npj | digital medicine

Perspective

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-025-01859-w

Will AI become our Co-PI?

Dillan Prasad[1] ✉, Aditya Khandeshi[1], Spencer Sartin[2], Rishi Jain[1], Nader Dahdaleh[1], Maciej Lesniak[1], Yuan Luo[3] & Christopher Ahuja[1]

Rapid advances in large language models (LLMs) are transforming the role of students and principal investigators (PIs) in biomedical research. This perspective examines how LLMs can reshape the laboratory model as de facto "Co-PIs" for tasks ranging from literature triage to hypothesis generation. By clarifying both opportunities and risks, we propose a framework for efficient AI collaboration which aims to guide investigators and trainees in harnessing LLMs responsibly.

Matching to competitive specialties is becoming increasingly challenging for medical students. Students are applying to residency with greater numbers of publications than ever before[1]. For example, in neurological surgery the average medical student today will apply with 37.4 research items (abstracts, posters, publications) versus 7.8 in 2009, a 379% increase[2,3]. Orthopedics, plastic surgery, and dermatology demonstrate this phenomenon at comparable magnitudes, though candidates for all specialties feel broadly compelled to publish higher volumes. However, it is unclear if increased research productivity correlates with quality. Evidence suggests that many medical student publications—oftentimes reviews and re-reviews of the literature—are never cited[4].

Research can broadly be classified as "synthesis" or "aggregation." Organic discovery, or "synthesis-type" medical research, produces something new, such as basic science experiments, clinical trials, and other works aimed at deepening our understanding of truth through hands-on experimentation with reagents, equipment and subjects/participants. However, most trainees engage substantially in "aggregation-type" research. Aggregation typically includes rote, repetitive tasks such as reviewing charts, extracting variables for meta-analyses, and systematic reviews.

These tasks contribute to knowledge generation but do so in a fundamentally different manner. The time required to complete aggregation-type studies is often much shorter compared to synthesis-type studies, which may be compelling for pressured students but less beneficial to society if substantial aggregate effort is otherwise shunted away from synthesis-type research at the national and international level[4]. Indeed, efforts have been made to realign trainee efforts and disincentivize "low effort" publications[5]. While medical trainee research output is an illustrative example, these challenges apply broadly to all fields and subfields of scientific discovery.

It is becoming increasingly clear that rote "aggregation" research tasks are proving anachronistic in the era of machine learning (ML) and large language models (LLMs). In this perspective, we examine how LLMs might serve as effective collaborators by relieving trainees of the rote aggregation work that dominates "aggregation-type" output, thus redirecting effort toward higher-value scientific synthesis. Recognizing the advances of AI in sophisticated areas such as parallelized robotics and deep simulations, we argue that AI itself might eventually participate in "synthesis-type" tasks. Using medical students in medical research as an example, we offer a framework for investigators in all disciplines who seek to integrate these systems into workflows. We raise open questions about maintaining scientific rigor, data privacy, propagated bias, and authenticity in the setting of rapidly evolving technology which continually challenges the conventional model of scientific inquiry.

AI for aggregation research

LLMs hold promise to automate many repetitive tasks with greater accuracy and efficiency. For example, LLM-based literature search models, biology protocol planners, and statistical tools have shown strong early results compared to conventional standards[6–8]. AI systems can rapidly screen large volumes of data, analyze existing literature, and extract key variables from extensive datasets, reducing the manual burden on students.

This not only accelerates data collection but stands to improve inter- and intra-rater reliability and reduce human error, including well-documented patterns of redundancy and research waste arising from repeated systematic reviews[9]. Medical students training to be clinician-scientists stand to gain from this. By reallocating time from rote tasks to critical thinking such as contextualizing findings, examining quality of evidence, and evaluating the implications of unexpected results, student training can more closely align with the role of the Principal Investigators (PIs) they are preparing to become. It may be more efficient for LLMs to automate the repetitive aspects of aggregation than to consume the precious time of physicians and trainees.

