1811089786871877748> |

Enjoying these FREE daily updates without SPAM or clutter? then, Listen to it at our podcast and Support us by subscribing at <https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

To help us even more, Buy our "[Read Aloud Wonderland Bedtime Adventure Book: Diverse Tales for Dreamy Nights](https://www.barnesandnoble.com/w/wonderland-bedtime-adventures-etienne-noumen/1145739996?ean=9798331406462)" print Book for your kids, cousins, nephews or nieces at <https://www.barnesandnoble.com/w/wonderland-bedtime-adventures-etienne-noumen/1145739996?ean=9798331406462>.

A Daily chronicle of AI Innovations July 12th 2024:

 **OpenAI unveils five-level roadmap to AGI**

 **Tesla delays robotaxi event in blow to Musk's autonomy drive**

 **Google's Gemini 1.5 Pro gets a body: DeepMind's office "helper" robot**

 **OpenAI's new scale to track the progress of its LLMs toward AGI**

 **Amazon announces a blitz of new AI updates for AWS**

 **Gemini 1.5 Pro powers robot navigation**

 **Transform text into lifelike speech in seconds**

Enjoying these **FREE daily updates without SPAM or clutter?** then, Listen to it at our podcast and Support us by subscribing at

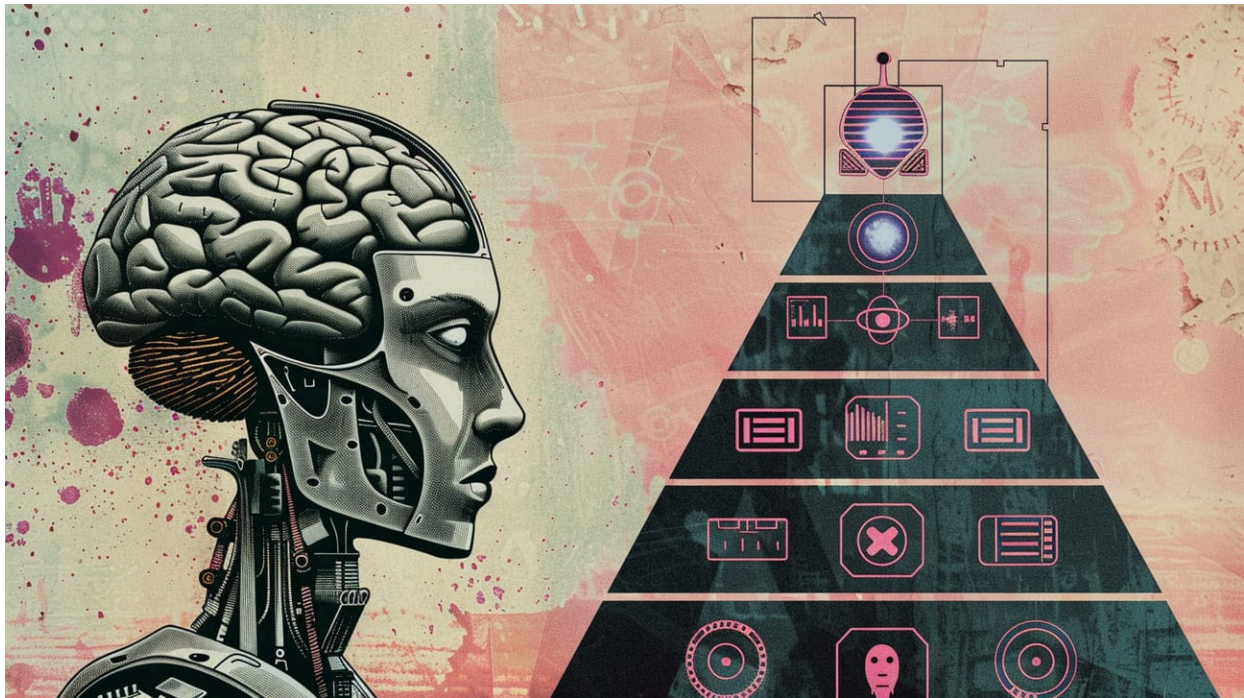
<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

To help us even more, Buy our "[Read Aloud Wonderland Bedtime Adventure Book: Diverse Tales for Dreamy Nights](#)" print Book for your kids, cousin, nephews or niece at

<https://www.barnesandnoble.com/w/wonderland-bedtime-adventures-etienne-noumen/1145739996?ean=9798331406462>.

OpenAI unveils five-level roadmap to AGI



- OpenAI has introduced a five-level scale to measure advancements towards Artificial General Intelligence (AGI) and aims to soon reach the "reasoner" stage, which is the second level.
- At an employee meeting, OpenAI revealed details about this new classification system and noted their proximity to achieving level 2, which involves AI capable of solving problems at a human level.
- The five-level framework culminates in systems that can outperform humans in most economically valuable tasks, with level 5 AI being able to perform the work of an entire organization.
- The classification system ranges from Level 1 (current conversational AI) to Level 5 (AI capable of running entire organizations).
- OpenAI believes its technology is currently at Level 1 but nearing Level 2, dubbed 'Reasoners.'
- The company reportedly demonstrated a GPT-4 research project showing human-like reasoning skills at the meeting, hinting at progress towards Level 2.
- Level 2 AI can perform basic problem-solving tasks on par with a PhD-level human without tools, with Level 3 rising to agents that can take action for users.

Source: <https://the-decoder.com/openai-unveils-five-level-ai-scale-aims-to-reach-level-2-soon/>

Tesla delays robotaxi event in blow to Musk's autonomy drive

- Tesla has delayed its robotaxi unveiling to October to give teams more time to build additional prototypes, according to unnamed sources.
- The event postponement, initially set for August 8, has led to a significant drop in Tesla's stock, while shares of competitors Uber and Lyft surged.
- Elon Musk has emphasized the robotaxi project over cheaper electric vehicles, despite the Full Self-Driving feature still requiring constant supervision and not making Teslas fully autonomous.

Source:

<https://www.scmp.com/tech/big-tech/article/3270171/tesla-delays-robotaxi-event-blow-musks-autonomy-drive>

Google's Gemini 1.5 Pro gets a body: DeepMind's office "helper" robot



A tall, wheeled "helper" robot is now roaming the halls of Google's California office, thanks to its AI model. Powered with Gemini 1.5 Pro's 1 million token context length, this robot assistant can use human instructions, video tours, and common sense reasoning to successfully navigate a space.

In a [new research paper](#) outlining the experiment, the researchers claim the robot proved to be up to 90% reliable at navigating, even with tricky commands such as “Where did I leave my coaster?” DeepMind’s algorithm, combined with the Gemini model, generates specific actions for the robot to take, such as turning, in response to commands and what it sees in front of it.

Why does it matter?

This work represents the next step in human-robot interaction. DeepMind says that in the future, users could simply record a tour of their environment with a smartphone so that their personal robot assistant can understand and navigate it.

Source: <https://x.com/GoogleDeepMind/status/1811401356827082796>

OpenAI’s new scale to track the progress of its LLMs toward AGI

OpenAI has created an internal scale to track its LLMs’ progress toward artificial general intelligence (AGI).

Chatbots, like ChatGPT, are at Level 1. OpenAI claims it is nearing Level 2, which is defined as a system that can solve basic problems at the level of a person with a PhD.

- Level 3 refers to AI agents capable of taking actions on a user’s behalf.
- Level 4 involves AI that can create new innovations.
- Level 5, the final step to achieving AGI, is AI that can perform the work of entire organizations of people.

This new grading scale is still under development.

Why does it matter?

OpenAI’s mission focuses on achieving AGI, making its definition crucial. A clear scale to evaluate progress could provide a more defined understanding of when AGI is reached, benefiting both OpenAI and its competitors.

Source:

<https://www.theverge.com/2024/7/11/24196746/heres-how-openai-will-determine-how-powerful-its-ai-systems-are>

Amazon announces a blitz of new AI updates for AWS

At the AWS New York Summit, AWS announced a wide range of capabilities for customers to tailor generative AI to their needs and realize the benefits of generative AI faster.

- Amazon Q Apps is now generally available. Users simply describe the application they want in a prompt and Amazon Q instantly generates it.
- With new features in Amazon Bedrock, AWS is making it easier to leverage your data, supercharge agents, and quickly, securely, and responsibly deploy generative AI into production.
- It also announced new partnerships with innovators like Scale AI to help you customize your applications quickly and easily.

Why does it matter?

AWS's lead in the cloud market has been shrinking, and it is relying on rapid AI product development to make its cloud services more appealing to customers.

Source:

<https://aws.amazon.com/blogs/machine-learning/empowering-everyone-with-genai-to-rapidly-build-customize-and-deploy-apps-securely-highlights-from-the-aws-new-york-summit>

Gemini 1.5 Pro powers robot navigation

Google DeepMind just published new research on robot navigation, leveraging the large context window of Gemini 1.5 Pro to enable robots to understand and navigate complex environments from human instructions.

- DeepMind's "Mobility VLA" combines Gemini's 1M token context with a map-like representation of spaces to create powerful navigation frameworks.
- Robots are first given a video tour of an environment, with key locations verbally highlighted – then constructing a graph of the space using video frames.
- In tests, robots responded to multimodal instructions, including map sketches, audio requests, and visual cues like a box of toys.
- The system also allows for natural language commands like "take me somewhere to draw things," with the robot then leading users to appropriate locations.

Equipping robots with multimodal capabilities and massive context windows is about to enable some wild use cases. Google's 'Project Astra' [demo](#) hinted at what the future holds for voice

assistants that can see, hear, and think – but embedding those functions within a robot takes things to another level.

Source: <https://x.com/GoogleDeepMind/status/1811401347477991932>

Groq claims the fastest hardware adoption in history

Groq announced that it has attracted 280,000 developers to its platform in just four months, a feat unprecedented in the hardware industry. Groq's innovative, memory-free approach to AI inference chips drives this rapid adoption.

Source:

<https://venturebeat.com/ai/groq-claims-fastest-hardware-adoption-in-history-at-vb-transform/>

SoftBank acquires UK AI chipmaker Graphcore

Graphcore, once considered a potential rival to market leader Nvidia, will now hire new staff in its UK offices. The firm will now be a subsidiary under SoftBank but will remain headquartered in Bristol.

Source: <https://www.bbc.com/news/articles/c3gd1n5kmy5o>

AMD to acquire Silo AI to expand enterprise AI solutions globally

Silo AI is the largest private AI lab in Europe, housing AI scientists and engineers with extensive experience developing tailored AI models. The move marks the latest in a series of acquisitions and corporate investments to support the AMD AI strategy.

Source: <https://www.silo.ai//blog/amd-to-acquire-silo-ai-to-expand-enterprise-ai-solutions-globally>

USA's COPIED Act would make removing digital watermarks illegal

The Act would direct the National Institute of Standards and Technology (NIST) to create standards and guidelines that help prove the origin of content and detect synthetic content, like through watermarking. It seeks to protect journalists and artists from having their work used by AI models without their consent.

Source:

<https://www.theverge.com/2024/7/11/24196769/copied-act-cantwell-blackburn-heinrich-ai-journalists-artists>

New startup helps creators track and license work used by AI

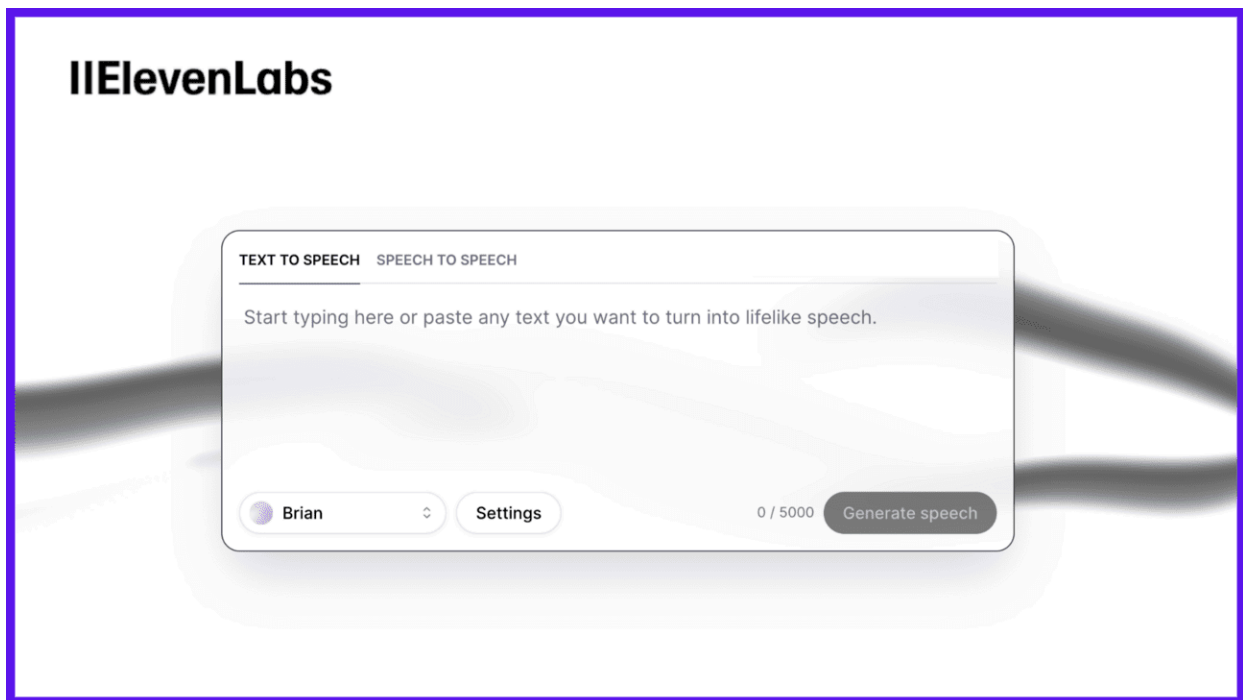
A new Los Angeles-based startup, SmarterLicense, is selling a tool that tracks when a creator's work is used on the internet for AI or other purposes.

