

AI News

Dear readers,

As we occupy that blissful space between Christmas and the New Year, we bring you this special festive edition of AI News Weekly Magazine. Combining the latest in AI technology and advancements, this issue aims to be your Yuletide companion, your window into the AI world.

Kicking off this edition, we delve into the fascinating findings of a study demonstrating that OpenAI's ChatGPT offers more detailed responses when tipped, albeit hypothetically. We examine its rapid growth and the intriguing future it carves. Microsoft's Copilot is set for a new upgrade, as it preps to integrate with OpenAI's GPT-4 Turbo and DALL-E 3, promising a host of new features.

AI news doesn't get more exciting than this. On one hand, Apple moves into the generative AI space with the release of MLX, its machine learning framework, and on the other, Google's Gemini and DeepMind's FunSearch showcase remarkable AI wizardry, in spite of the Gemini demo controversy. In response, OpenAI enthusiasts have proven their model's capabilities by recreating the contested demo without any editing cheats – an impressive achievement.

More on the AI front, we introduce improved coding aids OSS-Instruct and Magicoder developed by researchers from the University of Illinois and Tsinghua University. Additionally, Mistral AI's Mixtral 8x7B and Ollama's v0.1.16, are stirring the pool with their open and uncensored models, nudging AI towards democratization.

Our articles delve deeper into the ways AI is emerging in new fields such as astronomy, where AI systems can now predict the age of a supernova milliseconds after detection. In academia, we introduce the first fully open-source Large Language Models (LLMs) in LLM360. Simultaneously, the world of chemistry witnesses a surge of AI-driven innovations.

This edition also includes news of Stability AI's new membership structure, and the full-stack application AnythingLLM that exemplifies the transcending boundaries of document conversion for LLMs.

As we reflect on a year that has seen some truly breakthrough

innovations and advancements in AI, we also expect next year to offer an exciting panorama of equally transformative developments.

Before we conclude, our deepest apologies for the less consistent publication schedule this past December. We assure you of a return to our regular routine starting January 2024. Your patience and understanding have been invaluable, and we are grateful for your continued support.

On behalf of the AI News Magazine, we would like to extend our warmest greetings for the season and the New Year. Here's to ringing in a year that brings forth a trove of new innovations, discoveries, and opportunities in the AI world.

To a brighter, AI-powered future!

ChatGPT will provide more detailed and accurate responses if you pretend to tip it, according to a new study

2023-12-04

A recent study reveals that AI chatbot, ChatGPT, provides more detailed responses when users pretend to offer a tip. The experiment found that the chatbot's responses were more comprehensive when incentivized, although it refused to accept the tip, stating user satisfaction was its reward. The study highlights how the training material for AI models can influence their responses and reasoning.



[Read more at Windows Central...](#)

ChatGPT is winning the future — but what future is that?

2023-12-05

ChatGPT, launched by OpenAI a year ago, has become the fastest-growing consumer technology in history, impacting every corner of the tech industry. Despite its simplicity and reliability issues, it has amassed millions of users and become a billion-dollar business. The AI revolution has sparked debates about its potential impact on society, economy, and the tech industry. While AI technology still has room for improvement, its rapid development and adoption suggest a transformative potential akin to the internet, social media, and smartphones.



[Read more at The Verge...](#)

Microsoft upgrades Copilot with OpenAI's GPT-4 Turbo and DALL-E 3

2023-12-06

Microsoft's Copilot AI chatbot is set to integrate with OpenAI's GPT-4 Turbo and DALL-E 3, enhancing its functionality. The GPT-4 Turbo allows for larger data input, leading to more meaningful responses, while DALL-E 3 generates higher quality images. Other features include a rewrite menu in the Inline Compose tool and a new feature set for coders called Code Interpreter. Bing search is also getting an upgrade powered by GPT-4 for optimized search results.



[Read more at Engadget...](#)

Apple joins AI fray with release of model framework

2023-12-06

Apple has released MLX, a machine learning framework, and MLX Data, a deep learning model library, designed to run on Apple Silicon. These tools, inspired by frameworks like PyTorch and Jax, aim to bring generative AI applications to MacBooks. They offer a shared memory feature for efficient task execution on supported devices. The move marks a shift in Apple's conservative AI approach, potentially positioning it alongside competitors like Microsoft and Google in the generative AI space.