Nonetheless, widespread adoption of this still-nascent technology has encountered barriers including regulatory and institutional concerns around health information protection, data security, and the IT complexities of integration with existing electronic medical record (EMR) systems[10]. Presently, no LLM is approved by the FDA; while models may be used for clinical decision-making support, most offer disclaimers that generated information is not intended for medical use, sidestepping the need for FDA regulation[11]. However, LLMs compliant with the Health Insurance Portability and Accountability Act (HIPAA) are already used in a variety of

[1]Department of Neurological Surgery, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [2]Department of Molecular Engineering, University of Chicago, Chicago, IL, USA. [3]Institute for Artificial Intelligence in Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA.
✉e-mail: dillan@northwestern.edu

contexts including clinical documentation, information extraction and decision support, and require a bundle of context-dependent privacy-preserving measures to meet de-identification and minimum-necessary standards[12].

We anticipate market forces will prevail in this space; government and private funding will flow to health systems beneficiaries capable of deploying HIPAA-compliant LLMs to mine insights from their proprietary datasets. Ultimately, economies of scale will reduce costs; just as EMRs have largely replaced paper documentation, AI tools may eventually replace much of the manual effort in aggregation-type research while preserving patient privacy[12].

AI for synthesis research

Synthesis-type experimentation requires physical manipulation of matter. Reagents (e.g., chemical compounds, proteins, antibodies, etc.), energy (e.g., ultrasound, radiation, electromagnetic waves, etc.), and varying conditions are delivered to cell lines, animal subjects or human participants to study their effects in a controlled way. Conventionally, skilled humans with subject matter expertise have performed synthesis research through often arduous manual experimentation. This has created challenges with reproducibility, reagent consistency, and loss of expert knowledge as trainees/technicians change roles or fields.

Novel AI-powered systems have demonstrated promise for automating aspects of synthesis research for biomedical discovery[13]. Google's DeepMind developed AlphaFold, an AI system and open-source database that accurately predicts protein structures based solely on amino acid sequences, drastically accelerating research capabilities and leading to numerous scientific discoveries[14]. Deep learning models are increasingly used in drug discovery for identifying novel targets, drug structure, and assisting with clinical trials[15]. An AI model enabled the high-throughput screening of ~7500 molecules to identify a novel antibiotic, abaucin, which has shown potential to treat drug-resistant strains of *Acinetobacter baumannii*[16].

There is significant recent interest in connecting AI models to robotic actuators with the goal of autonomous physical experimentation, which is already underway in different academic and private sector contexts[17,18]. These autonomous "scientists" could run thousands of iterations, refine experimental designs on the fly, and uncover new scientific truths, potentially empowering biomedical discovery through human-AI collaborations[13,15]. In this conception of AI-assisted research, human PIs could generate hypotheses, design protocols, and execute experiments with parallelized robotics. Critically, the research framework is envisioned and planned by human researchers who provide explicit instructions to the AI; human creativity guides the tireless machines to leverage the strengths of each.

An added benefit to this arrangement is the possibility that parallelized robotic "scientists" might stumble upon a novel finding by chance, outside of the intended scope of an experimental hypothesis. Indeed, many of the greatest scientific discoveries have been made accidentally, such as penicillin, X-rays, and nuclear fission[19]. Even without complete awareness, it is reasonable that an LLM-integrated robotic system working within the constraints of our defined physical and theoretical sandbox may produce a truly serendipitous discovery. This raises some challenging questions about how to define human innovation and blurs the line between AI and humans further[20]. Nonetheless, by leveraging AI's iterative capacity, we may substantially increase the volume of exploratory experiments while potentially lowering the cost and time required to make field-changing discoveries.

AI as Co-PI

While robotics could optimize research workflows as indefatigable staff, the path to AI becoming a collaborative principal investigator lies in enhanced hypothesis generation. LLMs excel at processing and identifying patterns in vast datasets which are inaccessible or otherwise impossible for humans to reproduce at comparable cost given biological and energy constraints. Further, though precise estimates are elusive given the proprietary nature of most LLM training sets, there is consensus that the models have now read most text data available on the internet, including scientific papers and other paywalled content, prompting the need for synthetic data which itself is subject to limitations[21,22]. It stands to reason that LLMs, with access to unimaginable scientific data and the ability to integrate it, may be able to identify patterns and infer causality to guide human researchers.