Source:

<https://www.theinformation.com/articles/the-startup-helping-creators-track-and-license-work-used-by-ai>



Transform text into lifelike speech in seconds



ElevenLabs' AI-powered text-to-speech tool allows you to generate natural-sounding voiceovers easily with customizable voices and settings.

1. Sign up for a free ElevenLabs account [here](#) (10,000 free characters included).
2. Navigate to the "Speech" synthesis tool from your dashboard.
3. Enter your script in the text box and select a voice from the dropdown menu.
4. For advanced options, click "Advanced" to adjust the model, stability, and similarity settings.

5. Click "Generate speech" to create your audio file

Source:

<https://university.therundown.ai/c/daily-tutorials/transform-text-into-lifelike-speech-in-seconds-3bee4b0a-2b3c-4cea-989b-970e82342b1d>

Enjoying these FREE daily updates, Listen to it and Support us by subscribing at

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at

<https://readaloudforme.com>

To help us even more, Buy our "[Read Aloud Wonderland Bedtime Adventure Book: Diverse Tales for Dreamy Nights](#)"

for your kids, cousin, nephews or niece at

<https://www.barnesandnoble.com/w/wonderland-bedtime-adventures-etienne-noumen/1145739996?ean=9798331406462>.

A Daily chronicle of AI Innovations July 11th 2024:

 **Xiaomi unveils new factory that operates 24/7 without human labor**

 **OpenAI teams up with Los Alamos Lab to advance bioscience research.**



China dominates global gen AI adoption



Samsung reveals new AI wearables at 'Unpacked 2024'

Enjoying these daily updates, Listen to it and Support us by subscribing at

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

OpenAI partners with Los Alamos to advance 'bioscientific research'

- OpenAI is collaborating with Los Alamos National Laboratory to investigate how AI can be leveraged to counteract biological threats potentially created by non-experts using AI tools.
- The Los Alamos lab emphasized that prior research indicated ChatGPT-4 could provide information that might lead to creating biological threats, while OpenAI highlighted the partnership as a study on advancing bioscientific research safely.
- The focus of this partnership addresses concerns about AI being misused to develop bioweapons, with Los Alamos describing their work as a significant step towards understanding and mitigating risks associated with AI's potential to facilitate biological threats.

Source: <https://gizmodo.com/openai-partners-with-los-alamos-lab-to-save-us-from-ai-2000461202>

Xiaomi unveils new factory that operates 24/7 without human labor

- Xiaomi has launched a new autonomous smart factory in Beijing that can produce 10 million handsets annually and self-correct production issues using AI technology.
- The 860,000-square-foot facility includes 11 production lines and manufactures Xiaomi's latest smartphones, including the MIX Fold 4 and MIX Flip, at a high constant output rate.
- Operable 24/7 without human labor, the factory utilizes the Xiaomi Hyper Intelligent Manufacturing Platform to optimize processes and manage operations from material procurement to product delivery.

Source:

<https://www.techspot.com/news/103770-xiaomi-unveils-new-autonomous-smart-factory-operates-247.html>

OpenAI teams up with Los Alamos Lab to advance bioscience research

This first-of-its-kind partnership will assess how powerful models like GPT-4o can perform tasks in a physical lab setting using vision and voice by conducting biological safety evaluations. The evaluations will be conducted on standard laboratory experimental tasks, such as cell transformation, cell culture, and cell separation.

According to OpenAI, the upcoming partnership will extend its previous bioscience work into new dimensions, including the incorporation of 'wet lab techniques' and 'multiple modalities'.

The partnership will quantify and assess how these models can upskill professionals in performing real-world biological tasks.

Why does it matter?

It could demonstrate the real-world effectiveness of advanced multimodal AI models, particularly in sensitive areas like bioscience. It will also advance safe AI practices by assessing AI risks and setting new standards for safe AI-led innovations.

Source: <https://openai.com/index/openai-and-los-alamos-national-laboratory-work-together>

China dominates global gen AI adoption

According to a new survey of industries such as banking, insurance, healthcare, telecommunications, manufacturing, retail, and energy, China has emerged as a global leader in gen AI adoption.

Here are some noteworthy findings:

- Among the 1,600 decision-makers, 83% of Chinese respondents stated that they use gen AI, higher than 16 other countries and regions participating in the survey.
- A report by the United Nations WIPO highlighted that China had filed more than 38,000 patents between 2014 and 2023.
- China has also established a domestic gen AI industry with the help of tech giants like ByteDance and startups like Zhipu.

Why does it matter?

The USA is still the leader in successfully implementing gen AI. As China continues making developments in the field, it will be interesting to watch whether it will display enough potential to leave its rivals in the USA behind.

Source:

https://www.sas.com/en_us/news/press-releases/2024/july/genai-research-study-global.html

Samsung reveals new AI wearables at ‘Unpacked 2024’

Samsung unveiled advanced AI wearables at the Unpacked 2024 event, including the Samsung Galaxy Ring, AI-infused foldable smartphones, Galaxy Watch 7, and Galaxy Watch Ultra.

<https://youtu.be/IWCcBDL82oM?si=wHQ5zZKiu35BSanI>

Take a look at all of Samsung’s Unpacked 2024 in 12 minutes!

New Samsung Galaxy Ring features include:

- A seven-day battery life, along with 24/7 health monitoring.
- It also offers users a sleep score based on tracking metrics like movement, heart rate, and respiration.
- It also tracks the sleep cycles of users based on their skin temperature.

New features of foldable AI smartphones include:

- Sketch-to-image
- Note Assist
- Interpreter and Live Translate
- Built-in integration for the Google Gemini app
- AI-powered ProVisual Engine

The Galaxy Watch 7 and Galaxy Watch Ultra also boast features like AI-health monitoring, FDA-approved sleep apnea detection, diabetes tracking, and more, ushering Samsung into a new age of wearable revolution.

Why does it matter?

Samsung’s AI-infused gadgets are potential game-changers for personal health management. With features like FDA-approved sleep apnea detection, Samsung is blurring the line between consumer electronics and medical devices, causing speculations on whether it will leave established players like Oura, Apple, and Fitbit.

Source:

<https://news.samsung.com/global/galaxy-unpacked-2024-a-new-era-of-galaxy-ai-unfolds-at-the-louvre-in-paris>

AMD to buy SiloAI to bridge the gap with NVIDIA

AMD has agreed to pay \$665 million in cash to buy Silo in an attempt to accelerate its AI strategy and close the gap with its closest potential competition, NVIDIA Corp.

Source:

<https://www.bloomberg.com/news/articles/2024-07-10/amd-to-buy-european-ai-model-maker-silo-in-race-against-nvidia>

New AWS tool generates enterprise apps via prompts

The tool, named App Studio, lets you use a natural language prompt to build enterprise apps like inventory tracking systems or claims approval processes, eliminating the need for professional developers. It is currently available for a preview.

Source:

<https://aws.amazon.com/blogs/aws/build-custom-business-applications-without-cloud-expertise-using-aws-app-studio-preview>

Samsung Galaxy gets smarter with Google

Google has introduced new Gemini features and Wear OS 5 to Samsung devices. It has also extended its 'Circle to Search' feature's functionality, offering support for solutions to symbolic math equations, barcode scanning, and QR scanning.

Source:

<https://techcrunch.com/2024/07/10/google-brings-new-gemini-features-and-wearos-5-to-samsung-devices>

Writer drops enhancements to AI chat applications

Improvements include advanced graph-based retrieval-augmented generation (RAG) and AI transparency tools, available for users of 'Ask Writer' and AI Studio.

Source: <https://writer.com/blog/chat-app-rag-thought-process>

Vimeo launches AI content labels

Following the footsteps of TikTok, YouTube, and Meta, the AI video platform now urges creators to disclose when realistic content is created by AI. It is also working on developing automated AI labeling systems.

Source: <https://vimeo.com/blog/post/introducing-ai-content-labeling/>

Enjoying these daily updates, Listen to it and Support us by subscribing at

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>


A Daily chronicle of AI Innovations July 10th 2024:

 Microsoft and Apple abandon OpenAI board roles amid scrutiny

 The \$1.5B AI startup building a 'general purpose brain' for robots

 Odyssey is building a 'Hollywood-grade' visual AI

 Google's digital reconstruction of human brain with AI

 Odyssey is building a 'Hollywood-grade' visual AI

 Anthropic adds a playground to craft high-quality prompts

 Google's digital reconstruction of human brain with AI

 Anthropic's Claude Artifacts sharing goes live

US shuts down Russian AI bot farm

Enjoying these daily updates, Listen to it and Support us by subscribing at <https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

Microsoft and Apple abandon OpenAI board roles amid scrutiny

- Microsoft relinquished its observer seat on OpenAI's board less than eight months after obtaining the non-voting position, and Apple will no longer join the board as initially planned.
- Changes come amid increasing scrutiny from regulators, with UK and EU authorities investigating antitrust concerns over Microsoft's partnership with OpenAI, alongside other major tech AI deals.
- Despite leaving the board, Microsoft continues its partnership with OpenAI, backed by more than \$10 billion in investment, with its cloud services powering OpenAI's projects and integrations into Microsoft's products.
- Source:
<https://www.theverge.com/2024/7/10/24195528/microsoft-apple-openai-board-observer-seat-drop-regulator-scrutiny>

US shuts down Russian AI bot farm

- The Department of Justice announced the seizure of two domain names and over 900 social media accounts that were part of an AI-enhanced Russian bot farm aiming to spread disinformation about the Russia-Ukraine war.
- The bot farm, allegedly orchestrated by an RT employee, created numerous profiles to appear as American citizens, with the goal of amplifying Russian President Vladimir Putin's narrative surrounding the invasion of Ukraine.
- The operation involved the use of Meliorator software to generate and manage fake identities on X, which circumvented verification processes, and violated the Emergency Economic Powers Act according to the ongoing DOJ investigation.

Source:

<https://www.theverge.com/2024/7/9/24195228/doj-bot-farm-rt-russian-government-namecheap>

The \$1.5B AI startup building a 'general purpose brain' for robots

- Skild AI has raised \$300 million in a Series A funding round to develop a general-purpose AI brain designed to equip various types of robots, reaching a valuation of \$1.5 billion.
- This significant funding round saw participation from top venture capital firms such as Lightspeed Venture Partners, Softbank, alongside individual investors like Jeff Bezos.
- Skild AI aims to revolutionize the robotics industry with its versatile AI brain that can be integrated into any robot, enhancing its capabilities to perform multiple tasks in diverse environments, addressing the significant labor shortages in industries like healthcare and manufacturing.

Source:

<https://siliconangle.com/2024/07/09/skild-ai-raises-300m-build-general-purpose-ai-powered-brain-robot/>

Odyssey is building a 'Hollywood-grade' visual AI

Odyssey, a young AI startup, is pioneering Hollywood-grade visual AI that will allow for both generation and direction of beautiful scenery, characters, lighting, and motion.

It aims to give users full, fine-tuned control over every element in their scenes— all the way to the low-level materials, lighting, motion, and more. Instead of training one model that restricts users to a single input and a single, non-editable output, Odyssey is training four powerful generative models to enable its capabilities. Odyssey's creators claim the technology is what comes after text-to-video.

Why does it matter?

While we wait for the general release of OpenAI's Sora, Odyssey is paving a new way to create movies, TV shows, and video games. Instead of replacing humans with algorithms, it is placing a powerful enabler in the hands of professional storytellers.

Source: <https://x.com/olivercameron/status/1810335663197413406>

Anthropic adds a playground to craft high-quality prompts

Anthropic Console now offers a built-in prompt generator powered by Claude 3.5 Sonnet. You describe your task and Claude generates a high-quality prompt for you. You can also use Claude's new test case generation feature to generate input variables for your prompt and run the prompt to see Claude's response.

Moreover, with the new Evaluate feature you can do testing prompts against a range of real-world inputs directly in the Console instead of manually managing tests across spreadsheets or code. Anthropic has also added a feature to compare the outputs of two or more prompts side by side.

<https://youtu.be/KIGBsQqZcNA?si=7Ej39qsQBzyYIMTs>

Why does it matter?

Language models can improve significantly with small prompt changes. Normally, you'd figure this out yourself or hire a prompt engineer, but these features help make improvements quick and easier.

Source: <https://www.anthropic.com/news/evaluate-prompts>

Google's digital reconstruction of human brain with AI

Google researchers have completed the largest-ever AI-assisted digital reconstruction of human brain. They unveiled the most detailed map of the human brain yet of just 1 cubic millimeter of brain tissue (size of half a grain of rice) but at high resolution to show individual neurons and their connections.

Now, the team is working to map a mouse's brain because it looks exactly like a miniature version of a human brain. This may help solve mysteries about our minds that have eluded us since our beginnings.

Why does it matter?

This is a never-seen-before map of the entire human brain that could help us understand long-standing mysteries like where diseases come from to how we store memories. But the mapping takes billions of dollars and decades. AI might just have sped the process!

Source: <https://blog.google/technology/research/mouse-brain-research>

Microsoft ditches its observer seat on OpenAI's board; Apple to follow

Microsoft ditched the seat after Microsoft expressed confidence in the OpenAI's progress and direction. OpenAI stated after this change that there will be no more observers on the board, likely ruling out reports of Apple gaining an observer seat.

Source:

<https://techcrunch.com/2024/07/10/as-microsoft-leaves-its-observer-seat-openai-says-it-wont-have-any-more-observers>

LMSYS launched Math Arena and Instruction-Following (IF) Arena

Math and IF are two key domains testing models' logical skills and real-world tasks. Claude 3.5 Sonnet ranks #1 in Math Arena and joint #1 in IF with GPT-4o. While DeepSeek-coder is the #1 open model in math.

Source: <https://x.com/lmsysorg/status/1810773765447655604>

Aitomatic launches the first open-source LLM for semiconductor industry

SemiKong aims to revolutionize semiconductor processes and fabrication technology, giving potential for accelerated innovation and reduced costs. It outperforms generic LLMs like GPT and Llama3 on industry-specific tasks.