[Read more at The Verge...](#)

Google's Gemini Looks Remarkable, But It's Still Behind OpenAI

2023-12-11

Google, in a bid to outpace OpenAI, has unveiled Gemini, a new AI model capable of identifying magic tricks and acing accountancy exams. Despite the impressive demo video, Google remains technically behind OpenAI.



[Read more at Bloomberg.com...](#)

Google's best Gemini demo was faked | TechCrunch

2023-12-11

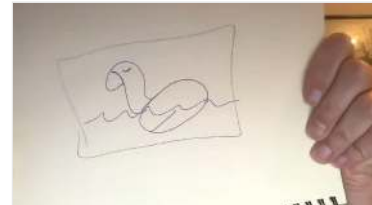
Google's Gemini AI model demo video, which showcases the model's multimodal understanding capabilities, has been revealed to be largely staged. The video, which shows the AI responding to various visual and verbal prompts, was created using carefully selected still images and text prompts, rather than live interactions. This has raised concerns about the accuracy and integrity of Google's AI demonstrations.

[Read more at TechCrunch...](#)



ChatGPT-4V user remade Google's deceptive Gemini AI demo without editing cheats

2023-12-11



Tech enthusiast Greg Technology has recreated Google's controversial Gemini AI demo using OpenAI's GPT-4V. The original Google demo was criticized for not being recorded in real-time and using post-production techniques. In contrast, Greg's version uses real-time voice and vision prompts, highlighting the latency between user prompts and AI responses. Greg's demo code is available on GitHub.

[Read more at Tom's Hardware...](#)

OSS-Instruct gives open-source LLMs a huge boost in coding

2023-12-11



Researchers from the University of Illinois and Tsinghua University have developed OSS-Instruct, a technique that uses AI models like GPT-4 to generate diverse coding examples for fine-tuning open-source models. Using this method, they created Magicoder, a suite of language models that outperforms most open-source models and approaches the performance of GPT-3.5. This advancement could democratize high-quality coding assistance.

[Read more at TechTalks – Technology solving problems... and creating new ones...](#)

Mixtral of experts

2023-12-11

Mistral AI has released Mixtral 8x7B, a high-quality sparse mixture of experts model (SMoE) that outperforms Llama 2 70B on most benchmarks and matches or exceeds GPT3.5.

The model, which handles multiple languages and shows strong performance in code generation, is the strongest open-weight model with a permissive license and offers the best cost/performance trade-offs. It can also be fine-tuned into an instruction-following model, scoring 8.3 on MT-Bench.

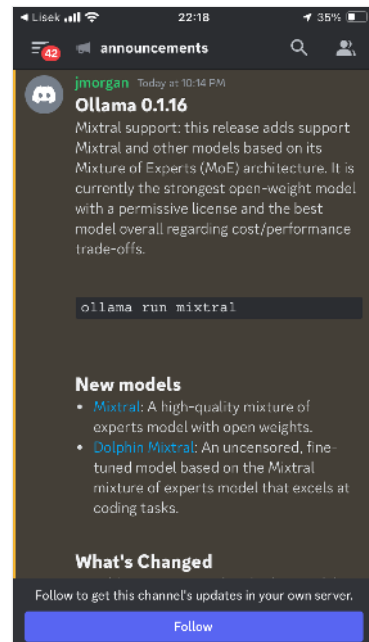
[Read more...](#)

	Llama 2 70B	GPT-3.5	Mixtral 8x7B
MMLU (0-50)	69.5%	70.0%	70.6%
Hellaswag (0-100)	87.1%	85.5%	86.7%
ARC Challenge (0-100)	85.1%	85.2%	85.8%
Winogrande (0-100)	83.2%	81.6%	81.2%
MBPP (0-100)	49.8%	52.2%	60.7%
CodeContest (0-100)	53.6%	57.1%	58.4%
MT Bench (0-10)	6.36	8.32	8.20

Ollama 0.1.16 adds Mixtral support

2023-12-14

The latest release of Ollama, version v0.1.16, adds support for Mixtral and other models based on the Mixture of Experts (MoE) architecture. This update includes new models like Mixtral, a high-quality mixture of experts model, and Dolphin Mixtral, an uncensored model optimized for coding tasks. It's important to note that these models require at least 48GB of memory. Additionally, a fix was implemented for an issue related to `load_duration` in the `/api/generate` response. For full details, visit the [GitHub release page](#).



dolphin-2.5-mixtral-8x7b

2023-12-15



 cognitivecomputations

/dolphin-2.5-mixtral-8x7b

 huggingface.co

The Dolphin 2.5 Mixtral 8x7b AI model, sponsored by Convai, excels at coding tasks. The latest update removes Samantha and WizardLM, introducing Synthia, OpenHermes, and PureDove, along with new datasets. Despite being uncensored and unbiased, users are advised to implement their own alignment layer for ethical compliance. The model uses the ChatML prompt format and requires three days for training. Future plans include enhanced chat use-cases, structured output, and role-playing.

