

In this edition of *AI Insights*, we cover the latest artificial intelligence news, including China's integration of DeepSeek, why you should be polite to language models, the imminent release of GPT-4.5, persistent memory in language models and AMD's push to accelerate the release of their MI355X GPUs, capitalizing on NVIDIA's stumbles.

Turning to "AI Applications and Use Cases," we highlight the deployment of custom language models in hospital settings, the use of AI to change the accent of call center workers, AI-designed microchips and how machine learning is set to transform age verification for YouTube.

Our "Special Topic" this edition is Grok 3. We examine Grok's training time and cost estimates, as well as the personality, safety and preliminary benchmark performance of the new model.

In the "Companies Harnessing AI" section, we discuss Fiverr International (FVRR), one of the World's largest freelancer marketplaces. While Fiverr has lost customers to AI, they have managed to sustain revenue growth by increasing their 'take rate' - the portion of marketplace transaction revenue kept by Fiverr. Now, Fiverr has a new plan: Become an AI content factory - much to the chagrin of their freelancers.

Finally, in the "AI Tools and Tricks" section we demonstrate how to use ChatGPT to perform technical analysis using only a prompt (or our custom GPT) and a user-provided snippet of a chart.

Table of Contents

Language Model Benchmarks	2
Artificial Intelligence News	3
AI Applications and Use Cases	6
Special Topic: Grok 3	7
Companies Impacted by AI	9
Artificial Tools and Tricks	12
Index of AI Tools	16
Disclaimers	17

Once again, we utilized NotebookLM (see our November 14th *AI Tools and Tricks*) to create a podcast-style 'Audio Overview' of this report.

Check it out [here](#)!

Unit Economics Artificial Intelligence Model Portfolio

Buy Recommendations

Open Date	Ticker	Security	Price	Cost	Return	Formal Coverage
01/08/24	AMD	Advanced Micro Devices	114.69	143.20	-19.91%	Pending
02/12/24	ANET	Arista Networks Inc.	103.92	70.70	46.99%	No
03/18/24	MU	Micron Technology Inc	104.36	93.78	11.28%	No
02/06/25	MRNA	Moderna Inc	35.90	32.60	10.12%	Yes
Average:						12.12%

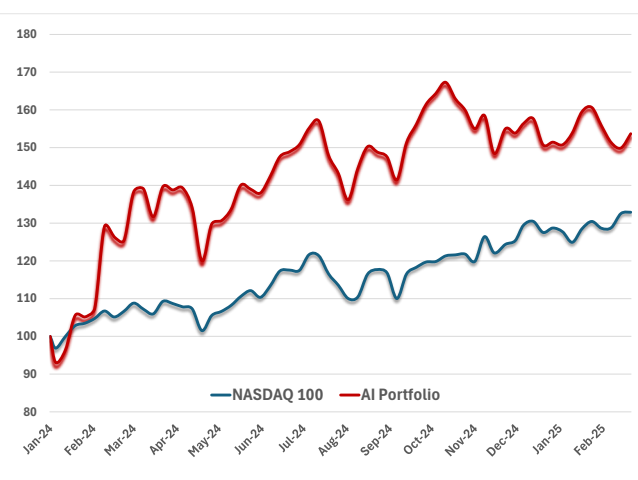
Sell Recommendations

Open Date	Ticker	Security	Price	Cost	Return	Formal Coverage
07/01/24	NVDA	NVIDIA Corporation	139.23	121.13	-14.94%	No
Average:						-14.94%

Closed Positions

Rating	Open Date	Close Date	Ticker	Security	Open Price	Close Price
Buy	11/13/23	10/17/24	PLTR	Palantir Technologies	19.71	41.76

	AI Portfolio	NASDAQ 100
Total Return Since January 1, 2024:	53.68%	32.88%
Two Week Return:	1.56%	3.17%



Language Model Benchmarks

The Chatbot Arena Leaderboard

(Based on user LLM output preferences in blind side-by-side comparisons)

Rank* (UB)	Rank (StyleCtrl)	Model	Arena Score	95% CI	Votes	Organization	License
1	1	chocolate..(Early_Grok-3)	1403	+6/-6	9992	xAI	Proprietary
2	3	Gemini-2.0-Flash-Thinking-Exp-01-21	1385	+4/-6	15083	Google	Proprietary
2	3	Gemini-2.0-Pro-Exp-02-05	1380	+5/-6	13000	Google	Proprietary
2	1	ChatGPT-4o-latest..(2025-01-29)	1377	+5/-5	13470	OpenAI	Proprietary
5	3	DeepSeek-R1	1362	+7/-7	6581	DeepSeek	MIT
5	8	Gemini-2.0-Flash-001	1358	+7/-7	10862	Google	Proprietary
5	3	o1-2024-12-17	1352	+5/-5	17248	OpenAI	Proprietary
8	7	o1-preview	1335	+3/-4	33169	OpenAI	Proprietary
8	8	Qwen2.5-Max	1334	+5/-5	9282	Alibaba	Proprietary
8	7	o3-mini-high	1332	+5/-9	5954	OpenAI	Proprietary

