From Knowledge to Discovery: Harnessing LLMs as Co-pilots in modern scientific research

Artificial Intelligence and scientific research

Artificial Intelligence,as a new general-purpose “method of invention” is poised to reshape the innovation process and the organization of R&D in both academic and commercial research [1].In order to drive widespread diffusion of AI in scientific research there are a number of significant challenges in software, algorithm design, workflows and data ecosystems that will need to be solved through co-ordinated public and private sector partnerships and a scientific community centric approach [2][3] .One specific AI algorithm that stands to play a pivotal role in this transformation is Large Language Models (LLMs) and its potential as a scientific-advisor / co-pilot for the end-to-end Research workflow.

The research workflow and the role of LLMs

The end to end research workflow, while diverse across domains, can be distilled into five core segments based on their function:

Knowledge Assimilation and Dissemination: Extract insights from the existing body of knowledge, identify novel research directions, and subsequently contribute new findings back.
Hypotheses Generation: Analyze scientific literature and proprietary data to form new hypotheses
Analytical Coordination: Orchestrate complex analytical and experimental tasks for hypothesis testing
Inference and Validation: Synthesize insights and validate robustness of findings
Discovery: Support the inclusion and replication of new insights as new additions to body of knowledge

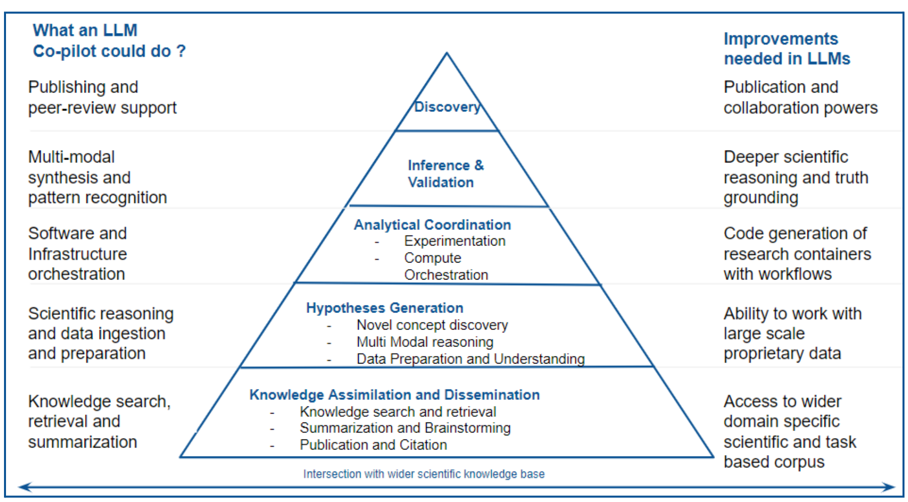

Figure1 : The different stages of the research workflow can be visualized as a pyramid with the tapering size of the layers corresponding to the current impact of AI based technologies in each layer. Additionally the opportunities and improvement areas for LLM based copilots in each layer are shown.

Author - Astitva Chopra (astitvac@gmail.com)

While various AI technologies could enhance discrete stages of this workflow, large language models (LLMs) stand apart in their potential as an integrated research co-pilot across the entire process. As LLMs scale, they exhibit strengthened reasoning, information retrieval, and domain adaptability [4][6], suggesting promise as a versatile research assistant. Recent work already demonstrates LLMs' capabilities for activities like literature synthesis, data analysis, and text generation. With systematic advancements in training, fine-tuning and domain specific task adaptation, LLMs could become a singular "Jarvis" [5] like AI agent capable of turbocharging scientific productivity.

LLMs as co-pilots for research : SoTA and Opportunities

To fully harness LLMs as scientific assistants their core cognitive capabilities, such as logical reasoning, knowledge mining and truth grounding must be enhanced. Recent research and improvements in training data curation, finetuning and hybrid architectures have already shown promising results with directional improvements. Below we lay out the current SoTA and further research directions for each stage of the research workflow.