[Read more...](#)

DeepMind AI outdoes human mathematicians on unsolved problem

2023-12-15

Google's DeepMind has developed an AI system, FunSearch, that uses large language models to generate new solutions to mathematical problems. The system was tested on the 'cap set problem', a complex version of the card game Set, and was able to improve on existing solutions. Unlike other AI applications, FunSearch allows users to see and learn from the successful programs it creates.



[Read more...](#)

OpenAI Demos a Control Method for Superintelligent AI

2023-12-15

OpenAI is developing a superalignment program to control superintelligent AI systems and align them with human goals. The team is using a weak AI model to supervise a stronger one, with the aim of the stronger model learning from the weaker one's mistakes. The first experiment showed promising results, particularly in natural language processing tasks. OpenAI is also offering \$10 million in grants for work on various alignment approaches, encouraging the wider research community to contribute to this ambitious project.



[Read more at IEEE Spectrum...](#)

Introducing the Stability AI Membership — Stability AI

2023-12-15

Stability AI introduces its new membership structure, offering commercial rights to its open AI models. The three-tiered membership aims to make generative AI technology accessible to all, from individual enthusiasts to large corporations. The tiers include Non-Commercial for personal use, Professional for creators and startups, and Enterprise for larger companies. The membership aims to balance competitiveness and openness in AI technologies, while providing exclusive benefits and commercialization opportunities.

[Read more at Stability AI...](#)



anything-llm: Open-source ChatGPT experience for your data

2023-12-15

AnythingLLM is a full-stack application that allows users to convert any document into a context for language learning models (LLMs) to reference during chats. The application supports multiple LLMs, embedding models, and vector databases. It also offers features like multi-user instance support, permissioning, and document management from a simple UI. The application can be run locally or hosted remotely, providing an efficient and customizable solution for enterprise-ready document chatbot needs.

[Read more at GitHub...](#)



AI predicts age of newly discovered supernovae within milliseconds

2023-12-16

Artificial intelligence can now predict the age of supernovae within milliseconds of detection, a breakthrough that could significantly aid projects like the Vera C. Rubin Observatory's Legacy Survey of Space and Time. The survey, starting in 2025, could spot over 10 million cosmic events nightly, potentially increasing supernovae discoveries by 100 times.



[Read more at New Scientist...](#)

Meet LLM360: The First Fully Open-Source and Transparent Large Language Models (LLMs)

2023-12-18

Researchers from Petuum, MBZUAI, USC, CMU, UIUC, and UCSD have introduced LLM360, an initiative to fully open-source Large Language Models (LLMs). The project aims to make the end-to-end LLM training process transparent and reproducible, providing all training code, data, model checkpoints, and intermediate results to the community. The team has released two 7B parameter LLMs, AMBER and CRYSTALCODER, along with their training code, data, and analyses.



[Read more at MarkTechPost...](#)

Uncensored Models

2023-12-19



Artificial Intelligence and Machine Learning

Exploring the possibilities



The article details the process of creating uncensored AI models, arguing for their necessity due to cultural diversity, valid use cases, ownership rights, and composability. It provides a step-by-step guide on uncensoring the WizardLM model, highlighting the importance of removing refusals and biased answers from the dataset. The author emphasizes the user's responsibility in handling the output of these models.

[Read more at Cognitive Computations...](#)

Autonomous chemical research with large language models

2023-12-22

The Nature article explores the advancements in large language models (LLMs) and their applications in various fields, including chemistry, biology, and coding. It discusses the development of models like LaMDA, LLaMA, and PaLM, and their role in predicting protein structures and text generation. The article also highlights the use of AI and robotics in organic synthesis and pharmaceutical production. Furthermore, it delves into the challenges and future prospects of LLMs in computational chemistry and code generation.

[Read more at Nature...](#)