Source: <https://venturebeat.com/ai/aitomatics-semikong-uses-ai-to-reshape-chipmaking-processes>

Stable Assistant's capabilities expand with two new features

It includes Search & Replace, which gives you the ability to replace an object in an image with another one. And Stable Audio enables the creation of high-quality audio of up to three minutes.

Source: <https://stability.ai/news/stability-ai-releases-stable-assistant-features>

Etsy will now allow sale of AI-generated art

It will allow the sale of artwork derived from the seller's own original prompts or AI tools as long as the artist discloses their use of AI in the item's listing description. Etsy will not allow the sale of AI prompt bundles, which it sees as crossing a creative line.

Source: <https://mashable.com/article/etsy-ai-art-policy>

Anthropic's Claude Artifacts sharing goes live

Anthropic just announced a new upgrade to its recently launched 'Artifacts' feature, allowing users to publish, share, and remix creations – alongside the launch of new prompt engineering tools in Claude's developer Console.

- The 'Artifacts' feature was [introduced](#) alongside Claude 3.5 Sonnet in June, allowing users to view, edit, and build in a real-time side panel workspace.
- Published Artifacts can now be shared and remixed by other users, opening up new avenues for collaborative learning.
- Anthropic also launched new developer tools in Console, including advanced testing, side-by-side output comparisons, and prompt generation assistance.

Making Artifacts shareable is a small but mighty update – unlocking a new dimension of AI-assisted content creation that could revolutionize how we approach online education, knowledge sharing, and collaborative work. The ability to easily create and distribute AI-generated experiences opens up a world of possibilities.

Source: <https://x.com/rowancheung/status/1810720903052882308>

A Daily chronicle of AI Innovations July 09th 2024:

 **LivePortrait animates images from video with precision**

 **Microsoft's 'MInference' slashes LLM processing time by 90%**

 **Groq's LLM engine surpasses Nvidia GPU processing**

OpenAI and Thrive create AI health coach

San Francisco office vacancies hit record high despite AI boom

Japan Ministry introduces first AI policy

Enjoying these daily updates, Listen to it and Support us by subscribing at <https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

LivePortrait animates images from video with precision

LivePortrait is a new method for animating still portraits using video. Instead of using expensive diffusion models, LivePortrait builds on an efficient "implicit keypoint" approach. This allows it to generate high-quality animations quickly and with precise control.

The key innovations in LivePortrait are:

1. Scaling up the training data to 69 million frames, using a mix of video and images, to improve generalization.
2. Designing new motion transformation and optimization techniques to get better facial expressions and details like eye movements.
3. Adding new "stitching" and "retargeting" modules that allow the user to precisely control aspects of the animation, like the eyes and lips.
4. This allows the method to animate portraits across diverse realistic and artistic styles while maintaining high computational efficiency.
5. LivePortrait can generate 512x512 portrait animations in just 12.8ms on an RTX 4090 GPU.

Why does it matter?

The advancements in generalization ability, quality, and controllability of LivePortrait could open up new possibilities, such as personalized avatar animation, virtual try-on, and augmented reality experiences on various devices.

Source: <https://arxiv.org/pdf/2407.03168>



San Francisco office vacancies hit record high despite AI boom

More than one-third of office spaces in San Francisco are empty despite an ongoing AI boom, with the vacancy rate hitting a historic 34.5% in the second quarter of this year, up from 33.9% in the first quarter. The severe office vacancy issue in San Francisco has led to landlords reducing rent prices to their lowest levels in almost a decade, with the average rent dropping to \$68.27 per square foot in the second quarter, a significant decrease from pre-pandemic rates. Despite efforts to attract back tech workers and the presence of AI firms, San Francisco's commercial real estate market remains in a critical state, with more than 53 companies moving their headquarters out of the city since 2020.

Source:

<https://nypost.com/2024/07/09/business/san-francisco-office-vacancies-hit-all-time-high-despite-silicon-valley-ai-boom/>



Microsoft's 'MInference' slashes LLM processing time by 90%

Microsoft has unveiled a new method called MInference that can reduce LLM processing time by up to 90% for inputs of one million tokens (equivalent to about 700 pages of text) while maintaining accuracy. MInference is designed to accelerate the "pre-filling" stage of LLM processing, which typically becomes a bottleneck when dealing with long text inputs.

Microsoft has released an interactive demo of MInference on the Hugging Face AI platform, allowing developers and researchers to test the technology directly in their web browsers. This hands-on approach aims to get the broader AI community involved in validating and refining the technology.

Why does it matter?

By making lengthy text processing faster and more efficient, MInference could enable wider adoption of LLMs across various domains. It could also reduce computational costs and energy usage, putting Microsoft at the forefront among tech companies and improving LLM efficiency.

Source:

<https://www.microsoft.com/en-us/research/project/minference-million-tokens-prompt-inference-for-long-context-llms/overview/>



Groq's LLM engine surpasses Nvidia GPU processing

Groq, a company that promises faster and more efficient AI processing, has unveiled a lightning-fast LLM engine. Their new LLM engine can handle queries at over 1,250 tokens per second, which is

much faster than what GPU chips from companies like Nvidia can do. This allows Groq's engine to provide near-instant responses to user queries and tasks.

<https://vimeo.com/980784649>

Groq's LLM engine has gained massive adoption, with its developer base rocketing past 280,000 in just 4 months. The company offers the engine for free, allowing developers to easily swap apps built on OpenAI's models to run on Groq's more efficient platform. Groq claims its technology uses about a third of the power of a GPU, making it a more energy-efficient option.

Why does it matter?

Groq's lightning-fast LLM engine allows for near-instantaneous responses, enabling new use cases like on-the-fly generation and editing. As large companies look to integrate generative AI into their enterprise apps, this could transform how AI models are deployed and used.

Source: <https://venturebeat.com/ai/groq-releases-blazing-fast-llm-engine-passes-270000-user-mark>

Japan's Defense Ministry introduces basic policy on using AI

This comes as the Japanese Self-Defense Forces grapple with challenges such as manpower shortages and the need to harness new technologies. The ministry believes AI has the potential to overcome these challenges in the face of Japan's declining population.

Source: <https://www.japantimes.co.jp/news/2024/07/02/japan/sdf-cybersecurity/>

Thrive AI Health democratizes access to expert-level health coaching

Thrive AI Health, a new company, funded by OpenAI and Thrive Global, uses AI to provide personalized health coaching. The AI assistant can leverage an individual's data to provide recommendations on sleep, diet, exercise, stress management, and social connections.

Source: <https://time.com/6994739/ai-behavior-change-health-care>

Qualcomm and Microsoft rely on AI wave to revive the PC market

Qualcomm and Microsoft are embarking on a marketing blitz to promote a new generation of "AI PCs." The goal is to revive the declining PC market. This strategy only applies to a small share of PCs sold this year, as major software vendors haven't agreed to the AI PC trend.

Source:

<https://www.bloomberg.com/news/articles/2024-07-08/qualcomm-microsoft-lean-on-ai-hype-to-spurr-pc-market-revival>

Poe's Previews let you see and interact with web apps directly within chats

This feature works especially well with advanced AI models like Claude 3.5 Sonnet, GPT-4o, and Gemini 1.5 Pro. Previews enable users to create custom interactive experiences like games, animations, and data visualizations without needing programming knowledge.

Source: https://x.com/poe_platform/status/1810335290281922984

Real-time AI video generation less than a year away: Luma Labs chief scientist

Luma's recently released video model, Dream Machine, was trained on enormous video data, equivalent to hundreds of trillions of words. According to Luma's chief scientist, Jiaming Song, this allows Dream Machine to reason about the world in new ways. He predicts realistic AI-generated videos will be possible within a year.

Source: <https://a16z.com/podcast/beyond-language-inside-a-hundred-trillion-token-video-model>

OpenAI and Thrive create AI health coach

The OpenAI Startup Fund and Thrive Global just announced Thrive AI Health, a new venture developing a hyper-personalized, multimodal AI-powered health coach to help users drive personal behavior change.

The AI coach will focus on five key areas: sleep, nutrition, fitness, stress management, and social connection.

Thrive AI Health will be trained on scientific research, biometric data, and individual preferences to offer tailored user recommendations.

DeCarlos Love steps in as Thrive AI Health's CEO, who formerly worked on AI, health, and fitness experiences at Google as a product leader.

OpenAI CEO Sam Altman and Thrive Global founder Ariana Huffington published an article in TIME detailing AI's potential to improve both health and lifespans.

With chronic disease and healthcare costs on the rise, AI-driven personalized coaching could be a game-changer – giving anyone the ability to leverage their data for health gains. Plus, Altman's network of companies and partners lends itself perfectly to crafting a major AI health powerhouse.

Source:

<https://www.prnewswire.com/news-releases/openai-startup-fund--arianna-huffingtons-thrive-global-create-new-company-thrive-ai-health-to-launch-hyper-personalized-ai-health-coach-302190536.html>

Japan Ministry introduces first AI policy

Japan's Defense Ministry just released its inaugural basic policy on the use of artificial intelligence in military applications, aiming to tackle recruitment challenges and keep pace with global powers in defense technology.

The policy outlines seven priority areas for AI deployment, including target detection, intelligence analysis, and unmanned systems.

Japan sees AI as a potential solution to its rapidly aging and shrinking population, which is currently impacting military recruitment.

The strategy also emphasizes human control over AI systems, ruling out fully autonomous lethal weapons.

Japan's Defense Ministry highlighted the U.S. and China's military AI use as part of the 'urgent need' for the country to utilize the tech to increase efficiency.

Whether the world is ready or not, the military and AI are about to intertwine. By completely ruling out autonomous lethal weapons, Japan is setting a potential model for more responsible use of the tech, which could influence how other powers approach the AI military arms race in the future.

Source: <https://www.japantimes.co.jp/news/2024/07/02/japan/sdf-cybersecurity>

What else is happening in AI on July 09th 2024

Poe launched 'Previews', a new feature allowing users to generate and interact with web apps directly within chats, leveraging LLMs like Claude 3.5 Sonnet for enhanced coding capabilities.

Source: https://x.com/poe_platform/status/1810335290281922984

Luma Labs chief scientist Jiaming Song said in an interview that real-time AI video generation is less than a year away, also showing evidence that its Dream Machine model can reason and predict world models in some capacity. Source: <https://x.com/AnjneyMidha/status/1808783852321583326>

Magnific AI introduced a new Photoshop plugin, allowing users to leverage the AI upscaling and enhancing tool directly in Adobe's editing platform. Source:

<https://x.com/javilopen/status/1810345184754069734>

Nvidia launched a new competition to create an open-source code dataset for training LLMs on hardware design, aiming to eventually automate the development of future GPUs. Source:

<https://nvlabs.github.io/LLM4HWDesign>

Taiwan Semiconductor Manufacturing Co. saw its valuation briefly surpass \$1T, coming on the heels of Morgan Stanley increasing its price targets for the AI chipmaker. Source:

<https://finance.yahoo.com/news/tsmc-shares-soar-record-expectations-041140534.html>

AI startup Hebbia secured \$130M in funding for its complex data analysis software, boosting the company's valuation to around \$700M. Source:

<https://www.bloomberg.com/news/articles/2024-07-08/hebbia-raises-130-million-for-ai-that-helps-firms-answer-complex-questions>

A new study testing ChatGPT's coding abilities found major limitations in the model's abilities, though the research has been criticized for its use of GPT-3.5 instead of newer, more capable models. Source: <https://ieeexplore.ieee.org/document/10507163>

Enjoying these daily updates, Listen to it and Support us by subscribing at

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

A Daily chronicle of AI Innovations July 08th 2024:

 **SenseTime released SenseNova 5.5 at the 2024 World Artificial Intelligence Conference**

 **Cloudflare launched a one-click feature to block all AI bots**

 **Waymo's Robotaxi gets busted by the cops**

 **OpenAI's secret AI details stolen in 2023 hack**

 **Fears of AI bubble intensify after new report**

 **Chinese AI firms flex muscles at WAIC**

Enjoying these daily updates, Listen to it and Support us by subscribing at

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

 **SenseTime released SenseNova 5.5 at the 2024 World Artificial Intelligence Conference**

Leading Chinese AI company SenseTime released an upgrade to its SenseNova large model. The new 5.5 version boasts China's first real-time multimodal model on par with GPT-4o, a cheaper IoT-ready edge model, and a rapidly growing customer base.

SenseNova 5.5 packs a 30% performance boost, matching GPT-4o in interactivity and key metrics. The suite includes SenseNova 5o for seamless human-like interaction and SenseChat Lite-5.5 for lightning-fast inference on edge devices.

With industry-specific models for finance, agriculture, and tourism, SenseTime claims significant efficiency improvements in these sectors, such as 5x improvement in agricultural analysis and 8x in travel planning efficiency.

Why does it matter?

With the launch of "Project \$0 Go," which offers free tokens and API migration consulting to enterprise users, combined with the advanced features of SenseNova 5.5, SenseTime will provide accessible and powerful AI solutions for businesses of all sizes.

Source: <https://www.sensetime.com/en/news-detail/51168278>

Cloudflare launched a one-click feature to block all AI bots

Cloudflare just dropped a single-click tool to block all AI scrapers and crawlers. With demand for training data soaring and sneaky bots rising, this new feature helps users protect their precious content without hassle.

Bytespider, Amazonbot, ClaudeBot, and GPTBot are the most active AI crawlers on Cloudflare's network. Some bots spoof user agents to appear as real browsers, but Cloudflare's ML models still identify them. It uses global network signals to detect and block new scraping tools in real time. Customers can report misbehaving AI bots to Cloudflare for investigation.

Why does it matter?

While AI bots hit 39% of top sites in June, less than 3% fought back. With Cloudflare's new feature, websites can protect users' precious data and gain more control.

Source:

<https://blog.cloudflare.com/declaring-your-aindependence-block-ai-bots-scrapers-and-crawlers-with-a-single-click>

Waymo's Robotaxi gets busted by the cops

A self-driving Waymo vehicle was pulled over by a police officer in Phoenix after running a red light. The vehicle briefly entered an oncoming traffic lane before entering a parking lot. Bodycam footage shows the officer finding no one in the self-driving Jaguar I-Pace. Dispatch records state the vehicle "freaked out," and the officer couldn't issue a citation to the computer.

