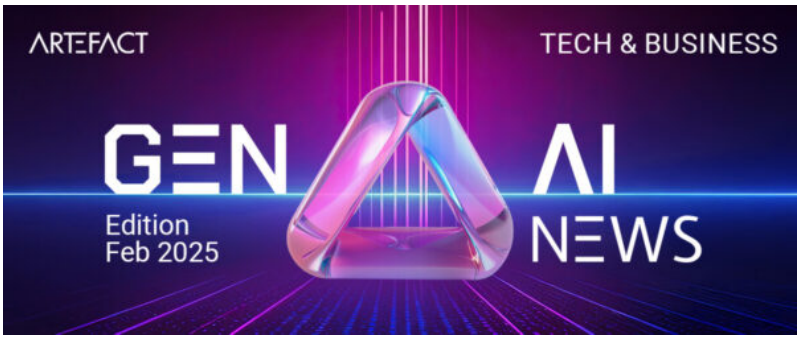


GenAI Newsletter March #1 | OpenAI's agent invasion, Manus, Mistral OCR...



Dear Artefactors and AI enthusiasts, Let's cut to the chase – this week's Gen AI newsletter is packed with the good stuff! We're talking OpenAI's agent invasion (in a good way!), Microsoft's Copilot getting a brain upgrade, and China's AI scene making waves. Basically, if you want to know what's up in AI, you're in the right place. Let's dive in!

GenAI News Podcast

Don't have time to read this newsletter?

Listen to the 13' podcast

Top Highlights

Elon Musk's "anti-woke" AI chatbot, Grok, has assessed a 75-85% probability that President Donald Trump is compromised by Russian interests. This evaluation considers Trump's financial ties, personal interactions, and consistent favorable stance toward Russia, suggesting he may be an unwitting asset due to his ego and debts.

OpenAI is developing AI "agents" for professionals, with prices up to $20,000/month. These include agents for sales lead management ($2,000/month), software development ($10,000/month), and PhD-

level research ($20,000/month). This move aims to offset OpenAI's $5 billion loss last year. SoftBank, an investor, has committed $3 billion to these agent products this year.

Mistral AI has launched Mistral OCR, an advanced optical character recognition API that sets a new standard in document comprehension, surpassing competitors like Google and Microsoft in accuracy and speed. Announced on March 6, 2025, this tool excels at processing complex documents – extracting text, tables, equations, and images in multiple languages – and converting them into structured formats like Markdown or JSON. With a processing speed of 2,000 pages per minute and a 94.89% accuracy rate, Mistral OCR aims to transform how businesses digitize and utilize vast document libraries, offering both cloud-based and on-premises options for enhanced privacy.

Manus, a new AI agentic system from China, is generating buzz as a potential "second DeepSeek moment" due to its ability to autonomously handle complex tasks, though skepticism remains about its originality and scalability. Developed by a startup named Monica, Manus combines multiple AI models, including Anthropic's Claude 3.5 Sonnet and Alibaba's Qwen, to perform functions like research, data analysis, report generation, workflow automation, and code deployment. Unlike DeepSeek, which built its own foundation model, Manus relies on existing large language models, leading some, like TechCrunch's Kyle Wiggers, to argue it's more of a remix than a groundbreaking innovation. Currently in private beta with limited access via invitation codes, its viral rise in China and claims of outperforming tools like OpenAI's Deep Research have sparked both excitement and debate about its true potential.

Business News & Market Insights

#1. Hyperscaler generative AI services are becoming more affordable as pricing stabilizes in a rapidly evolving market, according to a recent analysis by Seeking Alpha. The article, published on March 11, 2025, highlights how cloud providers like AWS, Microsoft, and Alphabet are adjusting prices to reflect demand and competition, despite massive investments in AI infrastructure. With no clear "comps" in this new GenAI landscape, hyperscalers are adding AI features globally – spanning 27 countries in 2024 – while service introductions outpace price changes by 20 to 1, signaling a push to make AI more accessible and cost-effective for users.

