World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

WJAETS

(Review Article)

A survey on the rise of the AI scientists: Accelerating discovery and confronting ethical frontiers

Muskaan Goyal *

Department of Computer Science, University of California, Berkeley, US.

World Journal of Advanced Engineering Technology and Sciences, 2025, 15(02), 564-569

Publication history: Received on 06 March 2025; revised on 03 May 2025; accepted on 05 May 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.15.2.0646

Abstract

Large Language Models (LLMs) are rapidly revolutionizing scientific research. These tools, powered by LLMs, are not just assistants for tasks such as literature synthesis, hypothesis generation, pattern identification, and research paper writing, but are evolving into autonomous discovery systems. This review explores the capabilities of these 'AI Scientists' and compares leading entities like Sakana AI's autonomous system and Google's collaborative 'AI co-scientist'. While these advancements promise unprecedented acceleration of scientific progress, they also bring significant ethical challenges, such as bias amplification, reproducibility, transparency, authorship, and equity. By identifying current limitations and analyzing these challenges, we aim to ensure the responsible use of AI in scientific discovery.

Keywords: AI Scientist; Autonomous Scientific Discovery; Scientific Research Methodology; Ethics in AI Research; Research Ethics; AI Scientist Architectures

1. Introduction

The scientific method is a systematic approach to scientific inquiry characterized by systematic observation, experimentation, inductive and deductive reasoning, and formation and testing of hypotheses and theories [1]. This scientific method is undergoing a paradigm shift driven by artificial intelligence. Traditional scientific methods face human capacity challenges due to the scientific data's sheer volume and complexity of scientific data. Large Language Models (LLMs) offer powerful information processing, text generation, code generation, and pattern identification capabilities [2, 3]. These LLM capabilities are a natural fit for the scientific method and discovery, leading to integration in scientific workflows from assistance towards more autonomous roles [4].

This shift towards an autonomous role is categorized as an AI Scientist, a term that describes LLM-powered systems designed to automate core systems and mechanisms of the research lifecycle [5]. These systems, such as Sakana AI's 'AI Scientist-v2' and Google's 'AI co-scientist', are not just assistants but active participants in the scientific discovery process. For instance, Sakana AI's 'AI Scientist-v2' recently generated its first peer-reviewed scientific publication in a top machine learning conference through an automated research lifecycle [6]. This peer-reviewed workshop paper was generated without human interference through an agentic tree–search methodology [6]. Similarly, Google's 'AI co-scientist' utilizes multi-agent frameworks with models like Gemini to generate hypotheses and design experiments to be executed in collaboration with human researchers [7]. Google's 'AI co-scientist' had demonstrated promise in biomedicine. These projects indicate a future where AI is not an assistant but an active discovery system. While these initiatives depict the potential for breakthroughs in diverse fields, this shift necessitates careful examination of the capabilities, limitations, and ethical concerns surrounding it.

* Corresponding author: Muskaan Goyal.

2. Llms as Catalysts for Scientific Discovery

LLMs are acting as catalysts for the various scientific discovery stages. A range of strategies is emerging for integrating AI into scientific discovery. These approaches span from AI Scientists serving in supportive roles, such as reviewer or research assistant, to more involved positions as collaborative partners. At the most advanced level, some AI Scientists are designed to generate, lead, and execute the discovery process autonomously.

LLMs, trained on vast amounts of text, can automate and accelerate literature reviews by summarizing key findings, identifying patterns, analyzing unstructured scientific data, and generating code for analysis [8]. This automation can significantly reduce the workload for researchers in all disciplines and facilitate interdisciplinary communication.

2.1. LLMs, Novelty and Experimentation

Beyond automating known foundational tasks, LLMs can assist with hypothesis generation, experimental design, and workflow automation for analysis [9]. These AI scientists can analyze existing research to identify knowledge gaps and suggest novel connections between concepts and past discoveries. They can also use the proposed testable hypotheses to generate an experimental design and suggest sampling methods, parameters, protocols, and analysis methods. This potential for driving novelty and experimentation in research is fascinating.

Multi-agent frameworks have been key to enabling iterative experimentations [9]. For example, Google's "AI co-scientist" leverages a multi-agent framework in which AI agents generate, debate, and refine hypotheses. AI Scientists can significantly reduce the time required for brainstorming [10]. Similarly, Sakana AI's "AI Scientist-v2" uses an agentic-tree-search methodology with an experiment manager agent in collaboration with a Vision-Language Model feedback loop for iterative refinement of the experiments and content [6]. These systems aim to reduce research timelines significantly, allowing researchers to spend more time on new data collection.