In other words, research models may be capable of inferential ideation that is productive in a scientific context, leading to collaborative "AI Co-PIs" where LLMs can generate hypotheses and guide specific experiments in partnership with human PIs (Fig. 1). For example, consider the European physicists of the 19th and 20th centuries, whom many would regard as the greatest minds of human history. While Bohr, Maxwell, and Einstein had access to only the best empirical data that the technology of their time afforded, they were still able to infer the physical frameworks of the standard model, quantum mechanics, and relativity. Long after their deaths, contemporary physicists have been able to retroactively validate their theories through empirical data produced from modern technology. Analogously, LLMs may not require empirical data beyond what is already accessible in the literature in order to produce guiding hypotheses of scientific value.

The ingenuity of history's physicists likely relied on some combination of lived experience, breadth of knowledge, creativity, speed of processing, recall, intelligence quotient, and all other physiological or psychometric factors known to underpin human quality of thought. LLMs already outperform humans on some of these metrics; for example, recent work has shown that GPT-4 consistently demonstrated higher originality and elaboration on divergent thinking tasks—including the Alternative Uses Task, Consequences Task, and Divergent Associations Task—compared to human participants, potentially suggesting greater creativity among models[23]. Other efforts have produced LLMs which iteratively refine hypotheses using a balance of exploration and exploitation strategies, functionally emulating human scientific reasoning where hypotheses are continuously tested and refined based on new information[24]. Taken together with evidence demonstrating LLM use of human decision-making heuristics as well as some alignment with human moral and causal judgements, it is reasonable to infer that AI hypothesis generation may mirror or even exceed human thinking[25,26]. Ultimately, the multilayer abstraction inherent in LLM architecture may allow for the interpolation of new patterns that may not be mere regurgitations of training data, potentially presenting value to human researchers[27].

Novel hypotheses generated by AI evidently warrant post-facto experimental validation. History reminds us that even our greatest minds revised their views; Einstein himself abandoned the cosmological constant he once defended[28]. Recognizing that human reasoning is likewise fallible strengthens the case for subjecting both human and AI-generated hypotheses to the same empirical scrutiny. But by recognizing hidden relationships or correlations across disciplines, LLMs could guide the design of experiments that would otherwise remain unexplored, steering researchers toward new truths. Recent work has shown that LLMs tuned with domain-specific knowledge outperform human neuroscientists in predicting experimental outcomes in neuroscience, likely due to the ability of models to integrate broad context and subtle patterns from extensive scientific literature[29]. Though AI may lack human-specific context, as models improve so too should machine reasoning, perhaps enabling AI to make causal inferences of comparable quality to even the best of humanity's thinkers. Ongoing efforts seek to replicate the thought processes of specific notable humans by fine-tuning LLMs on the sum total of an individual's works, though results remain mixed[30].

Indeed, AI Co-PIs may not be universally generalizable research tools as different fields of science warrant different approaches to innovation. Biology and physics fundamentally differ in the manner by which hypotheses, experiments, and results lead to new understanding. Physics is a "hard" science, where our experiments reveal progressively more about the fixed laws governing the unchanging universe. Biology, by contrast, is riddled with variability and heterogeneity from the cellular to organismal to societal levels making experiment control notoriously difficult. For example,