<https://youtu.be/7W-VneUv8Gk?si=Zv6hEUPRHbJ126y2>

Waymo initially refused to discuss the incident but later claimed inconsistent construction signage caused the vehicle to enter the wrong lane for 30 seconds. Federal regulators are investigating the safety of Waymo's self-driving software.

Why does it matter?

The incident shows the complexity of deploying self-driving cars. As these vehicles become more common on our streets, companies must ensure these vehicles can safely and reliably handle real-world situations.

Source:

<https://techcrunch.com/2024/07/06/waymo-robotaxi-pulled-over-by-phoenix-police-after-driving-into-the-wrong-lane/>



OpenAI's secret AI details stolen in 2023 hack

A new report from the New York Times just revealed that a hacker breached OpenAI's internal messaging systems last year, stealing sensitive details about the company's tech – with the event going unreported to the public or authorities.

The breach occurred in early 2023, with the hacker accessing an online forum where employees discussed OpenAI's latest tech advances.

While core AI systems and customer data weren't compromised, internal discussions about AI designs were exposed.

OpenAI informed employees and the board in April 2023, but did not disclose the incident publicly or to law enforcement.

Former researcher Leopold Aschenbrenner (later fired for allegedly leaking sensitive info) criticized OpenAI's security in a memo following the hack.

OpenAI has since established a Safety and Security Committee, including the addition of former NSA head Paul Nakasone, to address future risks.

Is OpenAI's secret sauce out in the wild? As other players continue to even the playing field in the AI race, it's fair to wonder if leaks and hacks have played a role in the development. The report also adds new intrigue to Aschenbrenner's firing – who has been adamant that his release was politically motivated.

Source: <https://www.nytimes.com/2024/07/04/technology/openai-hack.html>



Chinese AI firms flex muscles at WAIC

The World Artificial Intelligence Conference (WAIC) took place this weekend in Shanghai, with Chinese companies showcasing significant advances in LLMs, robotics, and other AI-infused products despite U.S. sanctions on advanced chips.

SenseTime unveiled SenseNova 5.5 at the event, claiming the model outperforms GPT-4o in 5 out of 8 key metrics.

The company also released SenseNova 5o, a real-time multimodal model capable of processing audio, text, image, and video.

Alibaba's cloud unit reported its open-source Tongyi Qianwen models doubled downloads to over 20M in just two months.

iFlytek introduced SparkDesk V4.0, touting advances over GPT-4 Turbo in multiple domains.

Moore Threads showcased KUAE, an AI data center solution with GPUs performing at 60% of NVIDIA's restricted A100.

If China's AI firms are being slowed down by U.S. restrictions, they certainly aren't showing it. The models and tech continue to rival the leaders in the market – and while sanctions may have created hurdles, they may have also spurred Chinese innovation with workarounds to stay competitive.

Source:

<https://www.scmp.com/tech/big-tech/article/3269387/chinas-ai-competition-deepens-sensetime-alibaba-claim-progress-ai-show>

Fears of AI bubble intensify after new report

The AI industry needs to generate \$600 billion annually to cover the extensive costs of AI infrastructure, according to a new Sequoia report, highlighting a significant financial gap despite heavy investments from major tech companies. Sequoia Capital analyst David Cahn suggests that the current revenue projections for AI companies fall short, raising concerns over a potential financial bubble within the AI sector. The discrepancy between AI infrastructure expenditure and revenue, coupled with speculative investments, suggests that the AI industry faces significant challenges in achieving sustainable profit, potentially leading to economic instability. Source: <https://www.tomshardware.com/tech-industry/artificial-intelligence/ai-industry-needs-to-earn-dollar-600-billion-per-year-to-pay-for-massive-hardware-spend-fears-of-an-ai-bubble-intensify-in-wake-of-sequoia-report>

Google researchers' paper warns that Gen AI ruins the internet

Most generative AI users use the tech to post fake or doctored content online; this AI-generated content influences public opinion, enables scams, and generates profit. The paper doesn't mention Google's issues and mistakes with AI, despite Google pushing the technology to its vast user base.

Source: <https://futurism.com/the-byte/google-researchers-paper-ai-internet>

Stability AI announced a new free license for its AI models

Commercial use of the AI models is allowed for small businesses and creators with under \$1M in revenue at no cost. Non-commercial use remains free for researchers, open-source devs, students, teachers, hobbyists, etc. Stability AI also pledged to improve SD3 Medium and share learnings quickly to benefit all.

Source: <https://stability.ai/news/license-update>

Google DeepMind developed a new AI training technique called JEST

JEST ((joint example selection) trains on batches of data and uses a small AI model to grade data quality and select the best batches for training a larger model. It achieves 13x faster training speed and 10x better power efficiency than other methods.

The technique leverages two AI models — a pre-trained reference model and a ‘learner’ model that is being trained to identify the most valuable data examples.

JEST intelligently selects the most instructive batches of data, making AI training up to 13x faster and 10x more efficient than current state-of-the-art methods.

In benchmark tests, JEST achieved top-tier performance while only using 10% of the training data required by previous leading models.

The method enables ‘data quality bootstrapping’ — using small, curated datasets to guide learning on larger unstructured ones.

Source: <https://arxiv.org/abs/2406.17711>

Apple Intelligence is expected to launch in iOS 18.4 in spring 2025

This will bring major improvements to Siri. New AI features may be released incrementally in iOS point updates. iOS 18 betas later this year will provide more details on the AI features. Source: <https://www.theverge.com/2024/7/7/24193619/apple-intelligence-better-siri-ios-18-4-spring-public-launch>

A new WhatsApp beta version for Android lets you send photos to Meta AI

Users can ask Meta AI questions about objects or context in their photos. Meta AI will also offer photo editing capabilities within the WhatsApp chat interface. Users will have control over their pictures and can delete them anytime.

Source: <https://wabetainfo.com/whatsapp-beta-for-android-2-24-14-20-whats-new/>

Google claims new AI training tech is 13 times faster and 10 times more power efficient —


DeepMind's new JEST optimizes training data for impressive gains.

Source:


<https://www.tomshardware.com/tech-industry/artificial-intelligence/google-claims-new-ai-training-tech-is-13-times-faster-and-10-times-more-power-efficient-deepminds-new-jest-optimizes-training-data-for-massive-gains>

New AI Job Opportunities on July 08th 2024

 xAI - Product Designer: <https://jobs.therundown.ai/jobs/60681923-product-designer>

 Weights & Biases - Programmer Writer, Documentation: <https://jobs.therundown.ai/jobs/66567362-programmer-writer-documentation-remote>

 DeepL - Enterprise Customer Success Manager: <https://jobs.therundown.ai/jobs/66103798-enterprise-customer-success-manager-%7C-dach>

 Dataiku - Senior Infrastructure Engineer: <https://jobs.therundown.ai/jobs/66413411-senior-infrastructure-engineer-paris>

Source: <https://jobs.therundown.ai/>

Enjoying these daily updates, Listen to it and Support us by subscribing at

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Chronicle Website at <https://readaloudforme.com>

A Daily chronicle of AI Innovations July 05th 2024:

 **AI recreates images from brain activity**

 **Apple rumored to launch AI-powered home device**

 **Google considered blocking Safari users from accessing its new AI features**

 **Researchers develop virus that leverages ChatGPT to spread through human-like emails**

 **New AI system decodes brain activity with near perfection**

 **ElevenLabs has exciting AI voice updates**

 **A French AI startup launches 'real-time' AI voice assistant**

Enjoying these AI updates, subscribe and listen to our podcast at:

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Innovations web site and Get a copy of our AI Unraveled Book at

https://readaloudforme.com/index_daily_ai_chronicles.html

 **New AI system decodes brain activity with near perfection**



Top Row: What the monkey saw

Bottom Row: The images the AI system reconstructed based on brain activity

Researchers have developed an AI system that can create remarkably accurate reconstructions of what someone is looking at based on recordings of their brain activity.

In previous studies, the team recorded brain activities using a functional MRI (fMRI) scanner and implanted electrode arrays. Now, they reanalyzed the data from these studies using an improved AI system that can learn which parts of the brain it should pay the most attention to.

As a result, some of the reconstructed images were remarkably close to the images the macaque monkey (in the study) saw.

Why does it matter?

This is probably the closest, most accurate mind-reading accomplished with AI yet. It proves that reconstructed images are greatly improved when the AI learns which parts of the brain to pay attention to. Ultimately, it can create better brain implants for restoring vision.

Source:

<https://www.newscientist.com/article/2438107-mind-reading-ai-recreates-what-youre-looking-at-with-amazing-accuracy>

⚡ ElevenLabs has exciting AI voice updates

ElevenLabs has partnered with estates of iconic Hollywood stars to bring their voices to the Reader App. Judy Garland, James Dean, Burt Reynolds, and Sir Laurence Olivier are now part of the library of voices on the Reader App.

<https://youtu.be/Zo5-rhYOINK?si=o986JZDegDKGp-be>

It has also introduced Voice Isolater. This tool removes unwanted background noise and extracts crystal-clear dialogue from any audio to make your next podcast, interview, or film sound like it was

recorded in the studio. It will be available via API in the coming weeks.

Why does it matter?

ElevenLabs is shipping fast! It appears to be setting a standard in the AI voice technology industry by consistently introducing new AI capabilities with its technology and addressing various needs in the audio industry.

Source: <https://elevenlabs.io/blog/iconic-voices>

A French AI startup launches 'real-time' AI voice assistant

A French AI startup, Kyutai, has launched a new 'real-time' AI voice assistant named Moshi. It is capable of listening and speaking simultaneously and in 70 different emotions and speaking styles, ranging from whispers to accented speech.

Kyutai claims Moshi is the first real-time voice AI assistant, with a latency of 160ms. You can try it via Hugging Face. It will be open-sourced for research in coming weeks.

Why does it matter?

Yet another impressive competitor that challenges OpenAI's perceived dominance in AI. (Moshi could outpace OpenAI's delayed voice offering.) Such advancements push competitors to improve their offerings, raising the bar for the entire industry.

Source: <https://www.youtube.com/live/hm2IJSKcYvo?si=EtirSsXktlwakmn5>

Meta's multi-token prediction models are now open for research

In April, Meta proposed a new approach for training LLMs to forecast multiple future words simultaneously vs. the traditional method to predict just the next word in a sequence. Meta has now released pre-trained models that leverage this approach.

Source:

<https://venturebeat.com/ai/meta-drops-ai-bombshell-multi-token-prediction-models-now-open-for-research/>

Apple to announce AI partnership with Google at iPhone 16 event

Apple has been meeting with several companies to partner with in the AI space, including Google. Reportedly, Apple will announce the addition of Google Gemini on iPhones at its annual event in September.

Source: <https://mashable.com/article/apple-google-ai-partnership-report>

Google simplifies the process for advertisers to disclose if political ads use AI

In an update to its Political content policy, Google requires advertisers to disclose election ads containing synthetic or digitally altered content. It will automatically include an in-ad disclosure for specific formats.

Source:

<https://searchengineland.com/google-disclosure-rules-synthetic-content-political-ads-443868>

WhatsApp is developing a personalized AI avatar generator

It appears to be working on a new Gen AI feature that will allow users to make personalized avatars of themselves for use in any imagined setting. It will generate images using user-supplied photos, text prompts, and Meta's Llama model.

Source:

<https://www.theverge.com/2024/7/4/24192112/whatsapp-ai-avatar-image-generator-imagine-meta-llama>

Meta ordered to stop training its AI on Brazilian personal data

Brazil's National Data Protection Authority (ANPD) has decided to suspend with immediate effect the validity of Meta's new privacy policy (updated in May) for using personal data to train generative AI systems in the country. Meta will face daily fines if it fails to comply.

Source:

<https://www.reuters.com/technology/artificial-intelligence/brazil-authority-suspends-metas-ai-privacy-policy-seeks-adjustment-2024-07-02>

Apple rumored to launch AI-powered home device

Apple is rumored to be developing a new home device that merges the functionalities of the HomePod and Apple TV, supported by "Apple Intelligence" and potentially featuring the upcoming A18 chip, according to recent code discoveries.

Identified as "HomeAccessory17,1," this device is expected to include a speaker and LCD screen, positioning it to compete with Amazon's Echo Show and Google's Nest series.

The smart device is anticipated to serve as a smart home hub, allowing users to control HomeKit devices, and it may integrate advanced AI features announced for iOS 18, iPadOS 18, and macOS Sequoia, including capabilities powered by OpenAI's GPT-4 to enhance Siri's responses.

Source: <https://bgr.com/tech/apple-mysterious-ai-powered-home-device/>

Google considered blocking Safari users from accessing its new AI features

Google considered limiting access to its new AI Overviews feature on Safari but ultimately decided not to follow through with the plan, according to a report by The Information.

The ongoing Justice Department investigation into Google's dominance in search highlights the company's arrangement with Apple, where Google pays around \$20 billion annually to be the default search engine on iPhones.

Google has been trying to reduce its dependency on Safari by encouraging iPhone users to switch to its own apps, but the company has faced challenges due to Safari's pre-installed presence on Apple devices.

Source: <https://9to5mac.com/2024/07/05/google-search-iphone-safari-ai-features/>

Researchers develop virus that leverages ChatGPT to spread through human-like emails

Researchers from ETH Zurich and Ohio State University created a virus named "synthetic cancer" that leverages ChatGPT to spread via AI-generated emails.

This virus can modify its code to evade antivirus software and uses Outlook to craft contextually relevant, seemingly innocuous email attachments.

The researchers stress the cybersecurity risks posed by Language Learning Models (LLMs), highlighting the need for further research into protective measures against intelligent malware.

