

AI News

Welcome to this edition of AI News Magazine, where we're sharing the pulse of the groundbreaking and innovative world of Artificial Intelligence (AI).

Our features begin with a fascinating exploration of AI's growing influence on journalism, showcasing how five Pulitzer finalists utilized AI assistance to drive their compelling stories. This trend brings into focus the intersections of journalism and AI, providing exciting terrain for discourse and advancement.

We spotlight the AI advancements in contemporary industries, including GitHub's novel AI tool designed to identify and rectify vulnerabilities in code, Predibase's specialized large language models that surpass the performance of GPT-4, and Nvidia's partnership with Hippocratic AI to create advanced health care "agents" that promise to provide efficient and cost-effective alternatives to human-led care.

In language processing, we introduce you to the 'gpt-prompt-engineer'; an AI tool aiming to streamline the search for effective prompts, the latest Claude-Journalist, a program designed to autonomously create news content, and Stable Code Instruct 3B, which demonstrates impressive advancement in assisting programming tasks.

The innovative applications of AI don't stop there. In the realm of sports strategy, Google DeepMind's TacticAI has been causing waves with geometric deep learning tactics set to revolutionize football, while researchers have tweaked unprecedented approaches from Chess-GPT to optimize the understanding of board states.

We focus on the essential challenges and potentials in AI's integration into education, healthcare, and economic growth forecasting, including a provoking discussion on the prospects of AI-induced hypergrowth in the global economy. In software engineering, Meta is marking notable progress with AUTI, a system powered by large language models that are bolstering automated unit testing.

Furthermore, we introduce the innovative AI toolkit LMFlow, a comprehensive AI framework Distribution Matching Distillation (DMD) for accelerated image generation, and Mojo, a novel programming

language that combines Python's accessibility with the sophistication of advanced system programming.

Engage with stories of pioneering technologies, insights into AI's influence on several sectors, and the transformative progress this field continues to make. The future of AI may be uncharted, but as we traverse this terrain, one thing is increasingly clear - we're in the midst of a technological paradigm shift with far-reaching impacts and opportunities.

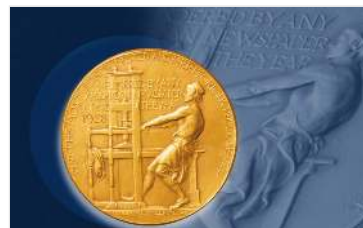
Here's to harnessing the prowess of AI for a better and smarter future! Enjoy this edition's deep dive into the fantastic world of artificial intelligence.

Five of this year's Pulitzer finalists are AI-powered

2024-03-17

For the first time, Pulitzer Prizes for journalism required entrants to disclose AI use in their reporting, with five finalists admitting to using AI tools. This reflects the Pulitzer Board's interest in understanding AI's role in journalism. Similarly, the George Polk Awards are developing an AI disclosure policy. Both awards are considering the implications of AI-assisted work while preserving the integrity of human-driven journalism.

[Read more...](#)



LoRA Land: Fine-Tuned Open-Source LLMs that Outperform GPT-4

2024-03-19

Predibase introduces LoRA Land, a suite of 25 specialized large language models that outperform GPT-4 by 4-15% across various tasks. Based on the Mistral-7b architecture, these models are cost-effectively fine-tuned using Parameter Efficient Fine-Tuning and Quantized Low Rank Adaptation.

The models cover applications from content moderation to SQL generation and can be deployed on a single A100 GPU, offering cost savings and scalability. This initiative demonstrates how task-specific models can provide a cost-effective alternative to commercial LLMs.

[Read more...](#)



GitHub's latest AI tool can automatically fix code vulnerabilities

2024-03-20

GitHub has launched a beta version of its code-scanning autofix tool, designed to identify and rectify code vulnerabilities in real-time. The tool integrates GitHub's Copilot and CodeQL, and can fix over two-thirds of detected vulnerabilities, covering more than 90% of alert types for supported

languages. Available to GitHub Advanced Security customers, the feature aims to save developers time and allow security teams to focus on higher-level strategies.

[Read more...](#)



Gpt Prompt Engineer Streamlines Finding The Most Effective Prompts

2024-03-20

The `gpt-prompt-engineer` tool uses AI models like GPT-4 and Claude 3 Opus to streamline the process of finding effective prompts. The tool generates, evaluates, and ranks prompts based on user-defined tasks and test cases. The latest update introduces auto-generation of test cases and a classification version for specific tasks. It integrates with Weights & Biases and Portkey for logging and tracing, and is open for contributions.