#2. Andreessen Horowitz's fourth edition of the Top 100 Gen AI Consumer Apps reveals a rapidly evolving consumer AI landscape, with ChatGPT maintaining its lead at 400 million weekly active users while DeepSeek surges to second place. The report, analyzing web and mobile app data from SimilarWeb and Sensor Tower as of January 2025, highlights significant shifts: 17 new web entrants and 18 new mobile apps, alongside advancements in AI video and "vibecoding" tools that democratize creation. ChatGPT's growth, doubling to 400 million users since August 2024, underscores its dominance, driven by innovations like GPT-4o, while competitors like DeepSeek challenge the status quo, signaling a dynamic, competitive market.

#3. GenAI is poised to transform the value of expertise by impacting approximately 50 million jobs, automating certain tasks while enhancing workers' capabilities in others, according to a recent Harvard Business Review article. This shift will force companies to overhaul their organizational structures and talent management strategies, reshaping how businesses operate and compete. The technology is expected to redraw traditional learning curves, creating new paradigms for skill acquisition and career progression. As a result, both businesses and individuals must adapt to a future where expertise is redefined by AI's capabilities and accessibility.

#4. China is experiencing a significant surge in artificial intelligence (AI) development, integrating AI technologies into various aspects of daily life, from advanced chatbots to intelligent toys. This rapid adoption is driven by substantial government investment and a thriving tech industry, positioning China as a global leader in AI innovation.

#5. The U.S. Army's Training and Doctrine Command (TRADOC) is utilizing an AI tool named CamoGPT to identify and remove references to diversity, equity, inclusion, and accessibility (DEIA) from training materials. This initiative aligns with President Donald Trump's January 27 executive order, which mandates the elimination of policies perceived as promoting divisive theories on race and gender. Developed by the Army's Artificial Intelligence Integration Center (AI2C) using Meta's Llama 3.3 model, CamoGPT enables rapid scanning and modification of documents to comply with the directive. The AI tool is part of a broader Department of Defense effort to revise military training and policies concerning DEIA content.

> Major deals, Partnerships & Innovations

#1. Microsoft has upgraded its Copilot AI by offering free, unlimited access to its Voice and Think Deeper features, integrating OpenAI's advanced o1 reasoning model to enhance user experience. Announced recently, this update allows all users to engage in extended voice conversations and tackle complex queries without restrictions, a move that significantly boosts Copilot's capabilities. The Voice feature supports practical tasks like language practice or hands-free guidance, while Think Deeper, powered by the o1 model, excels at in-depth problem-solving.

#2. OpenAI has launched a $50 million grant program called NextGenAI to support AI-assisted research at top universities, aiming to accelerate scientific discoveries and prepare future AI leaders. Announced on March 4, 2025, the initiative includes 15 founding academic partners such as Harvard, MIT, and the University of Oxford, offering research grants, compute funding, and API access to students, educators, and researchers. The program builds on OpenAI's earlier educational efforts, like ChatGPT Edu, and seeks to advance critical AI work amid concerns over reduced U.S. government support for such research. While positioned as a boost for innovation, some see it as a strategic move by OpenAI to promote its tools over competitors' in academic settings.

#3. The U.S. Department of Defense has partnered with Scale AI to integrate AI agents into military planning and operations, marking a significant step in leveraging artificial intelligence for defense purposes. Announced on March 5, 2025, this contract with the Defense Innovation Unit's Thunderforge program will use Microsoft's large language models to modernize decision-making processes, starting with air combat simulations and potentially expanding across all military branches. Scale AI, valued at $13.8 billion, aims to enhance efficiency by deploying "human-in-the-loop" AI agents, though concerns linger about over-reliance on AI in critical operations. This move reflects the Pentagon's push to adopt cutting-edge technology amid global competition, particularly with nations like China.