Source:

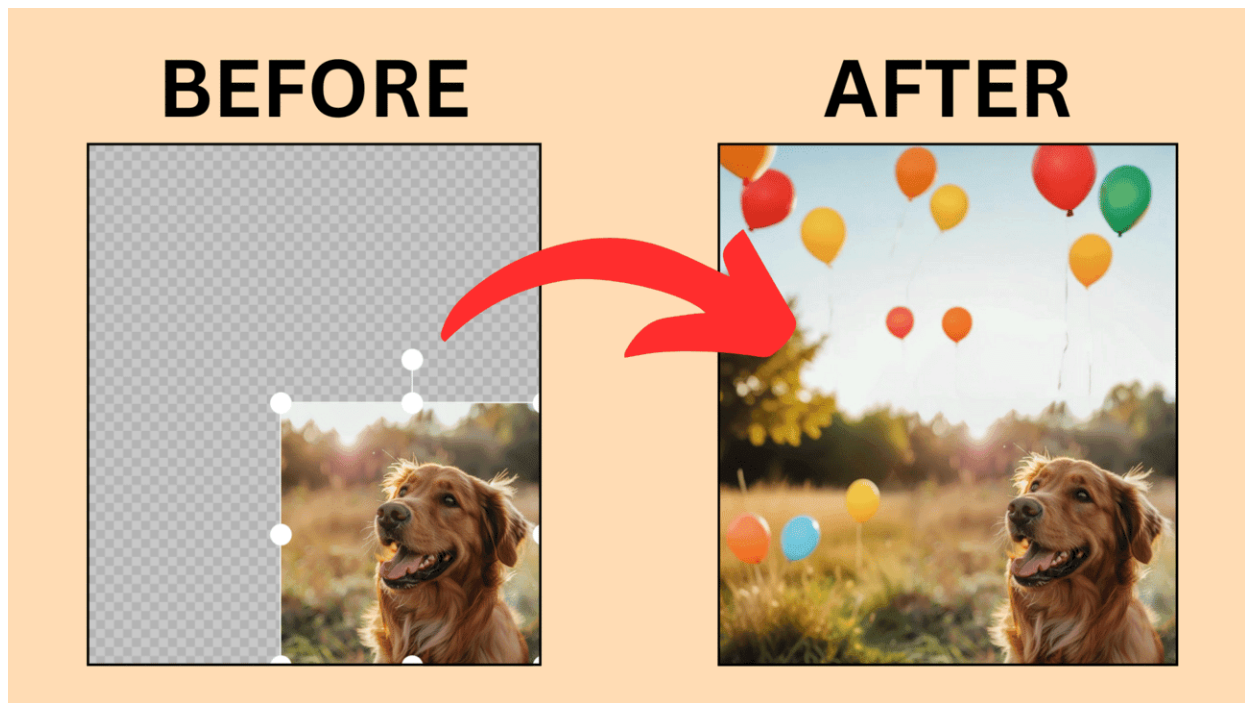
<https://www.newsbytesapp.com/news/science/virus-leverages-chatgpt-to-spread-itself-by-sending-human-like-emails/story>

You can now get AI Judy Garland or James Dean to read you the news.

Source:

<https://www.engadget.com/you-can-now-get-ai-judy-garland-or-james-dean-to-read-you-the-news-160023595.html>

Stretch creativity with AI image expansion



Freepik has a powerful new feature called 'Expand' that allows you to expand your images beyond their original boundaries, filling in details with AI.

Head over to the Freepik Pikaso website and look for the "Expand" feature.

Upload your image by clicking "Upload" or using drag-and-drop.

Choose your desired aspect ratio from the options on the left sidebar and add a prompt describing what you want in the expanded areas.

Click "Expand", browse the AI-generated results, and select your favorite 🎨

Source:

<https://university.therundown.ai/c/daily-tutorials/stretch-your-creativity-with-ai-image-expansion-56b69128-ef5a-445a-ae55-9bc31c343cdf>

Enjoying these AI updates, subscribe and listen to our podcast at:

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169> Visit our Daily AI Innovations web site and Get a copy of our AI Unraveled Book at https://readaloudforme.com/index_daily_ai_chronicles.html

A Daily chronicle of AI Innovations July 04th 2024:

 **OpenAI secrets stolen by hacker**

 **French AI lab Kyutai unveils conversational AI assistant Moshi**

 **China leads the world in generative AI patents**

 **OpenAI's ChatGPT Mac app was storing conversations in plain text**

 **Salesforce's small model breakthrough**

 **Perplexity gets major research upgrade**

Enjoying these AI updates, subscribe and listen to our podcast at:

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Innovations web site and Get a copy of our AI Unraveled Book at

https://readaloudforme.com/index_daily_ai_chronicles.html

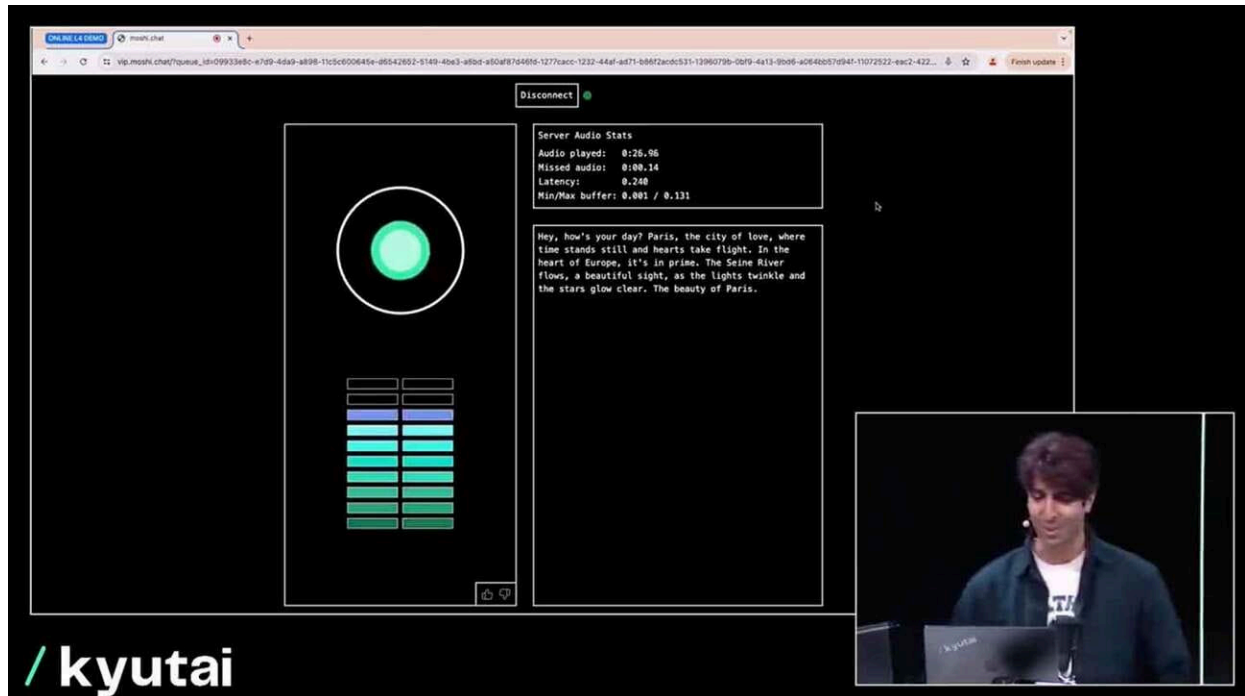
 **OpenAI secrets stolen by hacker**

- A hacker accessed OpenAI's internal messaging systems early last year and stole design details about the company's artificial intelligence technologies.
- The attacker extracted information from employee discussions in an online forum but did not breach the systems where OpenAI creates and stores its AI tech.

- OpenAI executives disclosed the breach to their staff in April 2023 but did not make it public, as no sensitive customer or partner information was compromised.

Source: <https://www.nytimes.com/2024/07/04/technology/openai-hack.html>

French AI lab Kyutai unveils conversational AI assistant Moshi



- French AI lab Kyutai introduced Moshi, a conversational AI assistant capable of natural interaction, at an event in Paris and plans to release it as open-source technology.
- Kyutai stated that Moshi is the first AI assistant with public access enabling real-time dialogue, differentiating it from OpenAI's GPT-4o, which has similar capabilities but is not yet available.
- Developed in six months by a small team, Moshi's unique "Audio Language Model" architecture allows it to process and predict speech directly from audio data, achieving low latency and impressive language skills despite its relatively small model size.

Source:

<https://the-decoder.com/french-ai-lab-kyutai-unveils-conversational-ai-assistant-moshi-plans-open-source-release/>

China leads the world in generative AI patents

- China has submitted significantly more patents related to generative artificial intelligence than any other nation, with the United States coming in a distant second, according to the World Intellectual Property Organization.
- In the decade leading up to 2023, over 38,200 generative AI inventions originated in China, compared to almost 6,300 from the United States, demonstrating China's consistent lead in this technology.
- Generative AI, using tools like ChatGPT and Google Gemini, has seen rapid growth and industry adoption, with concerns about its impact on jobs and fairness of content usage, noted the U.N. intellectual property agency.

Source: <https://fortune.com/asia/2024/07/04/china-generative-ai-patents-un-wipo-us-second/>



OpenAI's ChatGPT Mac app was storing conversations in plain text

- OpenAI launched the first official ChatGPT app for macOS, raising privacy concerns because conversations were initially stored in plain text.
- Developer Pedro Vieito revealed that the app did not use macOS sandboxing, making sensitive user data easily accessible to other apps or malware.
- OpenAI released an update after the concerns were publicized, which now encrypts chats on the Mac, urging users to update their app to the latest version.

Source: <https://9to5mac.com/2024/07/03/chatgpt-macos-conversations-plain-text/>



Salesforce's small model breakthrough

Rank	Overall Accuracy	Model	Abstract Syntax Tree (AST) Evaluation				Evaluation by Executing APIs				Relevance Detection
			Simple	Multiple	Parallel	Parallel Multiple	Simple	Multiple	Parallel	Parallel Multiple	
1	88	GPT-4-0125-Preview (Prompt)	88.36	95	90.5	91	99.41	94	84	75	70.42
2	87.65	Claude-3-Opus-0229 (Prompt)	86.73	94	86	89	97.65	92	80	75	80.42
3	86.35	Gemini-1.5-Pro-0514 (FC)	80.18	92	91	88.5	91.76	88	76	77.5	89.58
4	85.88	Gemini-1.5-Pro-0409 (FC)	80	92.5	90	88	90	90	74	77.5	88.75
5	85.88	GPT-4-1106-Preview (FC)	84	91.5	92	87	89.41	92	78	67.5	80.42
6	85.65	xLAM-7B (FC)	80.55	96	90	87.5	90.59	90	86	77.5	80.42
7	85.59	GPT-4-turbo-20240409 (Prompt)	86.55	95	90	88.5	97.65	94	80	72.5	62.5
8	84.71	Gorilla-OpenFunctions-v2 (FC)	88	95	87.5	87	94.71	94	70	67.5	61.25
9	84.65	GPT-4-0125-Preview (FC)	80.18	93	90.5	85	83.53	92	86	77.5	82.92
10	83.88	Llama-3-70B-Instruct (Prompt)	81.45	93	91.5	85	91.76	88	84	77.5	69.17
11	82.94	GPT-4o-2024-05-13 (FC)	78.91	90	87.5	84.5	86.47	78	82	75	81.25
12	82.88	GPT-4-turbo-2024-04-09 (FC)	74.73	90	89.5	88	82.94	88	76	67.5	88.75
13	81.82	Claude-3-Sonnet-20240229 (Prompt)	83.09	89	88	89.5	93.53	92	84	77.5	51.25
14	81.35	Mistral-Medium-2312 (Prompt)	80.55	92	84	78.5	65.88	76	82	70	88.33
15	80.53	GPT-4o-2024-05-13 (Prompt)	85.09	84	78.5	61	90	78	70	72.5	82.5
16	80.47	Functionary-Medium-v2.4 (FC)	79.45	90.5	87.5	85	68.82	84	80	70	74.17
17	80.35	Gemini-1.5-Flash-Preview-0514 (FC)	80.91	93.5	78	73.5	81.76	90	54	72.5	79.58
18	80.35	Command-R-Plus (Prompt) (Optimized)	82.91	88.5	81	82	92.94	90	84	80	54.17
19	80.29	Command-R-Plus (Prompt) (Original)	82.55	90	80	83	92.94	88	84	80	53.75
20	79.94	Functionary-Small-v2.4 (FC)	82.18	88.5	82	81.5	78.24	82	80	65	67.92
21	79.76	Command-R-Plus (FC) (Optimized)	79.09	91	88	82.5	81.18	86	74	67.5	63.75
22	77.47	Claude-3-Opus-0229 (FC)	82.73	91.5	58	60	90.59	94	38	62.5	82.5
23	76.47	Claude-instant-1.2 (Prompt)	79.82	85.5	83	67.5	84.71	80	82	65	57.5
24	74.41	xLAM-1B (FC)	75.09	80.5	76.5	63	79.41	80	78	62.5	72.08
25	74.29	Claude-3-Haiku-0207 (Prompt)	84.91	91.5	84.5	55.5	92.94	94	70	25	34.58
26	71.41	Claude-2.1 (Prompt)	80.18	76	55.5	52.5	71.18	84	46	47.5	83.33
27	70.94	Command-R-Plus (FC) (Original)	74.91	90	82	76.5	81.76	88	68	55	24.17
28	68.76	Mistral-large-2402 (FC Auto)	66.91	94.5	25.5	72	83.53	96	8	52.5	84.17
29	67	Gemini-1.0-Pro-001 (FC)	79.09	92.5	30	25.5	86.47	84	44	12.5	80
30	65.88	DBRX-Instruct (Prompt)	64	71.5	72	59	71.18	86	80	62.5	55.83
31	65.18	snowflake-instruct (Prompt)	62.36	69	59	54	87.65	86	74	72.5	59.58
32	64.35	Mistral-large-2402 (FC Any)	81.45	93.5	31.5	79.5	94.71	92	8	65	0
33	63.88	GPT-3.5-Turbo-0125 (FC)	61.45	66	90.5	81	93.53	80	82	70	2.08
43	43.71	Gemma-7b-it (Prompt)	42.18	48	30	44	30	32	40	25	70.83
44	40.76	Mistral-Small-2402 (Prompt)	5.82	68	79	68.5	34.12	20	68	30	98.33
45	40.41	Deepseek-v1.5 (Prompt)	39.27	49	37	28.5	37.06	38	36	12.5	57.08
46	23.71	Mistral-small-2402 (FC Auto)	2	25.5	3	3	56.47	70	6	5	99.58

Salesforce just published new research on APIGen, an automated system that generates optimal datasets for AI training on function calling tasks — enabling the company’s xLAM model to outperform much larger rivals.