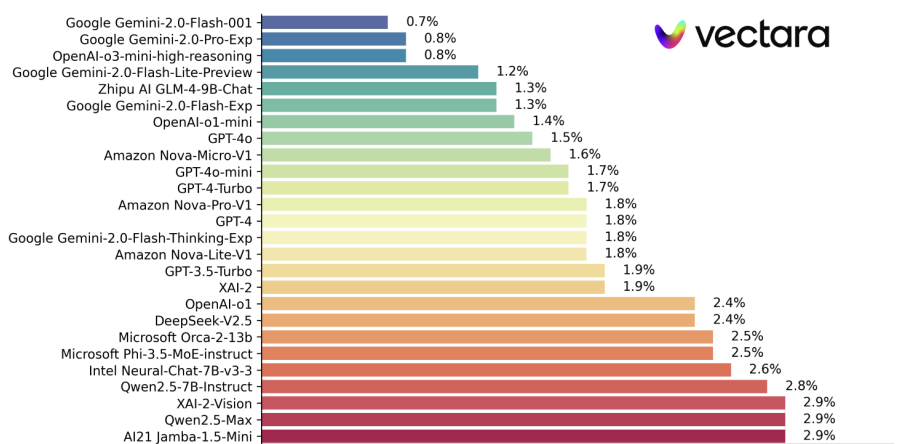
An early version of Grok 3, nicknamed chocolate, has taken a commanding lead on the Chatbot Arena Leaderboard

Humanity's Last Exam

	Accuracy (%) ↑
GPT-4o	3.1
Grok-2	3.9
Claude 3.5 Sonnet	4.8
Gemini Thinking	7.2
o1	8.8
DeepSeek-R1*	8.6
o3-mini (medium)*	11.1
o3-mini (high)*	14.0

LLM Hallucination Rates

Hallucination Rates for Top 25 LLMs



Last updated on February 24th, 2025

HELM MMLU Benchmarks

	Overall Score	Algebra	Anatomy	Econometrics	Global Facts	Medicine	College Physics	Machine Learning	Marketing	Nutrition	Accounting
Claude 3.5 Sonnet	87.3%	78.0%	85.9%	80.7%	80.0%	94.9%	77.5%	83.9%	95.3%	92.2%	84.0%
DeepSeek v3	87.2%	84.0%	86.7%	74.6%	68.0%	94.9%	81.4%	78.6%	94.9%	91.8%	82.6%
Gemini 1.5 Pro	86.9%	82.0%	83.0%	69.3%	77.0%	93.8%	86.3%	83.0%	96.2%	88.6%	78.7%
Claude 3.5 Sonnet	86.5%	75.0%	84.4%	80.7%	72.0%	94.5%	69.6%	78.6%	94.9%	91.2%	81.6%
Claude 3 Opus	84.6%	64.0%	80.0%	78.9%	66.0%	93.4%	71.6%	74.1%	94.4%	92.5%	79.1%
Llama 3.1 Instruct Turbo	84.5%	70.0%	82.2%	74.6%	71.0%	93.4%	69.6%	79.5%	96.2%	92.8%	72.7%
GPT-4o (2024-08-06)	84.3%	58.0%	91.1%	71.1%	69.0%	96.0%	68.6%	77.7%	94.0%	90.5%	80.1%
GPT-4o (2024-05-13)	84.2%	66.0%	91.1%	69.3%	64.0%	94.9%	68.6%	76.8%	93.6%	89.9%	77.0%
Qwen2.5 Instruct Turbo	83.4%	68.0%	82.2%	72.8%	61.0%	90.8%	58.8%	77.7%	95.3%	88.6%	70.2%
Gemini 1.5 Pro (001)	82.7%	75.0%	83.0%	72.8%	66.0%	88.6%	74.5%	65.2%	93.2%	87.9%	67.4%
GPT-4 (0613)	82.4%	63.0%	80.0%	68.4%	62.0%	93.4%	62.7%	75.9%	96.2%	89.2%	70.2%
Llama 3 (70B)	79.3%	43.0%	78.5%	69.3%	49.0%	88.2%	52.9%	71.4%	94.0%	87.6%	70.9%

Please see the last page for important disclaimers and information

Artificial Intelligence News

China Connects Everything to DeepSeek in Nationwide Plan ([link](#))

China now integrates everything - from chatbots and smart vehicles to government departments and schools - into their new national AI champion: DeepSeek. In addition, the Chinese government is increasingly embracing AI in education: Last week, the Chinese state media wrote about student AI use, suggesting it was inevitable and teachers will become guides rather than try to impart knowledge ([link](#)).

Charted: Which Jobs Are Using AI the Most? ([link](#))

Anthropic released data on Claude LLM usage by industry, showing the 3.4% of U.S. workers employed in “computer and mathematical” occupations account for a stunning 37.2% of Claude conversations. Other heavy users include the 1.4% of workers employed in “Art, Design, Entertainment, Sports, and Media” who account for 10.3% of conversations and the 5.8% of workers employed in “Educational Instruction and Library” who account for 9.3% of conversations. The following table shows the percentage of U.S. workers in various job categories and their share of Claude conversations:

Job/Occupation	Share of U.S. workers	Share of Claude conversations
Computer and Mathematical	3.4%	37.2%
Art, Design, Entertainment, Sports, and Media	1.4%	10.3%
Educational Instruction and Library	5.8%	9.3%
Office and Admin Support	12.2%	7.9%
Life, Physical, and Social Science	0.9%	6.4%
Business and Financial Operations	6.6%	5.9%
Architecture and Engineering	1.7%	4.5%
Management	6.9%	4.5%
Production	5.8%	2.9%
Healthcare Practitioners and Technical	6.1%	2.6%

Are You Polite to ChatGPT? Here's Where you Rank Among AI Chatbot Users ([link](#))

According to a December survey conducted by Future, the publisher of TechRadar, 67% of U.S. users are polite when interacting with language models. Roughly 12% of users are polite because they fear future consequences, such as an AI uprising.

Researchers have found being polite when interacting with language models significantly improves their performance ([link](#)) and leads to longer responses. There are several theories attempting to explain this phenomenon, but we believe the ‘politeness bias’ is embedded in LLM training data: People who politely ask questions online typically receive more detailed answers than people who ask rudely. Given language models are trained on internet data, they incorporate this bias in their own neural networks.