Stages of Scientific Research	Current SoTA	Improvement needed and potential research and solutioning directions
Knowledge Assimilation and Dissemination	Advanced search and retrieval capabilities Automated understanding summarization of multi-modal scientific literature Eg. **GPT-4**, **T5**, **Nougat**, **MedPaLM 2**	- Broadening context and working memory dimensions for effective knowledge extraction and integration with Knowledge Graphs [23] - Continued improvements in scientific reasoning and truth grounding through curated datasets[14] and novel prompting and refinement approaches. [15]
Insights Generation	Pattern recognition in vast literature Concept Discovery Eg. **SciBERT**, **ChemBERTa**,	- Improvements in reasoning through enhanced prompting (CoT,Self Consistency) [11] and improved task adapted benchmarks[22] - Utilizing innovative benchmarking tools for assessing scientific and mathematical reasoning capabilities during literature traversal [12]. - Merging external and proprietary datasets for expanded context awareness

 Author - Astitva Chopra (astitvac@gmail.com)

Analytical Coordination	Automated task orchestration Integration with computational tools and software Eg. **Langchain**, **Code Llama**, **WizardLM**	- Ability to integrate the increasing diversity of narrow AI models [3] and benchmarks [21] showing promise across different research domains - Utilization of scientific containerization [18] for orchestrating portable and repeatable research workflows - Ability to work stitch together external and open source datasets and models to allow for research exploration (Eg.Hugging Face model cards) - Continued improvements in code generation through enhanced fine-tuning, like EVOL-instruct [8] - Enhancing LLM connections with APIs [7] and Integration with experimental apparatus [13] for specialized computation or experimentation.
Inference and Validation	Statistical analysis capabilities through plugins Automated validation against existing knowledge base Eg. **GPT-4 Code Interpreter, LLM enabled Wolfram language**	- Training and evaluation on specialized scientific datasets and notations for advance pattern recognition[19] - Enhanced truth grounding (Possibly through a scientific computation language underpinning logical reasoning) [16] - Improvements in LLMs as Knowledge Bases and Retrieval Augmented Generation [9][10]
Discovery		- Improving models explainability [24] to validate inferences and ensure consistency with wider knowledge base - More work on general and transferable learning across domains and human aligned reasoning and ethics [25] - Support for peer-review process and replication

Partners - Ecosystem and Technology

Large language models (LLMs) represent a pivotal intermediary layer in integrating artificial intelligence into scientific research, with the potential to streamline many aspects of the research workflow. However, LLMs alone cannot achieve comprehensive assimilation of AI capabilities into the scientific domain. Academic research is an intrinsically collaborative endeavor, necessitating a collective approach for transformative impacts. Two key partner categories are vital for this integration:

Ecosystem partners, including funding agencies, research institutions, data providers, and academic publishers, are crucial for developing universally applicable AI tools tailored to researchers' needs and workflows. Engaging these entities will ensure AI capabilities align with real-world scientific workflows.[17][20]

Technology partners such as computational libraries, software frameworks, and infrastructure providers already play a crucial role in research computing. Partnering with these stakeholders is essential for constructing a holistic AI-for-science ecosystem. Integrating LLMs with existing scientific computing stacks will maximize interoperability and utility.

About the Author

Astitva is a member of the research computing team at UC Riverside. His journey began as a consultant, supporting the automation and digital transformation needs of diverse customers .He then spent almost a decade working at Google, helping integrate AI across various business and product domains. He led the development of an ML-based lead generation system and deployment of a conversational intelligence platform for the global ads business.

In recent years, he has shifted his focus to scientific computing and he played a critical role in the design of research computing solutions for Google Cloud as well as developing external partnerships.He co-founded the GCP Research Innovators program and played a critical role in the development of RAD Lab, both aimed at enhancing community-driven research. Currently, at UC Riverside, he is involved in the creation of an LLM-based lab assistant, reflecting his ongoing interest in merging AI with research methodologies.