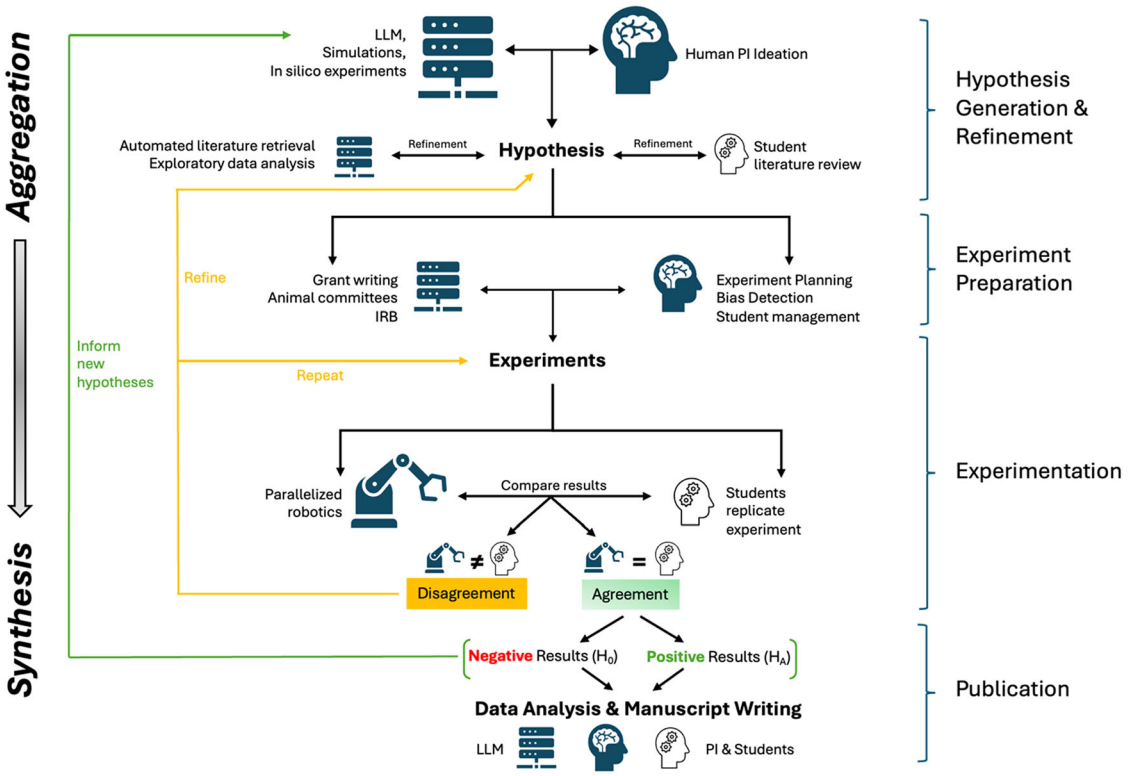

Fig. 1 | AI collaborators for efficient aggregation and synthesis in biomedical discovery.

the "antibody problem" refers to the paradox where even slight differences in experimental conditions—such as reagent lots, cell lines, or ambient conditions—can cause radically divergent antibody binding behaviors, making reproducibility an ongoing challenge.

A critic might argue that an AI Co-PI would be unable to make scientific progress under conditions of such variability in the biological sciences. However, AI-guided experiments and parallelized robotics might overcome variance through superior iterability and repetition. By scaling up the number and sophistication of experimental trials, AI-guided robotics can systematically probe these loose, variance-prone problems in biology until approximations become reliable truths. We already approximate this idea with high-throughput "omics" technologies, drug discovery, protein folding, and other contexts, illustrating how large-scale data generation can capture complexity that would overwhelm traditional human-led methods[13,14,31]. AI has demonstrated its ability to anticipate circumstantial challenges and adapt its strategies in other highly variable contexts as well, including disaster management and fraud prevention[32,33]. Variability and heterogeneity may reinforce, rather than undermine, the value of in-silico approaches for AI-PI collaborations in biomedical discovery.

Thinking scientifically

Scientific research hinges on transparency—replication is the cornerstone of confidence in the scientific process. A "black box" refers to a system where inputs and outputs can be observed, but internal mechanisms and interim stages remain hidden or incomprehensible. Many argue AI's "black box" nature is an insurmountable obstacle since we rarely grasp its precise reasoning and thought process, thus undermining the validity of the output.

For years, AI progress relied on pretraining scaling—training LLMs on ever-larger datasets at escalating computational costs[34,35]. However, new models like OpenAI's o1 and o3 and DeepSeek R1 introduce inference-time scaling (ITS), dynamically allocating reasoning power[36,37]. Rather than improving through scale, these models "think harder" at inference time, using chain-of-thought reasoning to refine outputs iteratively[38]. This mirrors human problem-solving, where complex tasks are broken into logical steps to improve accuracy[39]. With ITS and expanding agentic behaviors (e.g., function-calling, tool use), LLMs are approaching autonomous reasoning which further blurs the line between artificial and human cognition. However, when combined with "explainable AI" (XAI) methods including SHAP analysis, LIME, and counterfactual explanations, the next generation of models may be both smarter and more interpretable. This would be compelling justification for greater uptake of AI in academia, where scientists are rightly concerned about black box models leading knowledge generation astray.