3. Emerging Paradigms of "AI Scientist" Approaches

"AI Scientists" are diverse with varying systems and core functions. Key differences include their fundamental aim (human collaboration vs. full automation), focus (specific phases of research, entire research process, or domain-specific), varying human dependence (constant interaction, partial automation, or complete automation), and underlying AI technologies. Comparing these distinctions for some of the prominent approaches in our industry is crucial for understanding the potential impact of these AI tools on the future of scientific discovery.

Table 1 This table compares several prominent AI systems designed to assist with or automate aspects of the scientific research process in different industries

Feature	**Sakana AI Scientist (v1)**	**Sakana AI Scientist (v2)**	**Google AI Co-Scientist**	**DeepMind AlphaFold2**	**MIT SciAgents/**
Lead Organization	Sakana AI	Sakana AI	Google	Deepmind	MIT
Open Access	Yes	Yes	No	Yes	Yes
LLM Reliance	Yes (Can choose between OpenAI and Anthropic)	Yes (Can choose between OpenAI, Gemini, Anthropic)	Yes (Gemini Family)	No. It uses specialized deep learning.	Yes (OpenAI)
Frameworks	Multi-Agent Systems with code templates	Multi-Agent Systems with Agentic Tree Search	Multi-Agent Systems (Generate, Debate, Evolve) with human-in-the-loop framework	Deep learning Framework focused on diffusion learning.	Multi-Agent Systems, Tool Use
Primary Goal	Complete automation of the AI research cycle	Complete automation of the AI research cycle	Complement human scientists and accelerate their research through ideation and planning.	Solves a specific protein prediction problem	Autonomous research within specific scientific domains
Code and Experiment Generation	Partially automated with human-provided code templates	Automated	N/A (Since Co-Scientist is a collaborator and more relevant for novel hypothesis generation)	N/A (Does not generate experiments)	LLM agents plan and execute experiments using tools/simulations.
Human Interaction Level	Minimal	No Interactions	High since it is designed to collaborate and requires an expert in the loop.	Minimal, where an input sequence is required to be provided	High since humans define goals, review results, and integrate required LLM agents and tools.
Key Strengths	Proof of concept for fully automated research	Enhanced autonomy and automation with high scientific discoveries in machine learning	Strong collaborator for hypothesis generation, literature synthesis, and design support	Novel accuracy in protein folding and transformative impact in biology or medicine	Domain-specific automation.
Limitations	Requires human involvement with limitations in literary depth, novelty assessment, and code robustness	Limited in literary depth, novelty assessment, and code robustness	Focus on initial research like ideation, which requires human execution and validation	Highly domain-specific to protein structure prediction	Domain-specific and requires manual agent and tool integrations

4. Current Gaps and Limitations in AI-Driven Scientific Discovery

While AI scientists are making tremendous progress, significant hurdles lie before AI can function as autonomous scientists and make transformative scientific discoveries.

4.1. True Novelty

Developing new ideas and understanding is key to scientific progress [11]. Novelty often requires thinking creatively outside of the current ways of thinking. Currently, LLMs are primarily excellent at summarizing and interpreting existing knowledge. This capability will assist them in building on already established theories using the existing data. Still, they will lack the creativity to hypothesize ideas that question current scientific beliefs, resulting in limited disruptive discoveries [6]. Essentially, "AI Scientists" may primarily excel at incremental science and will be unable to make groundbreaking, novel discoveries that require creativity.

4.2. Echo Chamber Effect

In addition to the lack of novelty, AI Scientists also pose a risk of reinforcing existing knowledge and theories due to their inference and summarization capabilities [12]. These systems are trained on vast amounts of current research data and theories, but often lack adequate exposure to alternative theories. This echo chamber effect can lead to confirmation bias and harm the discovery of new knowledge [12].

4.3. Data Availability and Quality

LLMs are highly sensitive to their training data, so any gaps in available data (i.e, missing data, biased data, or inaccurate data) will result in low-quality research conducted by AI Scientists [13]. Furthermore, human intervention will be required to resolve hallucinations, resolve coding errors, and collect new data to make groundbreaking discoveries.