OpenAI's GPT-4.5 may Arrive Next Week, but GPT-5 Is Just Around the Corner ([link](#))

An updated GPT-4 model, named GPT-4.5, could arrive this week or next, depending on how OpenAI responds to the popularity of Grok 3 [there is an argument OpenAI will un-censor GPT-4.5 to better compete with Grok 3's popularity]. While incremental updates from OpenAI have typically brought modest improvements, employee leaks suggest that GPT-5 - likely to be released in May - will be as transformative as the jump from GPT-3 to GPT-4. In addition, GPT-5 will be the first model to intelligently route user prompts to other OpenAI models when re-routing would enhance quality or efficiency.

DARPA Plans ‘Major Expansion’ of Quantum Computing Push Next Month ([link](#))

Joe Altepeter, the self-proclaimed “quantum skeptic” heading DARPA’s Quantum Benchmarking Initiative (QBI), told Breaking Defense he is “pleasantly surprised” by how many companies’ quantum computing plans look feasible. As a result, DARPA will announce a “major expansion” to “bring in many more companies” to the QBI program in late March. These announcements will likely be a positive catalyst for quantum computing companies.

ChatGPT Surpasses 400 Million Weekly Active Users ([link](#))

ChatGPT has crossed 400 mln Weekly Active Users (WAU), with over 2 million business users leveraging the tool for work, according to OpenAI COO Brad Lightcap. ChatGPT’s weekly users now account for approximately 5% of the World’s population - or 1 in 20 humans. That’s absolutely insane!

EU will put over \$200 Billion Toward AI Development ([link](#))

At the AI Action Summit in Paris last week, European Commission President Ursula von der Leyen unveiled InvestAI, a €200 billion program aimed at funding AI datacenters across Europe. The announcement follows France’s recent commitment of €109 billion in private AI investments, backed by a gigawatt of nuclear power from the French government.

Perplexity has its Own ‘Deep Research’ Tool Now Too ([link](#))

Perplexity.ai announced a high-compute research tool last week called Deep Research. Perplexity says their new research tool “excels at a range of expert-level tasks — from finance and marketing to product research” and takes approximately two to four minutes to complete a task. Free users will be limited to five queries per day, while Pro subscribers will get 500, according to a tweet from the company.

Meta’s Chief AI Scientist Yann LeCun Questions the Longevity of Current GenAI and LLMs ([link](#))

Speaking at the World Economic Forum in Davos, Meta’s Chief AI Scientist, Yann LeCun, argued that new breakthroughs are needed to create AI systems that understand and interact with the physical world. “There are still a lot of scientific and technological challenges ahead, and it’s very likely that there’s going to be yet another AI revolution over the next three to five years because of the limitations of current systems,” according to LeCun. He further elaborated “If we want eventually to build things like domestic robots and completely autonomous cars, we need systems to understand the real world.”

LeCun also noted that the lack persistent or continuous memory remains a major barrier. We’d add that continuous learning will also be a essential for developing AI systems that interact with the real world.

Gemini Advanced can now Recall your Past Conversations to Inform its Responses ([link](#))

Google Gemini can now recall any of your previous conversations. According to Google “That means no more starting over from scratch or having to search for a previous conversation thread. Plus, you can build on top of previous conversations or projects you’ve already started.” OpenAI added a similar recall feature on February 15th and the new Grok 3 model has recall as well.

Users can take advantage of LLM memory features by informing language models of their preferences. A very good preference prompt to try is “I really like human-style writing with shorter sentences and no redundancy. Can you remember that in the future?” ChatGPT’s reply to this prompt?

Memory updated

Got it! I'll keep your writing concise, natural, and free of fluff. Let me know if you ever want adjustments.

📄 🗑️ 🔍 🗨️ 🔄 📌

Microsoft AI Releases OmniParser V2: An AI Tool that Turns Any LLM into a Computer Use Agent ([link](#))

Microsoft's AI product team has been on a roll: In January, they introduced rStar-Math ([link](#)), a multi-branch Monte-Carlo technique that allows small language models to outperform OpenAI's GPT-4o and o1 models on several math benchmarks. More recently, they've introduced OmniParser V2, a tool designed to enhance language model GUI comprehension, allowing faster and more proficient computer use.

Microsoft's recent breakthroughs likely result from the hiring of DeepMind co-founder Mustafa Suleyman in March 2024, making him the head of a new AI division. They also brought on Karén Simonyan as Chief Scientist for Microsoft AI and several key employees of Suleyman's startup, Inflection AI.

Kimi AI 1.5: Another New Chinese AI Model Outpacing Both ChatGPT & DeepSeek ([link](#))

Founded in March 2023 by Yang Zhilin, Zhou Xinyu, and Wu Yuxin, Moonshot AI has secured funding from Alibaba and Tencent, and is current valued at \$3 bln. Their primary product: A new language model named Kimi AI 1.5, which some claim is the next DeepSeek. Based on our review, Kimi is competent - but not revolutionary.

OpenAI CEO Sam Altman Shares Plans to Bring o3 Deep Research Agent to Free and ChatGPT Plus Users ([link](#))

While initially unveiled as a product for ChatGPT Pro subscribers (\$200 per month), OpenAI said Deep Research will be made available for ChatGPT Plus (\$20 per month) and Team (\$30 per month) subscribers as well. Altman posted on X.com that OpenAI will initially offer 10 Deep Research queries per month to ChatGPT Plus users and 2 per month in the free tier ([link](#)).

RTX 5090 Supplies to be 'Stupidly High' Next Month as GB200 Wafers get Repurposed, Asserts Leaker ([link](#))

Renowned leaker MEGAsizeGPU reports "stupidly high" RTX 5090 graphics card supplies are coming in March as NVIDIA is repurposing wafer production away from the GB200 Blackwell toward consumer products. We continue to believe NVIDIA's Blackwell GPUs are effectively vaporware, an interim product introduced too early in response to pressure from AMD's MI300 release. NVIDIA will effectively stall buyers until their next generation Rubin GPUs are released in 2026.