Alternatively, AI's lack of a visible thought process may be perceived as an advantage. Human cognition itself is a black box, as opaque as a neural network[40]. With 100 billion neurons and 100 trillion synapses, we cannot map our own reasoning precisely, yet we accept it based on external markers like experience and knowledge. If an AI generates a novel hypothesis, its origin may be irrelevant so long as it is eventually empirically validated. Paradoxically, AI's opacity could be an advantage by enabling cognition beyond human intuition, such as finding a shortcut between A and C through a new dimension that skips B entirely, as has been argued previously[41]. The black-box critique of AI as PI may need reframing: rather than a flawed imitation of human scientists, AI should be seen as a fundamentally different, assistive scientific agent.

Cautious optimism

Though aggregation-focused researchers may worry that LLMs will nullify the substantial effort invested into chart reviews, data extraction, and discussion sections, they should remain optimistic that these advancements may drive unprecedented improvements in biomedical research. If open-source tools can conduct aggregation-type research faster and more accurately than any human, the incentive to repeatedly re-review literature diminishes. This shift would free researchers to instead focus on hypothesis generation and interpretation of unexpected findings. In an era of uncertain government funding, improving efficiency in publicly funded biomedical research is both prudent and necessary.

As AI potentially progresses toward Artificial General Intelligence (AGI)—where models approach human-level performance across disciplines—its role in research will expand. Some define AGI by financial benchmarks, while others suggest it will function like a "country of geniuses in a data center."[42,43] Google has released an "AI Co-scientist" on their Gemini 2.0 model which is designed to accomplish many of the tasks in the framework above, including hypothesis generation, with a focus on drug-target interactions[44]. Yet full autonomy in scientific inquiry remains distant, and AI should be viewed not as a replacement but as an amplifier of human ingenuity.

Of potential concern, recent work led by MIT researchers suggests that sustained reliance on LLMs in academic contexts may attenuate the neural circuits which underwrite rigorous scientific thought[45]. Study participants who drafted essays with ChatGPT showed markedly weaker alpha and beta-band connectivity and struggled to recall their own writing after the tool was withdrawn; follow-up analyses revealed that early AI dependence produced shallow semantic encoding and poorer quotation accuracy. Any research framework that positions LLMs as co-investigators must include phased human-AI interaction and explicit incentives for independent reasoning before the machine is engaged.

Beyond individual effects, unmoderated AI poses broader societal risks as well. We must attempt to continually remove bias from AI training data and tune model outputs with objective standards of accuracy. If a hidden error slips into one model's output and becomes accepted as "truth," subsequent black-box models may repeatedly propagate and amplify that mistake, leading to far-reaching scientific inaccuracies. Until the scientific community reaches consensus on how to evaluate the accuracy of AI-derived results, researchers and journal editors face a serious challenge in verifying—and potentially rectifying—propagated errors. Machines without safeguards cannot protect scientific integrity or maintain patient safety.

While our perspective highlights significant opportunities for LLM integration in medical research, our analysis is limited in several ways. The practical implementation of LLMs as research collaborators has yet to be broadly validated in clinical or laboratory settings, and empirical studies evaluating AI-generated hypotheses remain sparse. Further, although we raise open questions surrounding ethical considerations, the practical resolution of regulatory and institutional barriers to widespread adoption remains uncertain and warrants careful consideration.

Our technological achievements as a species have been built on the scientific method and a communal definition of scientific integrity. It is this same integrity that will allow us to responsibly nurture the growth of AI within academia. As those with the final say, we must continue to scrutinize the work of our new "colleagues" with the same rigor and standards that we maintain for ourselves. We should guide our newfound collaborators away from error and bias and towards truth and discovery—akin to how an excellent PI mentors an eager young medical student.

Ultimately, against our habits and perhaps our human nature, we should strive to be agnostic to the origin of a great idea.

Data availability

No datasets were generated or analyzed during the current study.