Appendix

References

1. Cockburn, Iain M. et al. "The Impact of Artificial Intelligence on Innovation." IRPN: Innovation & Cyber Law & Policy (Topic) (2018)

2. Advanced Research Directions on AI for Science, Energy, and Security : Report on the U.S. Department of Energy (DOE) Summer 2022 Workshop Series on Artificial Intelligence (AI) for Science, Energy, and Security

3. Wang, Hanchen et al. "Scientific discovery in the age of artificial intelligence." Nature 620 (2023):

4. Brown, Tom B. et al. “Language Models are Few-Shot Learners.” ArXiv abs/2005.14165 (2020):

5. Zheng, Kai et al. “JARVIS: A Neuro-Symbolic Commonsense Reasoning Framework for Conversational Embodied Agents.” *ArXiv* abs/2208.13266 (2022)

6. Lee, Jinhyuk et al. “BioBERT: a pre-trained biomedical language representation model for biomedical text mining.” Bioinformatics 36 (2019): 1234 - 1240.

7. Liang, Yaobo et al. “TaskMatrix.AI: Completing Tasks by Connecting Foundation Models with Millions of APIs.” ArXiv abs/2303.16434 (2023)

8. Luo, Ziyang et al. “WizardCoder: Empowering Code Large Language Models with Evol-Instruct.” ArXiv abs/2306.08568 (2023)

9. Petroni, Fabio et al. “Language Models as Knowledge Bases?” ArXiv abs/1909.01066 (2019)

10. Lewis, Patrick et al. “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.” *ArXiv* abs/2005.11401 (2020)

11. Wang, Xuezhi et al. “Self-Consistency Improves Chain of Thought Reasoning in Language Models.” *ArXiv* abs/2203.11171 (2022): n. Pag.

12. *Papers with Code - SVAMP Benchmark (Math Word Problem Solving)*. (n.d.). *Papers with Code - SVAMP Benchmark (Math Word Problem Solving)*. https://paperswithcode.com/sota/math-word-problem-solving-on-svamp

13. Huang, Wenlong et al. “Language Models as Zero-Shot Planners: Extracting Actionable Knowledge for Embodied Agents.” *ArXiv* abs/2201.07207 (2022): n. Pag.

14. Taylor, Ross et al. “Galactica: A Large Language Model for Science.” *ArXiv* abs/2211.09085 (2022): n. pag.

15. Singhal, K. et al. “Towards Expert-Level Medical Question Answering with Large Language Models.” *ArXiv* abs/2305.09617 (2023): n. Pag.

16. Stephen Wolfram (2023), "ChatGPT Gets Its 'Wolfram Superpowers'!," Stephen Wolfram Writings. writings.stephenwolfram.com/2023/03/chatgpt-gets-its-wolfram-superpowers

17. (2023, January 3). National AI Research Resource (NAIRR) Task Force. https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf

18. Kurtzer, Gregory M. et al. "Singularity: Scientific containers for mobility of compute." *PLoS ONE* 12 (2017): n. Pag.

19. Weininger, David. "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules." *J. Chem. Inf. Comput. Sci.* 28 (1988): 31-36.

20. Tenopir, Carol et al. "Data Sharing by Scientists: Practices and Perceptions." *PLoS ONE* 6 (2011): n. Pag.

21. Thiyagalingam, Jeyan et al. "Scientific machine learning benchmarks." *Nature Reviews Physics* 4 (2021): 413 - 420.

22. Jin, Qiao et al. "PubMedQA: A Dataset for Biomedical Research Question Answering." *Conference on Empirical Methods in Natural Language Processing* (2019).

23. Pan, Shirui et al. "Unifying Large Language Models and Knowledge Graphs: A Roadmap." *ArXiv* abs/2306.08302 (2023): n. pag.

24. Arrieta, Alejandro Barredo et al. "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI." *ArXiv* abs/1910.10045 (2019): n. Pag.

25. Bubeck, Sébastien et al. "Sparks of Artificial General Intelligence: Early experiments with GPT-4." *ArXiv* abs/2303.12712 (2023): n. pag.