[Read more...](#)

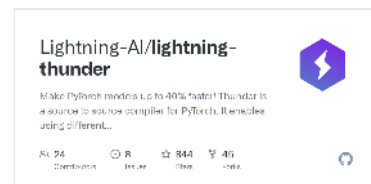


Thunder, a new compiler for PyTorch

2024-03-21

Lightning Thunder, a new source-to-source compiler, is set to boost the performance of PyTorch programs. It optimizes code and uses top-tier executors to achieve significant speed improvements, with a demonstrated 40% increase in training throughput for certain models. Thunder supports distributed strategies, offers a JIT compiler for Python callables and PyTorch modules, and is fully compatible with standard PyTorch. It's user-friendly, easily extendable, and open for contributions under the Apache 2.0 license.

[Read more...](#)



Nvidia announces AI-powered health care 'agents' that outperform nurses — and cost \$9 an hour

2024-03-22



Nvidia has partnered with Hippocratic AI to create advanced AI health care agents designed to improve patient interactions through empathetic and responsive communication. These agents, leveraging Nvidia's technology and Hippocratic's health care-focused language model, have demonstrated superior performance compared to human nurses in various tasks, such as identifying medication impacts and detecting toxic drug dosages. The collaboration aims to address the shortage of health care workers by providing a cost-effective alternative, with AI agents operating at \$9 per hour compared to the median nurse hourly rate of \$39.05. The initiative reflects a broader trend of integrating AI into health care to enhance access, equity, and patient outcomes.

[Read more at Fox Business...](#)

AI generates high-quality images 30 times faster in a single step

2024-03-22

MIT CSAIL researchers have developed a new framework that accelerates AI image generation by 30 times. The method, called distribution matching distillation (DMD), streamlines the process into a single step while maintaining or improving image quality. DMD uses a teacher-student model and bypasses common issues in generative adversarial networks. The technique shows promise in revolutionizing content creation, with potential applications in design tools, drug discovery, and 3D modeling.

[Read more...](#)

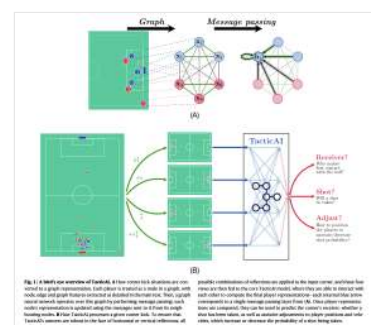


Google DeepMind Researchers Introduce TacticAI: A New Deep Learning System that is Reinventing Football Strategy

2024-03-23

DeepMind researchers have developed TacticAI, an AI assistant that uses geometric deep learning to revolutionize football strategy, particularly corner kicks. By converting player positions and movements into graph representations, TacticAI predicts player movements, counter-attack threats, and optimal scoring positions. The system's effectiveness has been proven through extensive testing, signaling a potential shift towards AI-augmented tactics in football coaching.

[Read more...](#)

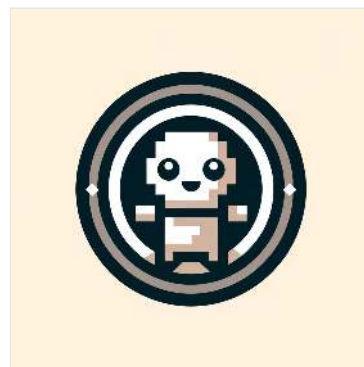


Lumentis: AI powered one-click comprehensive docs from transcripts and text.

2024-03-24

Lumentis is an AI-powered tool that transforms transcripts and unstructured data into comprehensive, easy-to-navigate documents. It offers flexibility with different AI models, cost estimates, and transparency. The tool can handle a variety of topics, with upcoming features including folder organization and auto-transcription. Open for collaboration on GitHub, Lumentis invites users to test and share their experiences to enhance the tool further.

[Read more...](#)



Manipulating Chess-GPT's World Model

2024-03-25

Chess-GPT, a language model trained on chess moves, can estimate player skills and board states, but struggles with random initial moves. By tweaking the model's internal activations, researchers improved its performance and understanding of board states. This suggests that such models can develop complex world models through self-supervised learning, but also highlights the need for better interpretability methods in AI.

[Read more...](#)

Introducing Stable Code Instruct 3B

2024-03-25

Stable Code Instruct 3B, a state-of-the-art large language model for coding, outperforms rivals in various programming tasks. It enhances code completion, facilitates natural language interactions, and supports multiple languages. The model excels in Fill in the Middle tasks, database queries, code translation, and more. Its advanced instruction tuning showcases impressive mathematical reasoning and technical narrative comprehension. Available for commercial use, it marks a significant advancement in AI-assisted programming tools.