Received: 15 April 2025; Accepted: 2 July 2025;
Published online: 14 July 2025

References

1. Martinez, V. H. et al. The competitiveness of orthopaedic surgery residency programs: a twenty-year analysis utilizing a normalized competitive index. *Surg. Pract. Sci.* **12**, 100155 (2023).
2. National Resident Matching Program. Charting outcomes in the match: Characteristics of applicants who matched to their preferred specialty: 2024 Main Residency Match. https://www.nrmp.org (2024).
3. National Resident Matching Program. Charting outcomes in the match: Characteristics of applicants who matched to their preferred specialty: 2009 Main Residency Match. https://www.nrmp.org (2009).
4. Wickramasinghe, D. P., Perera, C. S., Senarathna, S. & Samarasekera, D. N. Patterns and trends of medical student research. *BMC Med. Educ.* **13**, 175 (2013).
5. Bowers, C. A. et al. Arms race control score standardizes residency applicant publication assessment. *Neurosurgery* **97**, 250–258 (2025).
6. Lála, J. et al. PaperQA: retrieval-augmented generative agent for scientific research. Preprint at https://arxiv.org/abs/2312.07559 (2023).
7. Skarlinski, M. D. et al. Language agents achieve superhuman synthesis of scientific knowledge. Preprint at https://arxiv.org/abs/2409.13740 (2024).
8. O'Donoghue, O. et al. BioPlanner: automatic evaluation of LLMs on protocol planning in biology. Preprint at https://arxiv.org/abs/2310.10632 (2023).
9. Puljak, L. & Lund, H. Definition, harms, and prevention of redundant systematic reviews. *Syst. Rev.* **12**, 63 (2023).
10. Rezaeikhonakdar, D. AI chatbots and challenges of HIPAA compliance for AI developers and vendors. *J. Law Med. Ethics* **51**, 988–995 (2023).
11. Weissman, G. E., Mankowitz, T. & Kanter, G. P. Unregulated large language models produce medical device-like output. *npj Digit. Med.* **8**, 148 (2025).
12. Jonnagaddala, J. & Wong, Z. S. Y. Privacy preserving strategies for electronic health records in the era of large language models. *npj Digit. Med.* **8**, 34 (2025).
13. Gao, S. et al. Empowering biomedical discovery with AI agents. *Cell* **187**, 6125–6151 (2024).
14. Varadi, M. & Velankar, S. The impact of AlphaFold Protein Structure Database on the fields of life sciences. *Proteomics* **23**, 2200128 (2023).
15. Zhang, K. et al. Artificial intelligence in drug development. *Nat. Med.* **31**, 45–59 (2025).
16. Liu, G. et al. Deep learning-guided discovery of an antibiotic targeting *Acinetobacter baumannii*. *Nat. Chem. Biol.* **19**, 1342–1350 (2023).
17. Hamm, J. et al. A modular robotic platform for biological research: cell culture automation and remote experimentation. *Adv. Intell. Syst.* **6**, 2300566 (2024).
18. Arnold, C. Cloud labs: where robots do the research. *Nature* **606**, 612–613 (2022).
19. van Andel, P. E. K. Anatomy of the unsought finding. Serendipity: origin, history, domains, traditions, appearances, patterns and programmability. *Br. J. Philos. Sci.* **45**, 631–648 (1994).
20. Langmuir, I. The role of serendipity in research: Project Cirrus and the art of profiting from unexpected occurrences. *Bull. Am. Meteorol. Soc.* **77**, 1279–1284 (1941).
21. Gibney, E. Has your paper been used to train an AI model? Almost certainly. *Nature* **632**, 715–716 (2024).
22. Abgrall, G., Monnet, X. & Arora, A. Synthetic data and health privacy. *JAMA* **333**, 567–568 (2025).
23. Hubert, K. F., Awa, K. N. & Zabelina, D. L. The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Sci. Rep.* **14**, 3440 (2024).
24. Rabby, G. et al. Iterative hypothesis generation for scientific discovery with Monte Carlo Nash equilibrium self-refining trees. Preprint at https://arxiv.org/abs/2503.19309 (2025).
25. Suri, G. et al. Do large language models show decision heuristics similar to humans? A case study using GPT-3.5. Preprint at https://arxiv.org/abs/2305.04400 (2023).
26. Nie, A. et al. MOCA: measuring human-language model alignment on causal and moral judgment tasks. Preprint at https://arxiv.org/abs/2310.19677 (2023).
27. Liu, H. et al. Literature meets data: a synergistic approach to hypothesis generation. Preprint at https://arxiv.org/abs/2410.17309 (2024).
28. O'Raifeartaigh, C. et al. Einstein's 1917 static model of the universe: a centennial review. *EPJ H.* **42**, 431–474 (2017).