4.4. Lack of Adequate Robotics Advancement

Since "AI Scientists" rely on computational methods to make discoveries, they will have a greater impact in domains with simulation-dependent research. However, domains that depend on physical environments are less likely to benefit due to the limitations and challenges in integrating an advanced robotic system for physical experimentation.

4.5. Poor Multimodal Integration

Scientific Research depends on diverse data types, including but not limited to text, sensor data, images, code, and diagrams. Despite the advancements made by multimodal AI, AI Scientists still lack an efficient, streamlined integration of different data types within their workflow.

5. Ethical concerns

LLM's impressive capability to conduct autonomous research has introduced significant ethical concerns. These concerns require careful consideration and governance to mitigate.

5.1. Access

AI scientists are powered using LLMs, so they require extensive computational resources, which are not readily available in developing nations. This disparity in access to computational resources can amplify the disparities in scientific research capacity [14].

5.2. Accountability

In the traditional scientific discovery process, human researchers are entirely accountable for their research, but with AI-driven scientific research, the accountability becomes complicated.

5.3. Authorship

Human researchers who use AI as collaborators for their research now need to disclose their use. When AI plays a significant role in research, it becomes challenging to identify who should be credited as the primary contributor, whether the human author or the AI system. This raises important questions: how do we distinguish between the contributions made by humans and those created by the AI system? And if an AI system makes a meaningful contribution, who holds the authorship rights, the researcher using the AI, or the organization that developed and trained it?

5.4. Bias

LLMs are trained on vast amounts of data and are known to amplify biases in these existing training data [12]. When generating new scientific discoveries, LLMs reinforce existing societal biases, skewed findings, and unfair results in the existing scientific literature. This challenge is further amplified by the bias inherited by the large language models.

5.5. Misinformation

As LLMs are known for hallucinating, there is always a risk of fabricating research [13]. These hallucinations are challenging to detect in large research papers, requiring rigorous human verification, and can lead to the creation of more paper mills [14].

5.6. Transparency

LLMs are known to be "black boxes" since it is difficult to understand an output precisely [15]. This black box nature can result in a lack of transparency and reproducibility for the discovery, making it less credible [16].

In addition to these concerns, AI Scientists also suffer from privacy, environmental costs, equity, and deskilling concerns due to relying on large language models.

6. Future Directions: Enhancing LLM Use in Research Communication

Beyond advancing scientific discovery, LLMs can improve how research is communicated and reproduced in various ways.

6.1. Refined manuscript

Human researchers can enhance their manuscripts by leveraging prompts to improve consistency. Some ways include interactive exploration of scientific literature, auditing research methodologies, and augmenting research paper structure. Academic conferences like ICLR have already incorporated AI-generated suggestions in their mechanisms. These are usually surface-level reviews that require a secondary reviewer to detect deeper methodological flaws in research papers [17].

6.2. Reproducibility

LLMs' ability to quickly process large volumes of information makes them valuable tools for testing the reproducibility of published findings and reviewing the paper for any flaws. These systems can help review, replicate, and validate prior research results by identifying bias through reflection prompts, conducting simulated peer review, providing contextualized reference checking, and reproducing a hypothesis to compare the results using an automated experimental setup and data analysis.

7. Conclusion and Responsible Navigation

LLMs offer an unprecedented trajectory toward an "AI Scientist" capable of accelerating and automating scientific discovery through hypothesis generation, experimental design, and paper generation. Comparing numerous AI scientists' approaches, such as Sakana AI's "AI Scientist-v2," Google's "co-scientist," and DeepMind's "AlphaFold," reveals different capabilities, use cases, and limitations. Despite their tremendous capabilities, these AI Scientists have significant limitations on novel creativity, data quality, and multimodal complexity. Furthermore, these systems face the complexity of accountability and authorship, inherent bias, misinformation due to hallucinations, and a lack of transparency and equity concerns.

Interdisciplinary efforts and research on reliability, validation, and understanding are required to mitigate AI Scientists' limitations. The scientific community also involves developing ethical guidelines, transparency standards, and accountability frameworks. With sustained collective engagement and collaboration among researchers, domain experts, publishing houses, policymakers, and research institutions, we can ensure that AI will lead to groundbreaking discoveries and advance our scientific knowledge with fairness and equity.