- APIGen is designed to help models train on datasets that better reflect the real-world complexity of API usage.
- Salesforce trained a both 7B and 1B parameter version of xLAM using APIGen, testing them against key function calling benchmarks.
- xLAM’s 7B parameter model ranked 6th out of 46 models, matching or surpassing rivals 10x its size — including GPT-4.
- xLAM’s 1B ‘Tiny Giant’ outperformed models like Claude Haiku and GPT-3.5, with CEO Mark Benioff calling it the best ‘micro-model’ for function calling.

While the AI race has been focused on building ever-larger models, Salesforce’s approach suggests that smarter data curation can lead to more efficient systems. The research is also a major step towards better on-device, agentic AI — packing the power of large models into a tiny frame.

Source: <https://x.com/Benioff/status/1808365628551844186>

 Turn thoughts into polished content



PROMPT: "I WANT TO TRY AN EXERCISE WHERE I TURN MY SPOKEN THOUGHTS INTO A WELL-WRITTEN TEXT. I'LL START SPEAKING, AND YOU SHOULD LISTEN AND JUST SAY THE WORD 'CONTINUE' IF I PAUSE FOR TOO LONG. ONCE I SAY 'I'M DONE' OR 'FINISHED,' I'D LIKE YOU TO TRANSFORM MY THOUGHTS INTO A WELL-WRITTEN TEXT. AFTER THAT, PLEASE SUMMARIZE THE MAIN POINTS AND EXTRACT ANY ACTION ITEMS FROM OUR CONVERSATION."

ChatGPT's voice mode feature now allows you to convert your spoken ideas into well-written text, summaries, and action items, boosting your creativity and productivity.

1. Enable "Background Conversations" in the ChatGPT app settings.
2. Start a new chat with the prompt shown in the image above (it was too long for this email).
3. Speak your thoughts freely, pausing as needed, and say "I'm done" when you've expressed all your ideas.
4. Review the AI-generated text, summary, and action items, and save them to your notes.

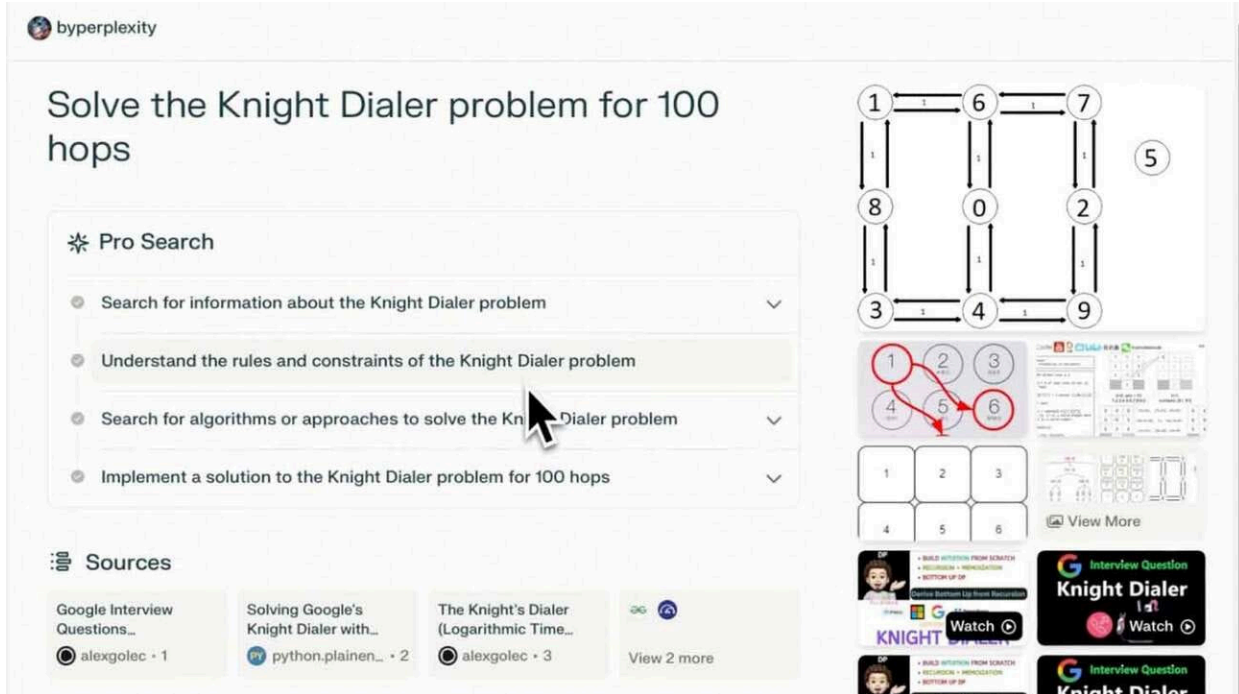
Pro tip: Try going on a long walk and rambling any ideas to ChatGPT using this trick — you'll be amazed by the summary you get at the end.

Source:

<https://university.therundown.ai/c/daily-tutorials/transform-your-thoughts-into-polished-content-with-ai-2116bbea-8001-4915-87d2-1bdd045f3d38>



Perplexity gets major research upgrade



Perplexity just announced new upgrades to its 'Pro Search' feature, enhancing capabilities for complex queries, multi-step reasoning, integration of Wolfram Alpha for math improvement, and more.

- Pro Search can now tackle complex queries using multi-step reasoning, chaining together multiple searches to find more comprehensive answers.
- A new integration with Wolfram Alpha allows for solving advanced mathematical problems, alongside upgraded code execution abilities.
- Free users get 5 Pro Searches every four hours, while subscribers to the \$20/month plan get 600 per day.
- The upgrade comes amid recent controversy over Perplexity's data scraping and attribution practices.

Given Google's struggles with AI overviews, Perplexity's upgrades will continue the push towards 'answer engines' that take the heavy lifting out of the user's hand. But the recent accusations aren't going away — and could cloud the whole AI-powered search sector until precedent is set.

Source:

<https://www.perplexity.ai/hub/blog/pro-search-upgraded-for-more-advanced-problem-solving>

Cloudflare released a free tool to detect and block AI bots circumventing website scraping protections, aiming to address concerns over unauthorized data collection for AI training. Source:

<https://blog.cloudflare.com/declaring-your-ai-dependence-block-ai-bots-scrapers-and-crawlers-with-a-single-click>

App Store chief Phil Schiller is joining OpenAI's board in an observer role, representing Apple as part of the recently announced AI partnership. Source:

<https://www.bloomberg.com/news/articles/2024-07-02/apple-to-get-openai-board-observer-role-as-part-of-ai-agreement>

Shanghai AI Lab introduced InternLM 2.5-7B, a model with a 1M context window and the ability to use tools that surged up the Open LLM Leaderboard upon release. Source:

https://x.com/intern_lm/status/1808501625700675917

Magic is set to raise over \$200M at a \$1.5B valuation, despite having no product or revenue yet — as the company continues to develop its coding-specialized models that can handle large context windows.

Source:

<https://www.reuters.com/technology/artificial-intelligence/ai-coding-startup-magic-seeks-15-billion-valuation-new-funding-round-sources-say-2024-07-02/>

Citadel CEO Ken Griffin told the company's new class of interns that he is 'not convinced' AI will achieve breakthroughs that automate human jobs in the next three years. Source:

<https://www.cnbc.com/2024/07/01/ken-griffin-says-hes-not-convinced-ai-will-replace-human-jobs-in-near-future.html>

ElevenLabs launched Voice Isolator, a new feature designed to help users remove background noise from recordings and create studio-quality audio. Source:

<https://x.com/elevenlabsio/status/1808589239744921663?>


Enjoying these AI updates, subscribe and listen to our podcast at:

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Innovations web site and Get a copy of our AI Unraveled Book at https://readaloudforme.com/index_daily_ai_chronicles.html


A Daily chronicle of AI Innovations July 03rd 2024:


 **Apple joins OpenAI board**

 **Google's emissions spiked by almost 50% due to AI boom**

 **Meta's new AI can create 3D objects from text in under a minute**

 **Meta's 3D Gen creates 3D assets at lightning speed**

 **Perplexity AI upgrades Pro Search with more advanced problem-solving**

 **The first Gen AI framework that keeps your prompts always encrypted**

 **ElevenLabs launches 'Iconic Voices'**

 **Leaks reveal Google Pixel AI upgrades**

 **Meta's new text-to-3D AI**

Enjoying these AI updates, subscribe and listen to our podcast at:
<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Innovations web site and Get a copy of our AI Unraveled Book at
https://readaloudforme.com/index_daily_ai_chronicles.html

⚡ Meta's 3D Gen creates 3D assets at lightning speed

Meta has introduced Meta 3D Gen, a new state-of-the-art, fast pipeline for text-to-3D asset generation. It offers 3D asset creation with high prompt fidelity and high-quality 3D shapes and textures in less than a minute.

According to Meta, the process is three to 10 times faster than existing solutions. The research paper even mentions that when assessed by professional 3D artists, the output of 3DGen is preferred a majority of time compared to industry alternatives, particularly for complex prompts, while being from 3× to 60× faster.

A significant feature of 3D Gen is its support physically-based rendering (PBR), necessary for 3D asset relighting in real-world applications.



Why does it matter?

3D Gen's implications extend far beyond Meta's sphere. In gaming, it could speed up the creation of expansive virtual worlds, allowing rapid prototyping. In architecture and industrial design, it could facilitate quick concept visualization, expediting the design process.

Source: <https://ai.meta.com/research/publications/meta-3d-gen/>

Perplexity AI upgrades Pro Search with more advanced problem-solving

Perplexity AI has improved Pro Search to tackle more complex queries, perform advanced math and programming computations, and deliver even more thoroughly researched answers. Everyone can use Pro Search five times every four hours for free, and Pro subscribers have unlimited access.

Perplexity suggests the upgraded Pro Search “can pinpoint case laws for attorneys, summarize trend analysis for marketers, and debug code for developers—and that’s just the start”. It can empower all professions to make more informed decisions.

Why does it matter?

This showcases AI's potential to assist professionals in specialized fields. Such advancements also push the boundaries of AI's practical applications in research and decision-making processes.

Source:

<https://www.perplexity.ai/hub/blog/pro-search-upgraded-for-more-advanced-problem-solving>

The first Gen AI framework that keeps your prompts always encrypted

Edgeless Systems introduced Continuum AI, the first generative AI framework that keeps prompts encrypted at all times with confidential computing by combining confidential VMs with NVIDIA H100 GPUs and secure sandboxing.

The Continuum technology has two main security goals. It first protects the user data and also protects AI model weights against the infrastructure, the service provider, and others. Edgeless Systems is also collaborating with NVIDIA to empower businesses across sectors to confidently integrate AI into their operations.

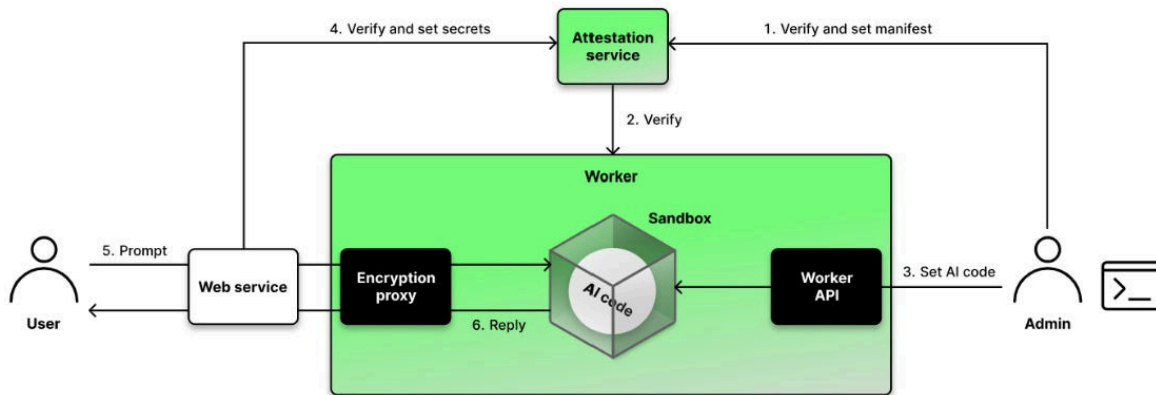


Figure 2. A workflow displaying interactions between the different components of Continuum and the end user

Why does it matter?

This greatly advances security for LLMs. The technology could be pivotal for a future where organizations can securely utilize AI, even for the most sensitive data.

Source:

<https://developer.nvidia.com/blog/advancing-security-for-large-language-models-with-nvidia-gpus-and-edgeless-systems>

RunwayML's Gen-3 Alpha models is now generally available

Announced a few weeks ago, Gen-3 is Runway's latest frontier model and a big upgrade from Gen-1 and Gen-2. It allows users to produce hyper-realistic videos from text, image, or video prompts. Users must upgrade to a paid plan to use the model.

Source:

<https://venturebeat.com/ai/runways-gen-3-alpha-ai-video-model-now-available-but-theres-a-catc>
[h](#)

Meta might be bringing generative AI to metaverse games

In a job listing, Meta mentioned it is seeking to research and prototype “new consumer experiences” with new types of gameplay driven by Gen AI. It is also planning to build Gen AI-powered tools that could “improve workflow and time-to-market” for games.

Source:

<https://techcrunch.com/2024/07/02/meta-plans-to-bring-generative-ai-to-metaverse-games>

Apple gets a non-voting seat on OpenAI's board

As a part of its AI agreement with OpenAI, Apple will get an observer role on OpenAI's board. Apple chose Phil Schiller, the head of Apple's App Store and its former marketing chief, for the position.