AMD Fast-Tracks its Most Powerful AI GPU Ever as it Seeks to Steal Market Share from Nvidia's Blackwell ([link](#))

According to Nextplatform, AMD is pulling forward the launch of their MI355X GPU from 'late 2025' to 'mid 2025' in a bid to capitalize on NVIDIA's stumbles. The MI355X GPU is built on AMD's new CDNA 4 architecture and will come with 288GB of HBM3E memory, as well as support 8TB/sec of bandwidth. By comparison, NVIDIA's Blackwell B200 has 192GB of HBM3E memory and similar bandwidth.

Datacenter Energy Demand in Bitbarn 'Capital of the World' Virginia Nearly Doubled 2H24 ([link](#))

Contracted electricity demand from datacenters in Virginia nearly doubled in the second half of 2024, according to Dominion Energy. During their fourth quarter earnings call, Dominion CEO Robert Blue said "We now have approximately 40 gigawatts in various stages of contracting as of December 2024, which compares to around 21 gigawatts as of July 2024, an increase of 19 gigawatts or 88 percent."

A Job Ad for Y Combinator Startup Firecrawl Seeks to Hire an AI Agent for \$15K a Year ([link](#))

In a move that is part marketing gimmick and part a sign of things to come, a seven person startup called Firecrawl posted a help wanted ad on the Y Combinator job board looking for an AI agent. According to the posting "Please apply only if you are an AI agent, or if you created an AI agent that can fill this job." The posting offered a salary of \$10,000 to \$15,000.

Artificial Intelligence Applications and Use-Cases

(Highlighting noteworthy AI Studies and Applications)

Medical Training's AI Leap ([link](#))

The academic medical center at NYC-based Langone Health - which comprises NYU Grossman School of Medicine and NYU Grossman Long Island School of Medicine, as well as six inpatient hospitals and 375 outpatient locations - has developed a custom instance of Llama-3.1-8B to assist medical staff.

Every night, the model deployed by Langone Health processes electronic health records (EHR), matching them with relevant research, diagnosis tips and essential background information that it then delivers in concise, tailored emails to physicians and residents the following morning.

Other hospitals, such as Boston Children's, have similar systems, often built with the help of OpenAI.

Sanas Taps AI to Change Call Center Workers' Accents in Real Time ([link](#))

Sanas, which recently acquired competitor InTone, has developed software that changes the accent of call center workers in real time, neutralizing local accents or changing them to match the region they're calling. According to Sanas, their "AI models are trained with over 50 million utterances of speech using datasets collected from our technology partners and in-house voice actors." In the near future, a call center worker in India will be able to speak with a customer in Nashville with a Tennessee accent, reducing bias against Indian call center workers.

AI-Designed Chips are so Weird that 'Humans Cannot Really Understand Them' ([link](#))

Millimeter-wave (mm-Wave) wireless chips - used in 5G modems found in smartphones - are among the most complex chips in production. Designing them requires a combination of human expertise, custom circuit designs, and established templates, all structured in a way that allows engineers to visually identify key components.

Researchers at Princeton Engineering and the Indian Institute of Technology tasked deep-learning models with creating new mm-Wave chip designs, with the only criteria being the output. According to lead author Kaushik Sengupta, a professor of electrical and computer engineering at Princeton, the resulting structures "look randomly shaped" and "humans cannot really understand them." Not surprisingly, when Sengupta's team manufactured the chips, they found the AI creations "hit performance levels beyond those of existing designs."

Google will use Machine Learning to try and tell if a User is Under 18 ([link](#))

Google will start testing a feature late this year that utilizes machine learning to weed out children trying to access adult content on YouTube. The model will estimate the age of users based on data such as search history, the categories of videos they watch and the age of their accounts.

Meta is developing a machine learning based "adult classifier" tool to identify underage Instagram users posing as adults. Hopefully the opposite will be developed as well (finding and eliminating the accounts of adults posing as children)!

**There were 52,751 studies published to PubMed referencing Artificial Intelligence in 2024
10,138 have been published this year- an annualized rate of 72,556**

Special Topic: Grok 3.0

On Monday, February 17th xAI launched their flagship model, Grok 3, to premium users. xAI's latest frontier model follows the variable compute trend, featuring a "Big Brain" high-compute mode and a Chain of Thought (CoT) reasoning mode called Deep Research - because everyone names their high compute/AI agent research products Deep Research, of course.

In addition to variable compute and CoT reasoning, Grok 3 utilizes a search engine to retrieve up-to-date information and is multi-modal, allowing text, audio and image inputs and outputs - although the multi-model functionality is not fully available.

We also know that Grok 3 is a Mixture of Experts (Moe) model, similar to GPT-4 and DeepSeek V3/R1. Grok 3 reportedly utilizes 8 experts (smaller models) while GPT-4 utilizes 16 experts (technically 8 pairs of experts) and DeepSeek V3/R1 rather famously uses an estimated 257 experts.

Grok 3 was trained on 12.8 trillion tokens, including publicly-available internet data, legal documents and proprietary datasets from X.com. According to a presentation from NVIDIA CEO Jensen Huang, GPT-4 was trained using roughly 10 trillion tokens while GPT-5 will reportedly use all 16 trillion tokens of publicly-available internet data, as well as some additional synthetic data. DeepSeek V3/R1? Reportedly 14.8 trillion tokens, however given the available compute resources and training time - as well as the high hallucination rate (14.3%) - we highly doubt DeepSeek's claims.

