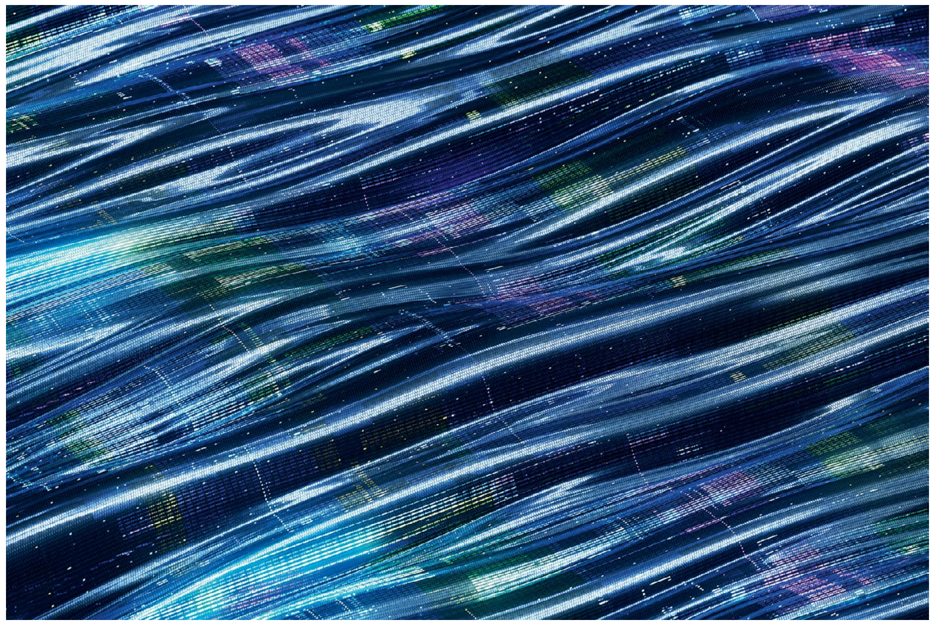
Editorial

https://doi.org/10.1038/s43588-025-00890-x

The rise of large language models

Check for updates

This issue of *Nature Computational Science* features a Focus that highlights both the promises and perils of large language models, their emerging applications across diverse scientific domains, and the opportunities to overcome the challenges that lie ahead.



CREDIT: EUGENE MYMRIN / MOMENT / GETTY IMAGES

Large language models (LLMs) are increasingly shaping the way we live and work. In everyday life, they assist with writing, translation, learning, communication, and so on — by making information more accessible and tools more efficient. LLMs also profoundly influence how knowledge is created and shared. In scientific research, for example, LLMs are transforming how research is conducted — from literature synthesis and hypothesis generation to experimental design and scientific code development. Their impact spans a wide range of disciplines, including life sciences and medicine, chemistry and materials science, physics, engineering, urban and Earth sciences, psychology, linguistics, and the humanities. As these models continue to evolve, they are not only enhancing existing methods but also unlocking new possibilities for scientific exploration. In this issue, we present a Focus that brings together expert perspectives from various fields to explore the opportunities, risks, and challenges of advancing and applying LLMs in scientific research.

Without a doubt, the transformer architecture has been central to the success of modern LLMs, powering models such as ChatGPT — whose website ranks among the most visited globally, as highlighted by Pedro Burgos in a News Feature. The transformer's self-attention mechanism enables the capture of long-range dependencies and contextual relationships far more effectively than earlier architectures such as recurrent neural networks. In a Perspective, Eva Portelance and Masoud Jasbi examine how non-symbolic generative artificial intelligence (AI) — particularly transformer-based LLMs — align with Chomsky's generative linguistic principles, demonstrating noteworthy linguistic capabilities that are reshaping language processing research. Similarly, in a Comment, Gabrielle O'Brien emphasizes how LLMs can assist with computer programming — the language humans use to interact with digital systems — and accelerate scientific workflows.

The success of LLMs extends far beyond language tasks, as highlighted by several examples in this Focus issue. In cognitive science, Ilia Sucholutsky and colleagues explore, in a Comment, the potential of LLMs in advancing the study of collective cognition — cognitive phenomena that emerge from social interaction between multiple individuals. They identify five roles for LLMs in cognitive research: participant, analyst, environment, interviewer, and facilitator, enabling the study of complexities that challenge conventional methodologies. Yong Li and colleagues envision, in their Perspective, LLMs as intelligent assistants in urban planning, capable of synthesizing ideas, generating designs, and evaluating planning outcomes to address growing urban complexities. In the humanities, Ted Underwood notes in a Comment that scholars are using LLMs to frame new research questions and even rethink how these models are trained, signaling a transformative dialogue between AI and historical inquiry. In chemistry and materials science, Gabe Gomes and collaborators, in their Perspective, emphasize opportunities for LLMs to support planning, optimization, data analysis, and automation, positioning them as active partners in chemical research. Collectively, these examples illustrate how LLMs are enhancing existing practices and opening new frontiers across various scientific domains.

Despite the successes and the promise, LLMs present substantial risks and challenges that the scientific community attends to and prepares strategies to mitigate. For instance, when using LLMs for social science inference, Lisa P. Argyle and colleagues point out, in a Perspective, the risks of unclear research goals and further emphasize the importance of two key guidelines: clearly defining the target of inference and determining when and under what conditions an inference is valid. In the context of AI coding tools, O'Brien warns that over-reliance on AI-generated code can lead to undetected semantic errors, particularly in scientific software, which is often untested and developed by scientists with limited programming expertise. In another Comment, Brett Beaulieu-Jones underscores that fairness and health equity remain underexplored when applying LLMs in healthcare, calling for clearer definitions and metrics to mitigate disparities. Additionally, global deployment of LLMs faces structural challenges, as noted in a News Feature on LLM developments in the Global

South, where limited hardware, training data, and capital investment risk widening existing inequalities. These examples illustrate that while progress is being made, ensuring reliability, fairness, and inclusivity in the development and application of LLMs remains an ongoing challenge.

It is worth mentioning that training and running LLMs on conventional digital processors such as GPUs is extremely energy intensive. As Yudeng Lin and Jianshi Tang note in their News & Views, deploying LLMs at scale can consume as much energy as an entire country, such as Ireland. This challenge stems from the enormous parameter space of LLMs and the limitations of the von Neumann architecture, which separates memory and computing units. The shuffling of vast amounts of data during training and inference leads to high energy costs and latency, creating critical constraints on further scaling of LLMs. To overcome these bottlenecks, alternative architectures such as in-memory computing (IMC) and neuromorphic computing hold the potential, as emphasized by Emre Neftci and colleagues in their Comment. Notably, their research Article, which is also published in this issue, demonstrates that IMC with gain-cell crossbar arrays can accelerate the attention mechanism in LLMs with 1.5 billion parameters, achieving a 100-fold speed-up and a 70,000-fold reduction in energy consumption compared to GPUs. These advances underscore the need to rethink hardware architectures to sustain LLM growth with energy efficiency and scalability.

With this Focus issue, we aim to underscore LLMs as catalysts for scientific discovery and interdisciplinary innovation, while acknowledging the substantial challenges they pose. Addressing these will require innovative hardware architectures, validation frameworks, and embedding ethical principles throughout their design and deployment — ensuring that LLMs drive breakthroughs that are not only powerful but also sustainable and equitable.

Published online: 24 September 2025