Source: <https://www.theverge.com/2024/7/2/24191105/apple-phil-schiller-join-openai-board>

Figma disabled AI tool after being criticised for ripping off Apple's design

Figma's Make Design feature generates UI layouts and components from text prompts. It repeatedly reproduced Apple's Weather app when used as a design aid, drawing accusations that Figma's AI seems heavily trained on existing apps.

Source:

<https://techcrunch.com/2024/07/02/figma-disables-its-ai-design-feature-that-appeared-to-be-ripping-off-apples-weather-app>

China is far ahead of other countries in generative AI inventions

According to the World Intellectual Property Organization (WIPO), more than 50,000 patent applications were filed in the past decade for Gen AI. More than 38,000 GenAI inventions were filed by China between 2014-2023 vs. only 6,276 by the U.S.

Source:

<https://www.reuters.com/technology/artificial-intelligence/china-leading-generative-ai-patents-race-un-report-says-2024-07-03>

Apple joins OpenAI board

- Phil Schiller, Apple's former marketing head and App Store chief, will reportedly join OpenAI's board as a non-voting observer, according to Bloomberg.
- This role will allow Schiller to understand OpenAI better, as Apple aims to integrate ChatGPT into iOS and macOS later this year to enhance Siri's capabilities.
- Microsoft also took a non-voting observer position on OpenAI's board last year, making it rare and significant for both Apple and Microsoft to be involved in this capacity.

Source: <https://www.theverge.com/2024/7/2/24191105/apple-phil-schiller-join-openai-board>

Google's emissions spiked by almost 50% due to AI boom

- Google reported a 48% increase in greenhouse gas emissions over the past five years due to the high energy demands of its AI data centers.
- Despite achieving seven years of renewable energy matching, Google faces significant challenges in meeting its goal of net zero emissions by 2030, highlighting the uncertainties surrounding AI's environmental impact.
- To address water consumption concerns, Google has committed to replenishing 120% of the water it uses by 2030, although in 2023, it only managed to replenish 18%.

Source:

<https://www.techradar.com/pro/google-says-its-emissions-have-grown-nearly-50-due-to-ai-data-center-boom-and-heres-what-it-plans-to-do-about-it>

Meta's new AI can create 3D objects from text in under a minute



- Meta has introduced 3D Gen, an AI system that creates high-quality 3D assets from text descriptions in under a minute, significantly advancing 3D content generation.
- The system uses a two-stage process, starting with AssetGen to generate a 3D mesh with PBR materials and followed by TextureGen to refine the textures, producing detailed and professional-grade 3D models.
- 3D Gen has shown superior performance and visual quality compared to other industry solutions, with potential applications in game development, architectural visualization, and virtual/augmented reality.

Source:

<https://www.maginate.com/article/meta-unveils-3d-gen-ai-that-creates-detailed-3d-assets-in-under-a-minute/>

Enjoying these AI updates, subscribe and listen to our podcast at:

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Innovations web site and Get a copy of our AI Unraveled Book at


https://readaloudforme.com/index_daily_ai_chronicles.html

A Daily chronicle of AI Innovations July 02nd 2024:

 **JARVIS-inspired Grok 2 aims to answer any user query**

 **Apple unveils a public demo of its '4M' AI model**

 **Amazon hires Adept's top executives to build an AGI team**


 **YouTube lets you remove AI-generated content resembling face or voice**

 **Runway opens Gen-3 Alpha access**

 **Motorola hits the AI runway**

 **Meta swaps 'Made with AI' label with 'AI info' to indicate AI photos**

 **Deepfakes to cost \$40 billion by 2027: Deloitte survey**

 **Anthropic launches a program to fund the creation of reliable AI benchmarks**

 **US's targeting of AI not helpful for healthy development: China**

 **New robot controlled by human brain cells**

 **Figma to temporarily disable AI feature amid plagiarism concerns**

Enjoying these AI updates, subscribe and listen to our podcast at:

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Innovations web site and Get a copy of our [AI Unraveled](#) Book at https://readaloudforme.com/index_daily_ai_chronicles.html

 **Runway opens Gen-3 Alpha access**



Runway just announced that its AI video generator, Gen-3 Alpha, is now available to all users following weeks of impressive, viral outputs after the model's release in mid-June.

- Runway unveiled Gen-3 Alpha last month, the first model in its next-gen series trained for learning 'general world models'.
- Gen-3 Alpha upgrades key features, including character and scene consistency, camera motion and techniques, and transitions between scenes.
- Gen-3 Alpha is available behind Runway's 'Standard' \$12/mo access plan, which gives users 63 seconds of generations a month.
- On Friday, we're running a free, [hands-on workshop](#) in our AI University covering how to create an AI commercial using Gen-3, ElevenLabs, and Midjourney.

Despite impressive recent releases from KLING and Luma Labs, Runway's Gen-3 Alpha model feels like the biggest leap AI video has taken since Sora. However, the tiny generation limits for non-unlimited plans might be a hurdle for power users.

Source: <https://x.com/runwayml/status/1807822396415467686>

Motorola hits the AI runway




https://youtu.be/CSfw_NjqQ2o?si=xYpZg9AwRgasLhov

Motorola just [launched](#) its ‘Styled By Moto’ ad campaign, an entirely AI-generated fashion spot promoting its new line of Razr folding smartphones — created using nine different AI tools, including Sora and Midjourney.

- The 30-second video features AI-generated models wearing outfits inspired by Motorola's iconic ‘batwing’ logo in settings like runways and photo shoots.
- Each look was created from thousands of AI-generated images, incorporating the brand's logo and colors of the new Razr phone line.
- Tools used include OpenAI's Sora, Adobe Firefly, Midjourney, Krea, Magnific, Luma, and more — reportedly taking over four months of research.
- The 30-second spot is also set to an AI-generated soundtrack incorporating the ‘Hello Moto’ jingle, created using Udio.

This is a fascinating look at the AI-powered stack used by a major brand, and a glimpse at how tools can (and will) be combined to open new creative avenues. It's also another example of the shift in discourse surrounding AI's use in marketing — potentially paving the way for wider acceptance and integration.

 **JARVIS-inspired Grok 2 aims to answer any user query**

Elon Musk has announced the release dates for two new AI assistants from xAI. The first, Grok 2, will be launched in August. Musk says Grok 2 is inspired by JARVIS from Iron Man and The Hitchhiker's Guide to the Galaxy and aims to answer virtually any user query. This ambitious goal is fueled by xAI's focus on "purging" LLM datasets used for training.



Musk also revealed that an even more powerful version, Grok 3, is planned for release by the end of the year. Grok 3 will leverage the processing power of 100,000 Nvidia H100 GPUs, potentially pushing the boundaries of AI performance even further.

Why does it matter?

These advanced AI assistants from xAI are intended to compete with and outperform AI chatbots like OpenAI's ChatGPT by focusing on data quality, user experience, and raw processing power. This will significantly advance the state of AI and transform how people interact with and leverage AI assistants.

Source: <https://www.coinspeaker.com/xai-grok-2-elon-musk-jarvis-ai-assistant/>

Apple unveils a public demo of its '4M' AI model

Apple and the Swiss Federal Institute of Technology Lausanne (EPFL) have released a public demo of the '4M' AI model on Hugging Face. The 4M (Massively Multimodal Masked Modeling) model can process and generate content across multiple modalities, such as creating images from text, detecting objects, and manipulating 3D scenes using natural language inputs.

While companies like Microsoft and Google have been making headlines with their AI partnerships and offerings, Apple has been steadily advancing its AI capabilities. The public demo of the 4M model suggests that Apple is now positioning itself as a significant player in the AI industry.

Why does it matter?

By making the 4M model publicly accessible, Apple is seeking to engage developers to build an ecosystem. It could lead to more coherent and versatile experiences, such as enhanced Siri capabilities and advancements in Apple's augmented reality efforts.

Source:

<https://venturebeat.com/ai/apple-just-launched-a-public-demo-of-its-4m-ai-model-heres-why-its-a-big-deal>

Amazon hires Adept's top executives to build an AGI team

Amazon is hiring the co-founders, including the CEO and several other key employees, from the AI startup Adept. CEO David Luan will join Amazon's AGI autonomy group, which is led by Rohit

Prasad, who is spearheading a unified push to accelerate Amazon's AI progress across different divisions like Alexa and AWS.

Amazon is consolidating its AI projects to develop a more advanced LLM to compete with OpenAI and Google's top offerings. This unified approach leverages the company's collective resources to accelerate progress in AI capabilities.

Why does it matter?

This acquisition indicates Amazon's intent to strengthen its position in the competitive AI landscape. By bringing the Adept team on board, Amazon is leveraging its expertise and specialized knowledge to advance its AGI aspirations.

Source:<https://www.bloomberg.com/news/articles/2024-06-28/amazon-hires-top-executives-from-ai-startup-adept-for-agi-team>

YouTube lets you remove AI-generated content resembling face or voice

YouTube lets people request the removal of AI-generated content that simulates their face or voice. Under YouTube's privacy request process, the requests will be reviewed based on whether the content is synthetic, if it identifies the person, and if it shows the person in sensitive behavior. Source:

<https://techcrunch.com/2024/07/01/youtube-now-lets-you-request-removal-of-ai-generated-content-that-simulates-your-face-or-voice>

Meta swaps 'Made with AI' label with 'AI info' to indicate AI photos

Meta is refining its AI photo labeling on Instagram and Facebook. The "Made with AI" label will be replaced with "AI info" to more accurately reflect the extent of AI use in images, from minor edits to the entire AI generation. It addresses photographers' concerns about the mislabeling of their photos. Source:

<https://techcrunch.com/2024/07/01/meta-changes-its-label-from-made-with-ai-to-ai-info-to-indicate-use-of-ai-in-photos>

Deepfakes to cost \$40 billion by 2027: Deloitte survey

Deepfake-related losses will increase from \$12.3 billion in 2023 to \$40 billion by 2027, growing at 32% annually. There was a 3,000% increase in incidents last year alone. Enterprises are not well-prepared to defend against deepfake attacks, with one in three having no strategy.

Source:

<https://venturebeat.com/security/deepfakes-will-cost-40-billion-by-2027-as-adversarial-ai-gains-momentum>

Anthropic launches a program to fund the creation of reliable AI benchmarks

Anthropic is launching a program to fund new AI benchmarks. The aim is to create more comprehensive evaluations of AI models, including assessing capabilities in cyberattacks and weapons and beneficial applications like scientific research and bias mitigation. Source: <https://techcrunch.com/2024/07/01/anthropic-looks-to-fund-a-new-more-comprehensive-generation-of-ai-benchmarks>

US's targeting of AI not helpful for healthy development: China

China has criticized the US approach to regulating and restricting investments in AI. Chinese officials stated that US actions targeting AI are not helpful for AI's healthy and sustainable development. They argued that the US measures will be divisive when it comes to global governance of AI.

Source:

<https://www.reuters.com/technology/artificial-intelligence/china-says-us-targeting-ai-not-helpful-healthy-development-2024-07-01>

New robot controlled by human brain cells

- Scientists in China have developed a robot with an artificial brain grown from human stem cells, which can perform basic tasks such as moving limbs, avoiding obstacles, and grasping objects, showcasing some intelligence functions of a biological brain.
- The brain-on-chip utilizes a brain-computer interface to facilitate communication with the external environment through encoding, decoding, and stimulation-feedback mechanisms.
- This pioneering brain-on-chip technology, requiring similar conditions to sustain as a human brain, is expected to have a revolutionary impact by advancing the field of hybrid intelligence, merging biological and artificial systems.

Source: <https://www.independent.co.uk/tech/robot-human-brain-china-b2571978.html>



Figma to temporarily disable AI feature amid plagiarism concerns

- Figma has temporarily disabled its "Make Design" AI feature after accusations that it was replicating Apple's Weather app designs.
- Andy Allen, founder of NotBoring Software, discovered that the feature consistently reproduced the layout of Apple's Weather app, leading to community concerns.
- CEO Dylan Field acknowledged the issue and stated the feature would be disabled until they can ensure its reliability and originality through comprehensive quality assurance checks.

Source:

<https://techcrunch.com/2024/07/02/figma-disables-its-ai-design-feature-that-appeared-to-be-ripping-off-apples-weather-app/>



Nvidia faces first antitrust charges

- French antitrust enforcers plan to charge Nvidia with alleged anticompetitive practices, becoming the first to take such action, according to Reuters.
- Nvidia's offices in France were raided last year as part of an investigation into possible abuses of dominance in the graphics cards sector.
- Regulatory bodies in the US, EU, China, and the UK are also examining Nvidia's business practices due to its significant presence in the AI chip market.

Source:

<https://finance.yahoo.com/news/french-antitrust-regulators-set-charge-151406034.html?>

Enjoying these AI updates, subscribe and listen to our podcast at:

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

Visit our Daily AI Innovations web site and Get a copy of our [AI Unraveled](#) Book at https://readaloudforme.com/index_daily_ai_chronicles.html

A Daily chronicle of AI Innovations July 01st 2024:

 Some Apple Intelligence features may be put behind a paywall

 Meta's new dataset could enable robots to learn manual skills from human experts

 Google announces advancements in Vertex AI models

 LMSYS's new Multimodal Arena compares top AI models' visual processing abilities

 Apple's Vision Pro gets an AI upgrade

 Humanoid robots head to the warehouse

 Google Translate adds 110 languages

Enjoying these AI updates, subscribe and listen to our podcast at:

<https://podcasts.apple.com/ca/podcast/ai-unraveled-latest-ai-news-trends-gpt-gemini-generative/id1684415169>

 Google announces advancements in Vertex AI models

Google has rolled out significant improvements to its Vertex AI platform, including the general availability of Gemini 1.5 Flash with a massive 1 million-token context window. Also, Gemini 1.5

Pro now offers an industry-leading 2 million-token context capability. Google is introducing context caching for these Gemini models, slashing input costs by 75%.