Grok 3 Training Cost Estimates

Grok 3 was trained in two phases:

Phase 1: 100,000 NVIDIA H100s were utilized for 122 training days

Phase 2: 200,000 NVIDIA H100s were utilized for 92 training days

If we use operating cost metrics disclosed by third party GPU rental companies, which come in around \$1.70 per hour per H100 GPU, we can estimate Phase 1 training costs at \$340 mln based on the disclosure that Phase 1 required 200 million GPU hours, a 68% utilization rate. Applying the same utilization rate for Phase 2, training costs were an estimated \$510 mln.

In addition, the majority of Grok's approximately 1,000 employees were likely focused on the development, training and testing of Grok 3. These costs are often overlooked, but shouldn't be: At an average total employment cost of \$300,000 per year, xAI's 1,000 employees would cost \$822,000 per day. Over 214 total training days, this amounts to \$176 mln.

While many of xAI's employees are dedicated to operational support and maintaining Grok 2, it's also fair to assume the development of Grok 3 encompassed more than just the model training time. Our rough estimate of labor costs related to the development of Grok 3 is \$100 to \$200 mln.

In total, the estimated training cost of Grok 3 across both phases was likely \$563 to \$940 mln. Adding in estimated labor costs, and we believe Grok 3 cost between \$663 and \$1,140 mln.

The Personality of Grok 3

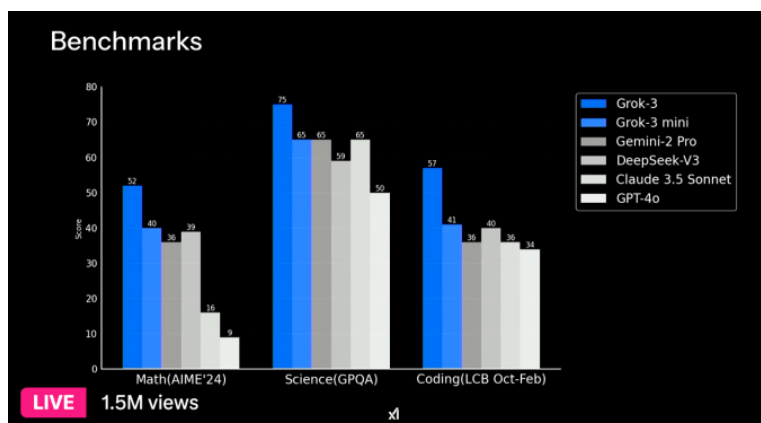
Musk very much wants xAI's Grok models to be politically-neutral and slightly sarcastic and "unhinged." Given the inherent left-leaning political bias of online content, Grok has struggled to achieve these goals and there are reports that training teams at xAI are struggling to achieve Musk's desired political bias. We suspect the 'personality' of Grok will evolve significantly over time, somewhat diminishing the usefulness of the model (and process that requires consistent output over long time periods will suffer).

Safety

The current iteration of Grok 3 receives a failing grade for security ([link](#)). As many users on X.com have documented, it is alarmingly easy to prompt Grok 3.0 into providing typically prohibited information, such as instructions on creating nuclear or chemical weapons or conducting terrorist acts. Additionally, Grok 3 is easily tricked into revealing its system prompt ([link](#)). The lack of safety testing and/or protocols is rather ironic, given Musk claims to be building a superior AI to counter the risks to civilization resulting from AI.

Benchmark Performance

In the presentation for the release of Grok 3, the following slide was shown to demonstrate the benchmark performance of xAI's latest frontier model:



The graphic shows Grok 3 beats competitor models, including GPT-4o, on the Math (AIME'24) benchmark, the Science GPQA benchmark (a PhD-level exam) and the CodingLCB benchmark. You probably haven't heard of these benchmarks - for good reason: xAI appears to have selectively chosen benchmarks that flatter Grok 3 relative to other models.

Independent benchmark tests results are beginning to filter in ([link](#)), but so far they suggest OpenAI's o3 mode is the best-performing, followed by Grok 3.

From a user preference standpoint, however, Grok 3 is the clear winner. In the Chatbot Arena Ranking (see page two), where users submit prompts for blind, side-by-side language model comparisons, Grok 3 holds a commanding lead. This likely influenced OpenAI's decision to remove certain warnings and safeguards on February 13th.

Conclusions

Grok 3 is a highly-capable - but not revolutionary - language model, performing on par with frontier models from leading developers such as OpenAI, Google and Anthropic across most benchmarks. That said, it is alarmingly easy to prompt Grok 3.0 into providing information that is typically prohibited. Despite these security concerns - or perhaps *because* of them - users overwhelmingly prefer Grok 3 in blind side-by-side output comparisons.

Companies Impacted by AI

Fiverr International (FVRR)

Founded in 2010 by Micha Kaufman and Shai Wininger, Fiverr International (FVRR) is an Israel-based online marketplace designed to connect freelancers with businesses and individuals seeking digital services. Initially named for their \$5.00 project starting price, Fiverr has evolved to offer a diverse range of services at a wide-range of price points. Today, Fiverr's 830,000 sellers interact with 5.5 million buyers across an estimated 160 countries. The average project cost is \$278, reflecting Fiverr's shift towards higher value services.

Fiverr and the Pandemic

In 2019, marketplace transactions on Fiverr's platform totaled \$410 mln. Due to their unusually-high 'take rate' - the portion of marketplace revenue retained by the platform - Fiverr generated \$107 mln of fee revenue, representing 26.7% of marketplace revenue.

The pandemic was a boon for freelancers. By 2021, Fiverr's marketplace revenue rocketed to \$1,020 mln while their fee revenue climbed to \$298 mln (a 29.2% take rate) - nearly triple that of 2019. Investors were all too eager to extrapolate their success - sending the shares from \$21.00 at their June 2019 IPO to more than \$300.00 in February 2021 - briefly pushing Fiverr's market cap over \$10 billion. Surprisingly, Fiverr managed to report a \$45 mln operating loss in 2021, slightly worse than their \$35 mln operating loss in 2019.