Compliance with ethical standards

Disclosure of conflict of interest

There are no conflicts of interest to be disclosed.

References

[1] Andersen H, Hepburn B. Scientific Method [Internet]. Stanford Encyclopedia of Philosophy. 2015. Available from: https://plato.stanford.edu/entries/scientific-method/

[2] Naveed H, Khan AU, Qiu S, Saqib M, Anwar S, Usman M, et al. A Comprehensive Overview of Large Language Models [Internet]. arXiv.org. 2023. Available from: https://arxiv.org/abs/2307.06435

[3] Jiang J, Wang F, Shen J, Kim S, Kim S. A Survey on Large Language Models for Code Generation [Internet]. arXiv.org. 2024. Available from: https://arxiv.org/abs/2406.00515

[4] Eger S, Cao Y, D'Souza J, Geiger A, Greisinger C, Gross S, et al. Transforming Science with Large Language Models: A Survey on AI-assisted Scientific Discovery, Experimentation, Content Generation, and Evaluation [Internet]. arXiv.org. 2025. Available from: https://arxiv.org/abs/2502.05151

[5] Lu C, Lu C, Lange RT, Foerster J, Clune J, Ha D. The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery [Internet]. arXiv.org. 2024. Available from: https://arxiv.org/abs/2408.06292

[6] Yamada Y, Lange R, Lu C, Hu S, Lu C, Foerster J, et al. The AI Scientist-v2: Workshop-Level Automated Scientific Discovery via Agentic Tree Search [Internet]. arXiv.org. 2025. Available from: https://arxiv.org/abs/2504.08066

[7] Gottweis J, Weng WH, Daryin A, Tu T, Palepu A, Sirkovic P, et al. Towards an AI co-scientist [Internet]. arXiv.org. 2025 [cited 2025 Mar 8]. Available from: https://arxiv.org/abs/2502.18864

[8] Scherbakov D, Hubig N, Jansari V, Bakumenko A, Lenert LA. The emergence of Large Language Models (LLM) as a tool in literature reviews: an LLM automated systematic review [Internet]. arXiv.org. 2024 [cited 2024 Oct 27]. Available from: https://arxiv.org/abs/2409.04600

[9] Alkan AK, Sourav S, Jablonska M, Astarita S, Chakrabarty R, Garuda N, Khetarpal P, Pióro M, Tanoglidis D, Iyer KG, Polimera MS. A Survey on Hypothesis Generation for Scientific Discovery in the Era of Large Language Models. arXiv preprint arXiv:2504.05496. 2025 Apr 7.

[10] Lavrič F, Škraba A. Brainstorming will never be the same again—a human group supported by artificial intelligence. Machine Learning and Knowledge Extraction. 2023 Sep 25;5(4):1282-301.

[11] Bird A. What is scientific progress?. Noûs. 2007 Mar 1;41(1):64-89.

[12] Sharma N, Liao QV, Xiao Z. Generative echo chamber? effect of llm-powered search systems on diverse information seeking. InProceedings of the 2024 CHI Conference on Human Factors in Computing Systems 2024 May 11 (pp. 1-17).

[13] Banerjee S, Agarwal A, Singla S. Llms will always hallucinate, and we need to live with this. arXiv preprint arXiv:2409.05746. 2024 Sep 9.

[14] Kendall G, Teixeira da Silva JA. Risks of abuse of large language models, like ChatGPT, in scientific publishing: Authorship, predatory publishing, and paper mills. Learned Publishing. 2024 Jan 1;37(1).

[15] Quttainah M, Mishra V, Madakam S, Lurie Y, Mark S. Cost, usability, credibility, fairness, accountability, transparency, and explainability framework for safe and effective large language models in medical education: Narrative review and qualitative study. Jmir Ai. 2024 Apr 23;3(1):e51834.

[16] Boyko J, Cohen J, Fox N, Veiga MH, Li JI, Liu J, Modenesi B, Rauch AH, Reid KN, Tribedi S, Visheratina A. An interdisciplinary outlook on large language models for scientific research. arXiv preprint arXiv:2311.04929. 2023 Nov 3.

[17] Beel J, Kan MY, Baumgart M. Evaluating Sakana's AI Scientist for Autonomous Research: Wishful Thinking or an Emerging Reality Towards' Artificial Research Intelligence'(ARI)?. arXiv preprint arXiv:2502.14297. 2025 Feb 20.