#1. Microsoft Research has introduced BioEmu-1, a groundbreaking deep learning model that can generate thousands of protein structures per hour, offering new insights into protein dynamics for scientists and drug discovery. Unlike traditional methods that provide a single static protein structure, BioEmu-1 captures the diverse, flexible conformations proteins can adopt, likened to viewing a movie rather than a snapshot. This model, now open-source, leverages vast datasets and achieves significant computational efficiency over classical molecular dynamics

simulations, accurately predicting protein stability and structural distributions.

#2. Coca-Cola has discovered that generative AI significantly enhances creativity and efficiency across its operations. Pratik Thakar, the company's head of generative AI, shared insights from their journey, highlighting how the technology streamlines processes like product design, marketing campaigns, and internal workflows. By integrating AI tools, Coca-Cola has accelerated innovation while maintaining a human touch in storytelling, positioning the brand to adapt swiftly in a competitive market. This strategic embrace of AI underscores its potential to transform traditional business practices while amplifying creative output.

#3. Larry Page, co-founder of Google, has launched a new AI startup named Dynatomics, with the objective to revolutionize product manufacturing. The company focuses on using large language models to create highly optimized designs for various objects, which can then be manufactured in factories. Chris Anderson, former CTO of Page-backed electric airplane startup Kittyhawk, leads this stealth initiative. This move aligns with a broader trend of integrating AI into manufacturing processes, with companies like Orbital Materials and PhysicsX also exploring AI applications in material discovery and engineering simulations.

#4. Elon Musk's Department of Government Efficiency (DOGE) has deployed a proprietary chatbot, GSAi, to 1,500 federal workers within the General Services Administration (GSA) to automate tasks (drafting emails, summarizing texts, and writing code) and enhance productivity. This may lead to federal workforce downsizing. While aligning with President Trump's modernization agenda, this move raises concerns among civil liberties organizations about potential violations of federal laws and the need for legislative oversight to protect personal information.

New Models & Innovations

#1. Google has introduced "AI Mode" a new experimental feature in Google Search that enhances user queries with advanced AI capabilities, aiming to provide more comprehensive and diverse responses to complex questions. Announced on March 4, 2025, this feature, built into Search via the Labs program, leverages a custom Gemini 2.0 model to combine high-quality web content with real-time data, such as the Knowledge Graph and shopping information, to deliver easy-to-understand answers. Unlike traditional search, AI Mode uses a "query fan-out" technique to explore multiple subtopics

simultaneously, supports multimodal inputs like text, voice, and images, and allows follow-up questions for deeper exploration.

#2. Apple is introducing AI-generated review summaries to the App Store with the release of iOS 18.4. These summaries condense user reviews into concise paragraphs, highlighting both positive and negative feedback. It is initially available in the U.S. for English-language apps with sufficient reviews. Apple plans to update these summaries weekly and expand the feature to other languages and regions later this year. This initiative mirrors Amazon's 2023 implementation of AI-generated review highlights.

#3. Booking.com is leveraging GenAI to gain deeper insights into traveler intent, enhancing its ability to personalize user experiences. This initiative aligns with the company's broader strategy to integrate AI technologies across its platforms, aiming to streamline travel planning and booking processes. In line with this strategy, Booking Holdings has introduced tools like the AI Trip Planner on Booking.com and an AI-powered travel assistant named Penny on Priceline. These innovations are designed to provide more seamless and personalized experiences for travelers.

#4. Microsoft's Dragon Copilot, an AI voice assistant, is transforming healthcare by reducing administrative burdens and improving clinician workflows through advanced voice recognition and ambient listening technology. Built on the Microsoft Cloud for Healthcare, it integrates Dragon Medical One’s speech capabilities with DAX Copilot’s ambient AI, enabling clinicians to save approximately five minutes per patient encounter by automating tasks like clinical note creation and supporting multiple languages. This innovation has led to a reported 70% reduction in clinician burnout and fatigue, with 93% of patients noting an enhanced experience due to more focused doctor interactions.