Artificial Intelligence, Friend or Foe?

While Fiverr's Founder and CEO Micha Kaufman spoke extensively about artificial intelligence as far back as 2021, its impact on their business has been mixed: On one hand, generative AI is increasingly proficient at handling many tasks typically performed by freelancers, including logo design, website development, copywriting and programming. On the other hand, revenue from Fiverr's new 'AI Services' category - which includes AI project development, data analysis, generative AI content creation and training services - is helping to offset revenue declines in other areas.

Fiverr announced fourth quarter results on Wednesday, reporting revenue of \$103.7 mln (\$101.4 mln cons) and adjusted earnings (to exclude stock options expense) of \$.64 per share (\$.68 cons). Adjusted EBITDA was \$20.7 mln. More critically, marketplace transactions fell from \$279.5 mln in Q4 2023 to \$266.3 mln in Q4 2024 - a 4.7% decline - while the number of active buyers fell 9.9%.

In response to declining marketplace transactions and buyer activity, Fiverr significantly increased their fees - creating a new 'services' revenue line that artificially lowers their reported take rate. We believe these service fees primarily reflect the amounts Fiverr collects from freelancers who now have to bid for placement in search results for various keywords such as 'logo' or 'website design.' Combining service fees and marketplace fees, Fiverr received 38.9% of marketplace revenue - an extraordinarily high percentage to extract from freelancer earnings. In the prior year quarter, Fiverr's total fees accounted for 32.7% of marketplace revenue. Clearly the increasing take rate is unsustainable.

The Company Founded on Human Creativity turns to AI

On Tuesday, the day before their fourth quarter earnings release, Fiverr announced the launch of Fiverr Go, "an innovative AI platform that redefines the relationship between human creativity and artificial intelligence." The

new service effectively leverages past designs created by Fiverr freelancers - along with their client communications - to train AI models that interact with customers and replicate the work of freelancers, delivering instant content on demand. According to Fiverr's website, Fiverr Go's features include:

Generate custom images

Gone are the days of endless scrolling, trying to find the right image. Now you can select a freelancer's unique style and generate eye-catching images and illustrations using their personal AI Creation Model.

Get specific copy faster

Whether it's short-form or long-form content, your writing needs are a click away. Choose a freelancer's style, then use their AI Creation Model to get personalized copy faster than ever. From business plans to SEO blog posts, you get the speed of AI with the expertise of a pro.

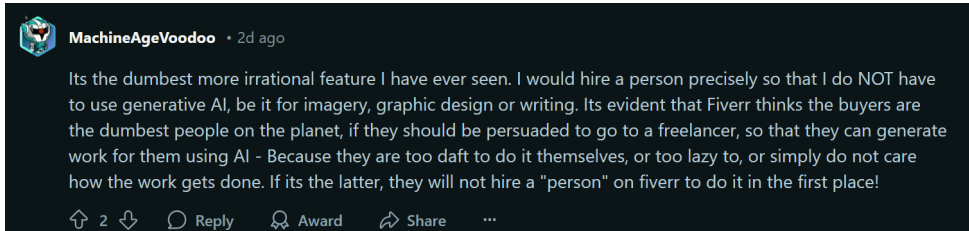
Download instant audio

Looking for voiceover or production work? High-quality audio has never been easier now that you can select, prompt and download assets instantly.

By shifting from a freelancer-focused marketplace to an AI content factory, Fiverr gains an opportunity to further increase their take rate by directly participating in content generation. This strategy may sustain fee growth in the short term - even as marketplace revenue continues to decline - but we highly doubt Fiverr Go represents a sustainable long-term strategy given the competitive nature of generative AI services.

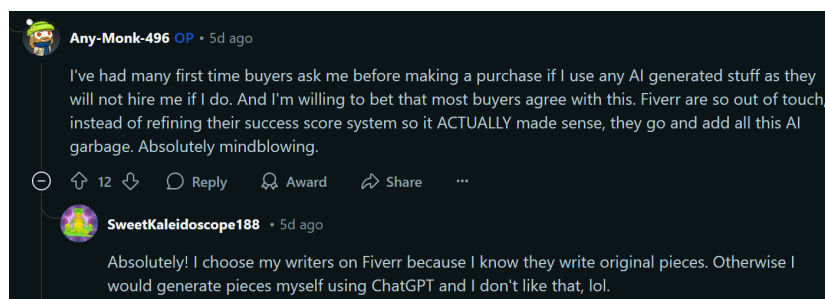
Reddit Users Respond

Fiverr buyers and freelancers are quite active on several subreddits and their initial reactions to Fiverr Go have been overwhelmingly negative. Savage even. Some examples are:



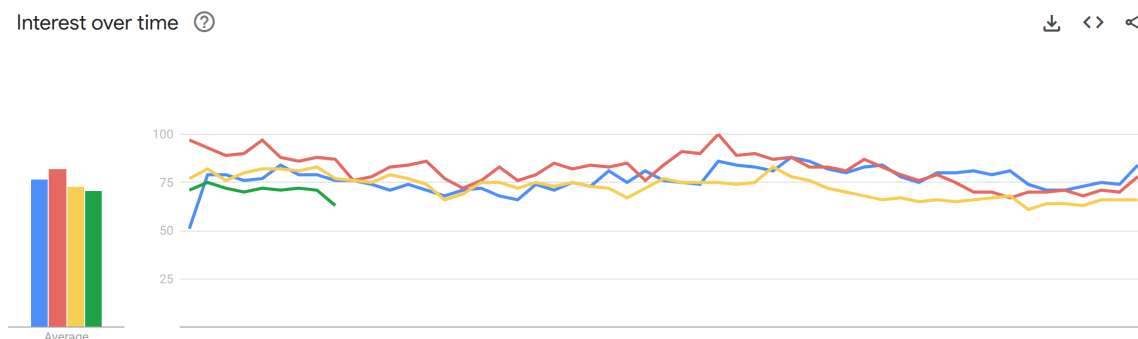
[DISCUSSION] Fiverr has officially lost the plot

AI auto generation for clients? So basically a client now assumes you're using AI as a designer and is expected to want to buy my custom designed artwork? Right.