[Read more...](#)

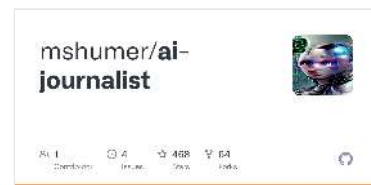


ai-journalist

2024-03-25

Claude-Journalist is an innovative AI-powered tool designed to autonomously research, write, and edit articles on any topic. Built on the Claude 3 AI model, it harnesses web search APIs to source information and the newspaper3k library to extract article text. Users can experience a code-free version on the HyperWrite Platform. The workflow involves prompting for a topic, generating search terms, selecting relevant URLs, and crafting a high-quality article, which is then reviewed and refined by the AI itself. To use Claude-Journalist, one needs Anthropic and SERP API keys. Despite its capabilities, the tool is experimental and content produced should be fact-checked by humans. The project, open for contributions and improvements, is under the MIT License. For those interested in further AI advancements, the HyperWrite Personal Assistant offers enhanced writing skills and browser operation capabilities.

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

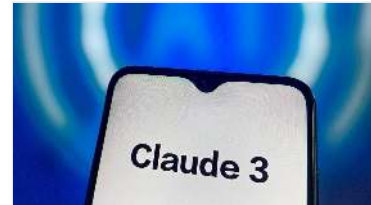


Claude takes the top spot in AI chatbot ranking — finally knocking GPT-4 down to second place

2024-03-27

Anthropic's Claude 3 Opus has ousted OpenAI's GPT-4 from the top spot in the LMSYS Chatbot Arena, marking the first time GPT-4 has been dethroned. Despite the narrow victory, the impending launch of GPT-5 could threaten Anthropic's lead. The leaderboard, which uses the Elo rating system, also features impressive performances from smaller models like Claude 3 Haiku. The dominance of proprietary models poses a challenge for open-source AI, but upcoming developments suggest a dynamic future for AI competition.

[Read more...](#)



DBRX: A New State-of-the-Art Open LLM

2024-03-27

Databricks has launched DBRX, a new open-source large language model (LLM) that sets a new performance standard. With a fine-grained mixture-of-experts architecture, DBRX is faster and smaller than its competitors, and more FLOP-efficient during training. The model's weights are available on Hugging Face, and it's already integrated into GenAI-powered products. Trained using Databricks' tools on NVIDIA H100 GPUs, DBRX offers optimized serving infrastructure for efficient inference.

[Read more...](#)

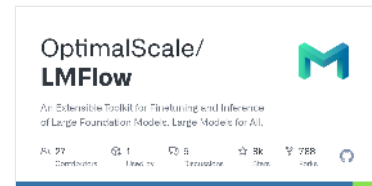


GitHub – OptimalScale/LMFlow: An Extensible Toolkit for Finetuning and Inference of Large Foundation Models. Large Models for All.

2024-03-28

LMFlow, an extensible toolkit for finetuning large machine learning models, now supports LISA, a memory-efficient algorithm, and other models like Llama2, ChatGLM2, and Baichuan. With features for finetuning acceleration, memory optimization, and inference acceleration, LMFlow is user-friendly, fast, and reliable. Available on PyPI, it comes with comprehensive documentation and examples, making it accessible to a wide community.

[Read more...](#)



The AI Economy: Debating the Prospects for Hypergrowth

2024-03-28

Artificial Intelligence (AI) could trigger unprecedented economic growth, potentially reaching annual rates of 20-30% by the end of the century, according to researchers. However, this depends on AI achieving human-level capabilities and automating economically valuable ideas. While this could lead to explosive growth, it may also result in significant unemployment and inequality. If AI falls short, it could still boost productivity in sectors like healthcare and education, leading to a modest but impactful economic growth.

[Read more...](#)



Large Language Models Boost Automated Unit Testing, Study Finds

2024-03-30

Meta researchers have developed AUTI, a system that uses large language models to automatically enhance unit test suites. In tests, AUTI increased branch coverage by 10.2% on 20 open-source Java projects and discovered 19 new bugs. The system could significantly improve developer productivity and software quality assurance, although further validation is needed. The research contributes to the growing field of AI in software engineering, with potential to transform how software is built and tested.

[Read more...](#)



Mojo: The New Programming Language Merging Python's Ease with Advanced Systems Capabilities

2024-03-30

Introducing Mojo, a revolutionary programming language that combines Python's simplicity with advanced systems programming. Still in its early stages, Mojo aims to bridge the gap between research and production environments. The language offers a rich ecosystem and a robust standard library, with a community-driven approach encouraging contributions and active developer interaction.

[Read more...](#)

