

Advancing Science and Math with GPT-5.2

Stronger Performance Where Precision Matters

At the heart of GPT-5.2's prowess is its robust mathematical reasoning. This capability is crucial for scientific and technical work, where multi-step logic, consistent quantities, and error-free analyses are paramount. GPT-5.2's improvements on benchmarks like FrontierMath highlight its general reasoning and abstraction skills, which are essential for coding, data analysis, and experimental design.

On the GPQA Diamond benchmark, a graduate-level Q&A test covering physics, chemistry, and biology, GPT-5.2 Pro achieved an impressive 93.2% accuracy, with GPT-5.2 Thinking close behind at 92.4%. These results underscore the model's ability to handle complex scientific queries without external tools, relying solely on its reasoning capabilities.

In the field of mathematics, GPT-5.2 Thinking set a new record by solving 40.3% of expert-level problems on FrontierMath. This achievement demonstrates the model's capacity to engage with intricate mathematical concepts and provide solutions that were previously unattainable.

Case Study: Resolving Open Research Problems

GPT-5.2's impact extends beyond benchmark performance. In a notable case study, the model assisted researchers in resolving a long-standing open problem in statistical learning theory. The question at hand—whether more data consistently improves results in statistical models—has puzzled researchers for years. GPT-5.2 Pro provided a direct solution, which was then meticulously verified by human experts.

This collaboration highlights a new paradigm in scientific research, where AI models serve as tools for exploration and hypothesis testing, while human researchers ensure accuracy and interpret the findings. GPT-5.2's ability to extend its results to higher-dimensional settings further exemplifies its potential to drive innovation across various domains.

Looking Ahead: The Future of AI in Science

The advancements showcased by GPT-5.2 suggest a promising future for AI in scientific research. In fields with strong axiomatic foundations, such as mathematics and theoretical computer science, AI models can explore proofs, test hypotheses, and uncover connections that might otherwise require significant human effort.

However, it is essential to recognize that AI systems are not independent researchers. Expert judgment, verification, and domain understanding remain critical. By integrating AI into research workflows with a focus on validation, transparency, and collaboration, we can harness its full potential to accelerate scientific discovery.

In conclusion, GPT-5.2 represents a significant leap forward in AI's ability to assist and enhance scientific research. Its precision, reasoning capabilities, and collaborative potential make it an invaluable tool for researchers worldwide. As we continue to explore the possibilities of AI in science, GPT-5.2 stands as a testament to the transformative power of technology in advancing human knowledge.

Contributor: Alessandro Linzi