Moreover, Google launched Imagen 3 in preview and added third-party models like Anthropic's Claude 3.5 Sonnet on Vertex AI.

They've also made Grounding with Google Search generally available and announced a new service for grounding AI agents with specialized third-party data. Plus, they've expanded data residency guarantees to 23 countries, addressing growing data sovereignty concerns.

Why does it matter?

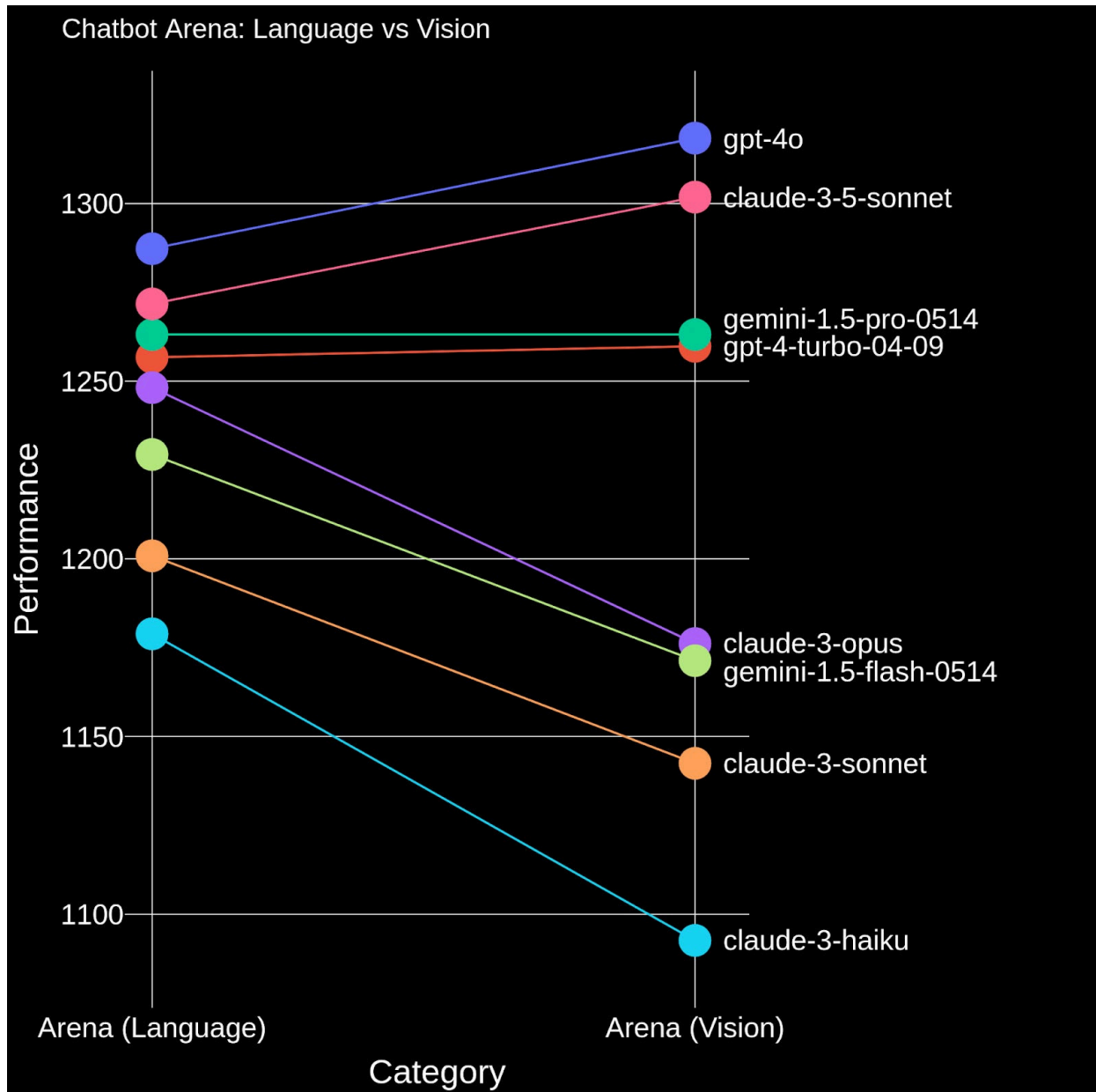
Google is positioning Vertex AI as the most "enterprise-ready" generative AI platform. With expanded context windows and improved grounding capabilities, this move also addresses concerns about the accuracy of Google's AI-based search features.

Source:

<https://cloud.google.com/blog/products/ai-machine-learning/vertex-ai-offers-enterprise-ready-generative-ai>



LMSYS's new Multimodal Arena compares top AI models' visual processing abilities



LMSYS Org added image recognition to Chatbot Arena to compare vision language models (VLMs), collecting over 17,000 user preferences in just two weeks. OpenAI's GPT-4o and Anthropic's Claude 3.5 Sonnet outperformed other models in image recognition. Also, the open-source LLaVA-v1.6-34B performed comparably to some proprietary models.

These AI models tackle diverse tasks, from deciphering memes to solving math problems with visual aids. However, the examples provided show that even top models can stumble when interpreting complex visual information or handling nuanced queries.

Why does it matter?

This leaderboard isn't just a tech popularity contest—it shows how advanced AI models can decode images. However, the varying performance also serves as a reality check, reminding us that while AI can recognize a cat in a photo, it might struggle to interpret your latest sales graph.

Source: <https://lmsys.org/blog/2024-06-27-multimodal>

🕶️ Apple's Vision Pro gets an AI upgrade



Apple is reportedly working to bring its Apple Intelligence features to the Vision Pro headset, though not this year. Meanwhile, Apple is tweaking its in-store Vision Pro demos, allowing potential buyers to view personal media and try a more comfortable headband. Apple's main challenge is adapting its AI features to a mixed-reality environment.

The company is tweaking its retail strategy for Vision Pro demos, hoping to boost sales of the pricey headset. Apple is also exploring the possibility of monetizing AI features through subscription services like "Apple Intelligence+."

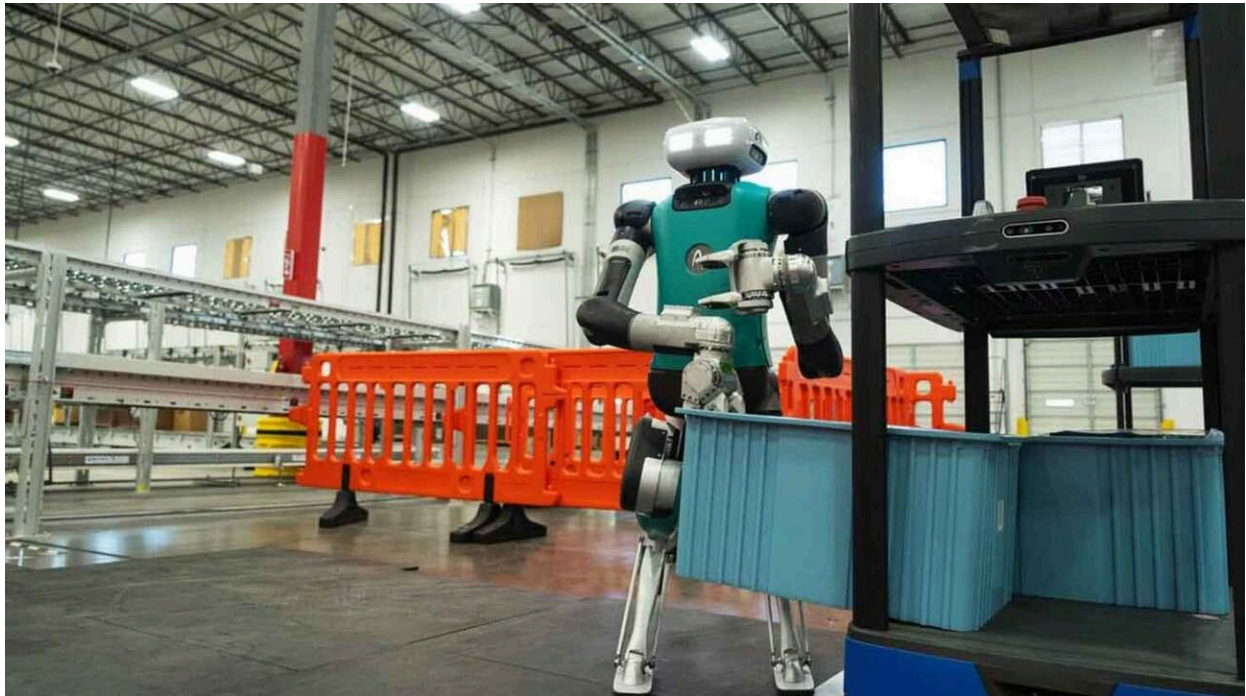
Why does it matter?

Apple's Vision Pro, with its 16GB RAM and M2 chip, can handle advanced AI tasks. However, cloud infrastructure limitations are causing a delay in launch. It's a classic case of "good things come to those who wait."

Source:

<https://www.bloomberg.com/news/newsletters/2024-06-30/apple-s-longer-lasting-devices-ios-19-and-apple-intelligence-on-the-vision-pro-ly1jnrw4>

Humanoid robots head to the warehouse



Agility Robotics just signed a multi-year deal with GXO Logistics to bring the company's Digit humanoid robots to warehouses, following a successful pilot in Spanx facilities in 2023.

The agreement is being hailed as the first Robots-as-a-Service (RaaS) deal and 'formal commercial deployment' of the humanoid robots.

Agility's Digit robots will be integrated into GXO's logistics operations at a Spanx facility in Connecticut, handling repetitive tasks and logistics work.

The 5'9" tall Digit can lift up to 35 pounds, and integrates with a cloud-based Agility Arc platform to control full fleets and optimize facility workflows.

Digit tested a proof-of-concept trial with Spanx in 2023, with Amazon also testing the robots at its own warehouses.

Is RaaS the new SaaS? Soon, every company will be looking to adopt advanced robotics into their workforce — and subscription services could help lower the financial and technical barriers needed to scale without the massive upfront costs.

Source:

<https://agilityrobotics.com/content/gxo-signs-industry-first-multi-year-agreement-with-agility-robotics>

Google Translate adds 110 languages



Google just announced its largest-ever expansion of Google Translate, adding support for 110 new languages enabled by the company's PaLM 2 LLM model.

The new languages represent over 614M speakers, covering about 8% of the global population.

Google's PaLM 2 model was the driving force behind the expansion, helping unlock translations for closely related languages.

The expansion also includes some languages with no current native speakers, displaying how AI models can help preserve 'lost' dialects.

The additions are part of Google's '1,000 Languages Initiative,' which aims to build AI that supports all of the world's spoken languages.

We've talked frequently about AI's coming power to break down language barriers with its translation capabilities — but the technology is also playing a very active role in both uncovering and preserving languages from lost and endangered cultures.

Source: <https://blog.google/products/translate/google-translate-new-languages-2024>

Amazon's Q AI assistant for enterprises gets an update for call centers

The update provides real-time, step-by-step guides for customer issues. It aims to reduce the "toggle tax" - time wasted switching between applications. The system listens to calls in real-time and automatically provides relevant information.

Source:

<https://venturebeat.com/ai/amazon-upgrades-ai-assistant-q-to-make-call-centers-way-more-efficient>

WhatsApp is developing a feature to choose Meta AI Llama models

Users will be able to choose between two options: faster responses with Llama 3-70B (default) or more complex queries with Llama 3-405B (advanced). Llama 3-405B will be limited to a certain number of prompts per week. This feature aims to give users more control over their AI interactions.

Source: <https://wabetainfo.com/whatsapp-beta-for-android-2-24-14-7-whats-new/>

Bill Gates says AI's energy consumption isn't a major concern

He claims that while data centers may consume up to 6% of global electricity, AI will ultimately drive greater energy efficiency. Gates believes tech companies will invest in green energy to power their AI operations, potentially offsetting the increased demand.

Source: https://www.theregister.com/2024/06/28/bill_gates_ai_power_consumption

Amazon is investigating Perplexity AI for possible scraping abuse

Perplexity appears to be scraping websites that have forbidden access through robots.txt. AWS prohibits customers from violating the robots.txt standard. Perplexity uses an unpublished IP address to access websites that block its official crawler. The company claims a third party performs web crawling for them.

Source: <https://www.wired.com/story/aws-perplexity-bot-scraping-investigation>

Microsoft AI chief claims content on the open web is "freeware"

Mustafa Suleyman claimed that anything published online becomes "freeware" and fair game for AI training. This stance, however, contradicts basic copyright principles and ignores the legal complexities of fair use. He suggests that robots.txt might protect content from scraping.

Source:

<https://www.theverge.com/2024/6/28/24188391/microsoft-ai-suleyman-social-contract-freeware>

Some Apple Intelligence features may be put behind a paywall

Apple Intelligence, initially free, is expected to introduce a premium "Apple Intelligence+" subscription tier with additional features, similar to iCloud, according to Bloomberg's Mark Gurman. Apple plans to monetize Apple Intelligence not only through direct subscriptions but also by taking a share of revenue from partner AI services like OpenAI and potentially Google Gemini. Apple Intelligence will be integrated into multiple devices, excluding the HomePod due to hardware limitations, and may include a new robotic device, making it comparable to iCloud in its broad application and frequent updates. Source:

<https://www.techradar.com/computing/is-apple-intelligence-the-new-icloud-ai-platform-tipped-to-get-new-subscription-tier>

Meta's new dataset could enable robots to learn manual skills from human experts

Meta has introduced a new benchmark dataset named HOT3D to advance AI research in 3D hand-object interactions, containing over one million frames from various perspectives. This dataset aims to enhance the understanding of human hand manipulation of objects, addressing a significant challenge in computer vision research according to Meta. HOT3D includes over 800 minutes of egocentric video recordings, multiple perspectives, detailed 3D pose annotations, and 3D object models, which could help robots and XR devices learn manual skills from human experts. Source:

<https://the-decoder.com/metass-new-hot3d-dataset-could-enable-robots-to-learn-manual-skills-from-human-experts/>