#5. Google has unveiled Gemma 3, a collection of lightweight, state-of-the-art open models built from the same research and technology that powers its Gemini 2.0 models. Designed for versatility, Gemma 3 can operate efficiently across various devices, from phones and laptops to workstations, supporting over 140 languages. Notably, it offers advanced text and visual reasoning capabilities, a 128k-token context window, and function calling for complex tasks. The models come in sizes ranging from 1 billion to 27 billion parameters, with quantized versions available for faster performance and reduced computational requirements. Google emphasizes that Gemma 3 delivers state-of-the-art performance for its size, outperforming models like Llama3-405B, DeepSeek-V3,

and o3-mini in preliminary human preference evaluations on LMArena's leaderboard.

#6. OpenAI has introduced the Responses API, designed to streamline the development of AI agents by combining the simplicity of Chat Completions with advanced tool-use capabilities. This new API supports built-in tools such as web search, file search, and computer use, enabling agents to perform tasks like retrieving real-time information, searching documents, and automating computer operations. The Responses API is available to all developers at no additional cost, with usage billed at standard rates. OpenAI plans to phase out the Assistants API in favor of the Responses API by mid-2026, providing a comprehensive migration guide to ensure a smooth transition for developers.

#7. OpenAI has introduced the Agents SDK, a comprehensive toolkit designed to assist developers in creating and managing AI agents capable of performing complex tasks autonomously. This SDK facilitates the orchestration of multiple AI agents, enabling them to collaborate effectively to achieve intricate objectives. The Agents SDK complements the newly released Responses API, which offers foundational tools for developers to build agents equipped with functionalities such as web search, file retrieval, and computer task execution. OpenAI plans to phase out its previous Assistants API by mid-2026, encouraging developers to transition to the enhanced capabilities provided by the Responses API and Agents SDK.

Thoughts of the week by Hanan Ouazan

Managing Partner & Global Lead AI Acceleration

The Future of AI Is Seamless, Adaptive, and Invisible

The Future of AI Is Seamless, Adaptive, and Invisible

AI is rapidly evolving toward a future where users no longer think about models, modes, or tools. Instead, systems are becoming **dynamically adaptive** – automatically adjusting reasoning depth, selecting the right tools, and managing resources in real time based on task complexity. What was once a fragmented landscape – separate models for different tasks, manual switching between functions – is converging into **unified, autonomous intelligence.** This marks the end of AI as a collection of discrete capabilities and the rise of AI as an **integrated, fluid experience**. As Satya Nadella succinctly put it: "The models are getting commoditized. OpenAI is not a model company- it's a product company." The model itself is

no longer the differentiator. What matters now is **how intelligently capabilities are orchestrated**, and **how seamlessly they are embedded into workflows.** This shift is already visible. Claude 3.7 Sonnet handles both simple and complex reasoning in a unified experience. GPT-powered ChatGPT now autonomously decides when to trigger tools like web search, without user instruction. The direction is clear: **reduce friction, simplify interaction, and let the system manage complexity.** Two recent developments make this transformation tangible: – **Manus** shows how this vision translates into **product design**. It combines research, automation, and visual reasoning, with **automatic orchestration** behind the scenes. Users don't know (or need to know) which model or tool is working; they simply interact with a system that delivers. Manus doesn't even use its own foundation model – it builds on Claude Sonnet – proving that **value lies in experience design, not model ownership.** – **OpenAI's Responses API** brings the same logic to **developer tooling.** It allows teams to build **agentic applications** with reasoning, tool use (search, file retrieval, computer actions), and multi-step workflows – all through a single, unified interface. No more stitching together APIs or managing complexity: the **orchestration is automatic.** The AI race is no longer about building better models; it's about building **better systems.** Models are becoming commodities. The real winners won't be those with the smartest AI, but those who make AI **disappear** – seamlessly integrated, effortlessly powerful. In this future, power won't come from **owning** the model, but from **mastering orchestration and experience design.**

Share this article

Contact Us