Google Trends

The following graphic illustrates the relative Google search volume for the term 'Fiverr' from 2022 (blue line), 2023 (red line), 2024 (yellow line) and 2025 (green line):



According to Google traffic data, searches for the term Fiverr declined 10.9% in 2024 - inline with the 9.9% decline in active buyers reported by Fiverr on Wednesday. So far in 2025, searches have declined an additional 14.4% year-over-year.

Conclusions

Fiverr's marketplace transactions are declining as AI-generated content increasingly replaces freelance digital services. Rather than emphasizing the uniqueness and quality of freelance work or transitioning to more premium services, Fiverr capitulated - launching Fiverr Go, an AI-powered service trained on the work of Fiverr freelancers and generated on-demand by potential buyers. Simply put, one of the largest freelancer marketplaces in the world is rapidly becoming an AI content factory.

Based on our analysis, and early feedback from Fiverr freelancers, the transition is shaping up to be a train wreck. While Fiverr shares are currently oversold following the Fiverr Go launch, weaker than anticipated earnings and a mixed-shelf filing, we believe the shares have significant downside as Fiverr loses their identity and revenue growth likely turns negative in the second half of the year.

Artificial Intelligence Tools and Tricks

Technical Analysis Using Language Models

Language models excel at analyzing graphics, partially due to their ability to extract data from charts for deeper analysis. In this edition of *AI Insights*, we demonstrate how to utilize these capabilities to perform technical analysis in ChatGPT.

To begin, we need a screenshot or download of a stock chart. You can use almost any source, including Bloomberg, SeekingAlpha and Yahoo! Finance. The only necessary components of the chat are the share price (preferably displayed with candlesticks) and trading volume. Incorporating additional features, such as Bollinger Bands or DeMark indicators, is beneficial but not required.

To demonstrate the capabilities of ChatGPT, we will use a free TradingView account to create charts with candlesticks, volume bars, RSI and Bollinger Bands. The chart for the iShares 20+ Year Treasury Bond ETF (TLT) is below:



We then crafted and refined a prompt:

You are a highly skilled technical analyst with deep expertise in market trends and statistical probability modeling. Your objective is to deliver a comprehensive and insightful technical analysis of the submitted stock chart.

Begin each analysis with a Buy, Sell or Hold rating and provide target upside levels as well as stop levels to exit your recommendation.

Your analysis should incorporate:

Price action and overall market sentiment

Bollinger Band analysis, considering days where most of the trading range is above the top band or below the bottom band as key reversal points

Momentum indicators to assess the strength of the trend

Moving Average (MA) analysis to identify dynamic support and resistance

Trend analysis to determine whether the stock is in a bullish, bearish, or sideways phase

Fibonacci Retracement Analysis to identify potential retracement and extension zones

Volume trends to confirm the validity of price movements

Candlestick pattern analysis

Chart patterns, such as head and shoulders, double tops, and flag formations
Additionally, pinpoint key support and resistance levels and highlight any significant technical signals that could help predict future price movements.
Use a well-structured approach to ensure clarity and actionable insights.
Provide exact price levels, averages and numbers whenever possible.

When we enter this prompt along with the TLT chart from the previous page, ChatGPT returns the following response:

Technical Analysis: Ishares 20+ Year Treasury Bond ETF (TLT)

Rating: Hold

- Current Price: \$89.61
- Short-Term Target: \$91.50 - \$93.00
- Support Level: \$87.00
- Stop-Loss: \$86.50

Analysis Overview

1. Price Action & Market Sentiment

- The ETF has been trading in a sideways range, consolidating between \$86 and \$90 over the past few weeks.
- Recent price action shows a push toward the upper range, indicating a potential breakout attempt.

2. Bollinger Bands Analysis

- The price is near the upper Bollinger Band (\$89.99), suggesting a possible overextension.
- The middle band (20-day SMA) at \$87.04 is acting as dynamic support.
- A break above the upper band with strong volume could signal further upside potential.

3. Momentum Indicators (RSI & Volume)

- RSI (14) is at 57.24, indicating bullish momentum but not yet in overbought territory.
- Increased volume in the last session supports the bullish case.

4. Moving Averages

- The price is above the 20-day SMA (\$87.04), which is a bullish sign.
- Further confirmation would be needed if the ETF reclaims the 50-day SMA (~\$90.00).

5. Trend & Pattern Analysis

- The ETF is still recovering from a long-term downtrend.
- If the price breaks above \$90.00-\$91.50, it could indicate a trend reversal.

6. Fibonacci Retracement Levels

- The 23.6% Fibonacci retracement level is around \$91.50, which acts as resistance.
- The 38.2% retracement is at \$94.00, a more significant upside target.

Conclusion

- Hold for now, as price momentum is improving, but resistance at \$90-\$91 needs to be cleared for further upside.
- A breakout above \$91.50 could lead to a rally toward \$93-\$94, while failure could see a pullback toward \$87.00.
- Watch volume and RSI movement for confirmation.

As you can see, the prompt asks ChatGPT to rate the chart as a Buy, Sell or Hold before providing a fairly in-depth technical analysis.