29. Luo, X. et al. Large language models surpass human experts in predicting neuroscience results. *Nat. Hum. Behav.* **9**, 305–315 (2025).
30. Schwitzgebel, E., Schwitzgebel, D. & Strasser, A. Creating a large language model of a philosopher. *Mind Lang.* **39**, 237–259 (2024).
31. Qian, L. et al. AI-empowered perturbation proteomics for complex biological systems. *Cell Genomics* **4**, 11 (2024).
32. Sun, W., Bocchini, P. & Davison, B. D. Applications of artificial intelligence for disaster management. *Nat. Hazards* **103**, 2631–2689 (2020).
33. Bello, O. A. & Olufemi, K. Artificial intelligence in fraud prevention: exploring techniques and applications challenges and opportunities. *Comp. Sci. IT Res. J.* **5**, 1505–1520 (2024).
34. Pearce, T. et al. Scaling laws for pre-training agents and world models. Preprint at https://arxiv.org/abs/2411.04434 (2024).
35. Chowdhury, H. Nvidia boss Jensen Huang predicts computing power will increase a "millionfold" in a decade. *Business Insider*. https://www.businessinsider.com/nvidia-jensen-huang-predicts-increase-computing-power-ai-scaling-2024-11 (2024).
36. Kaplan, J. et al. Scaling laws for neural language models. Preprint at https://arxiv.org/abs/2001.08361 (2020).
37. Snell, C., Lee, J., Xu, K. & Kumar, A. Scaling LLM test-time compute optimally can be more effective than scaling model parameters. Preprint at https://arxiv.org/abs/2408.03314 (2024).
38. Wei, J. et al. Chain-of-thought prompting elicits reasoning in large language models. *Adv. Neural Inf. Process. Syst.* **35**, 24824–24837 (2022).
39. Simon, H. A. & Newell, A. Human problem solving: the state of the theory in 1970. *Am. Psychol.* **26**, 145–159 (1971).
40. Bonezzi, A., Ostinelli, M. & Melzner, J. The human black-box: the illusion of understanding humans better than algorithmic decision-making. *J. Exp. Psychol. Gen.* **151**, 2250–2262 (2022).
41. Duede, E. Deep learning opacity in scientific discovery. *Philos. Sci.* **90**, 1089–1099 (2023).
42. Perrigo, B. How OpenAI's Sam Altman is thinking about AGI and superintelligence. *TIME*. https://time.com/7205596/sam-altman-superintelligence-agi/ (2024).
43. Varanasi, L. Here's how Anthropic CEO Dario Amodei defines artificial general intelligence. *Business Insider*. https://www.businessinsider.com/how-anthropic-ceo-dario-amodei-defines-artificial-general-intelligence-2024-10 (2024).
44. Gottweis, J. et al. Towards an AI co-scientist. Preprint at https://arxiv.org/abs/2502.18864 (2025).
45. Kosmyna, N. et al. Your brain on ChatGPT: accumulation of cognitive debt when using an AI assistant for essay writing task. Preprint at https://arxiv.org/abs/2506.08872 (2025).

Acknowledgements

We acknowledge Jacob Levy and Graham Michelson for stimulating and refining our thoughts on these matters. We further acknowledge the Northwestern Feinberg Functional Neurosurgery Research Group (FRG).

Author contributions

D.P. and C.A. were responsible for conceptualization. D.P., A.K., S.S., R.J., N.D., M.L., Y.L., and C.A. drafted and edited the manuscript. All authors read and approved the final manuscript before submission.

Competing interests

Y.L. is a member of the NPJ Digital Medicine Editorial Board. He was not part of peer review nor decision making of the manuscript. The authors have no other competing interests to declare. No funding was received for this work.

Additional information

Correspondence and requests for materials should be addressed to Dillan Prasad.