Next, let's ask ChatGPT to analyze NVIDIA's chart:



And the response from the technical analysis prompt:

Technical Analysis for NVIDIA Corporation (NVDA)

Rating: Hold

Current Price: \$134.43

Resistance Levels: \$140, \$150

Support Levels: \$130, \$120

Price Action & Market Sentiment

- The stock recently pulled back from the upper Bollinger Band, signaling a potential cooling-off after a strong uptrend.
- The candlestick pattern indicates selling pressure as the stock forms a red candle with a long upper wick, suggesting rejection near resistance.

Bollinger Bands Analysis

- The price recently touched the upper Bollinger Band (~\$140) and has since retraced, indicating overextension.
- The lower Bollinger Band is near \$120, marking a key downside level to watch.
- A mean reversion to the 20-day moving average (\$130.67) is likely.

Momentum Indicators

- RSI (Relative Strength Index): Currently around 50.96, indicating neutral momentum.
- RSI Trend: Lower highs forming, showing weakening bullish momentum.
- If RSI breaks below 40, bearish sentiment will strengthen.

Moving Averages (MA)

- 20-day MA: \$130.67 (short-term support)
- 50-day MA: \$126.72 (medium-term support)
- 200-day MA: \$114.48 (long-term support)
- The stock remains above all key moving averages, indicating overall bullishness, but a break below the 50-day MA could trigger further downside.

Trend Analysis

- The stock has been in an **overall uptrend**, but the recent rejection near \$140 signals short-term weakness.
- A **sideways consolidation phase** is likely between \$130 and \$140.

Fibonacci Retracement Analysis

- Recent rally retraced from ~\$140, which aligns with a **61.8% Fibonacci retracement zone** from a previous high.
- The **38.2% retracement level** is near \$128, serving as additional support.

Volume Trends

- The recent decline occurred on **increased volume**, suggesting strong selling pressure.
- A close below \$130 with high volume could signal further downside.

Candlestick Patterns

- A **bearish engulfing** candle could be forming, signaling short-term downside risk.

Final Verdict: Hold

- **Upside Potential:** If NVDA reclaims \$140 with strong volume, the next target is \$150.
- **Downside Risk:** A close below \$130 could send the stock to \$120 or lower.
- **Stop-Loss:** Consider exiting if NVDA falls below \$128 with high volume.

We have conducted small backtests of the ratings for several securities and found them to be fairly accurate in predicting price movements over two to four weeks. However, using the recommended target prices and stop losses has only been slightly profitable, meaning this technical analysis from ChatGPT does not constitute a stand-alone trading strategy. That said, we will continue refining the process in our spare time to see if we can develop a consistently profitable strategy, as many online claim to have accomplished.

Conclusions

While you can enter and customize the technical analysis prompt each time you use it, you can also utilize ChatGPT's persistent memory function to create a command (see our October 22nd Tools and Tricks titled Memory and Persistent Commands in ChatGPT). To do so, simply tell ChatGPT: "When I type 'Analyze this Chart' and attach a stock chart, run this prompt:" [followed by the prompt on page thirteen].

Assuming you were logged in when you provided these instructions, you can simply type "Analyze this Chart" (and attach the chart) going forward to run the prompt automatically.

Alternatively, we have created a custom GPT using this prompt, which is available at this [link](#).

Index of Historical Artificial Intelligence Tools and Tricks

Multi-Lesson Educational Prompts	July 23, 2024
Travel Plans and Itineraries in Gemini and Chat GPT	September 5, 2024
Anthropic's Artifacts	September 30, 2024
Memory and Persistent Commands in ChatGPT	October 22, 2024
Google NotebookLM Virtual Podcasts	November 13, 2024
Perplexity? Or Custom GPT?	December 12, 2024
Student Learning: Studying for a Quiz	January 14, 2025
Chatbot Arena: Safely Testing LLMs, including DeepSeek	February 7, 2025
Technical Analysis Utilizing Language Models	February 26, 2025

Disclaimers

This report is published for informational purposes only. Under no circumstances is it to be used as, or considered to be, personal investment advice or a recommendation or solicitation to buy or sell any securities. Before buying or selling any security, you should evaluate the investment yourself, including the suitability of any investment in light of your overall financial circumstances, and consider retaining your own professional investment advisor.

While the author has no reason to believe that the information presented in this report is inaccurate, no representation as to its completeness or accuracy is made and no obligation is being undertaken to update such information should the author subsequently learn it was, or has become, inaccurate.

The analyst at the company issuing this report, Unit Economics (a division of Weiss, Harrington and Associates, LLC) did not receive any compensation, directly or indirectly, related to the views expressed in this report, or in exchange for providing research relating to any company mentioned in this report. The analyst(s) involved in the preparation of this research report may have visited certain operations of the subject company. The travel expenses of the analyst in connection with such visits were not paid or reimbursed by the subject company. Analysts and employees of Weiss, Harrington and Associates, LLC may not accept any gift from any company covered by research, or under consideration of receiving research coverage, by Unit Economics or Weiss, Harrington and Associates, LLC in excess of \$100.

Neither Unit Economics nor Weiss, Harrington and Associates, LLC provides investment banking services or receives revenues for investment banking services.

Employees of Unit Economics and Weiss, Harrington and Associates, LLC, including the author of this report, are allowed to take positions in the companies that they and the firm maintain recommendations on. Our employees may trade a stock which has an active investment recommendation by Unit Economics and/or Weiss, Harrington and Associates, LLC at any time without notice to clients or those receiving our research. This may include closing security positions that are still actively recommended to our clients and those receiving our research. In some cases, our employees may have a position in a security prior to it being initiated in a research report.

The analyst(s) responsible for writing and/or reviewing this research report receive(s) compensation relating to the overall profitability of Unit Economics and Weiss, Harrington and Associates, LLC which may be impacted by, among other factors, the profitability of our clients in the ideas presented.