

Towards Agentic AI for Science: Hypothesis Generation, Comprehension, Quantification, and Validation

Lifu Huang

UC Davis
Davis, California, USA
lfuhuang@ucdavis.edu

Danai Koutra

University of Michigan
Ann Arbor, Michigan, USA
dkoutra@umich.edu

Adithya Kulkarni

Virginia Tech
Blacksburg, Virginia, USA
aditkulk@vt.edu

Temiloluwa Prioleau

Dartmouth College
Hanover, New Hampshire, USA
tprioleau@dartmouth.edu

Qingyun Wu

Pennsylvania State University
University Park, Pennsylvania, USA
qingyun.wu@psu.edu

Yujun Yan

Dartmouth College
Hanover, New Hampshire, USA
yujun.yan@dartmouth.edu

Yaoqing Yang

Dartmouth College
Hanover, New Hampshire, USA
yaoqing.yang@dartmouth.edu

James Zou

Stanford University
Stanford, California, USA
jamesz@stanford.edu

Dawei Zhou

Virginia Tech
Blacksburg, Virginia, USA
zhoud@vt.edu

CCS Concepts

• **Computing methodologies** → **Knowledge representation and reasoning**; • **Information systems** → *Decision support systems*.

Keywords

Agentic AI, Hypothesis generation, Hypothesis validation

ACM Reference Format:

Lifu Huang, Danai Koutra, Adithya Kulkarni, Temiloluwa Prioleau, Qingyun Wu, Yujun Yan, Yaoqing Yang, James Zou, and Dawei Zhou. 2025. Towards Agentic AI for Science: Hypothesis Generation, Comprehension, Quantification, and Validation. In *Companion Proceedings of the ACM Web Conference 2025 (WWW Companion '25)*, April 28-May 2, 2025, Sydney, NSW, Australia. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3701716.3717754>

1 Overview of the Workshop

AI models for science [14] have the potential to harness large datasets, accelerate scientific discoveries and transform numerous fields. For example, AlphaFold [13] has accurately predicted protein structures, significantly reducing research time in biology and medicine. Similarly, AI has been used to identify compounds for diseases like COVID-19 [10] and Cancer [21], dramatically shortening the timelines for breakthroughs in healthcare. In mathematics, AlphaTensor [7] has discovered faster algorithms for matrix-matrix multiplication than Strassen's algorithm. AI systems have been applied in fields such as healthcare [10, 21], material science [15], physics [19], mathematics [7], education [16], psychology [18], and software development [20]. However, these models are often developed ad-hoc. In contrast, agentic AI [14] has recently emerged as a

systematic approach to advancing scientific discovery, offering the ability to autonomously perform complex tasks and interact with the researchers.

To enhance the scientific discovery process with AI, it is essential to first understand the key stages where AI-driven solutions can have a substantial impact. Scientific discovery typically involves several key steps [11]: hypothesis generation, quantification, experimentation, data analysis, and verification. Each of these stages is crucial for building a robust body of knowledge. Hypothesis generation involves formulating potential explanations or predictions, quantification expresses these ideas in measurable terms, experimentation gathers data under controlled conditions, data analysis extracts meaningful insights, and verification confirms findings through replication or additional evidence. Each step requires substantial human effort, which makes AI-driven solutions highly appealing [14]. However, integrating AI into these processes presents several challenges [2]. In hypothesis generation, AI may struggle to propose truly novel ideas beyond what it has been trained on [17]. Quantification is difficult when it comes to accurately modeling complex phenomena [5]. Experimentation raises concerns as AI might suggest impractical experiments or overlook critical variables [6]. In data analysis, the risks include overfitting, bias, and misinterpreting correlations [4]. Finally, verification is complicated by issues of replicability and consistency, making it harder to validate AI-driven results [8]. Overcoming these challenges is essential for AI to effectively contribute to the advancement of scientific discovery.

Our mission is to foster interdisciplinary collaboration to develop fully autonomous AI systems, addressing challenges like benchmark datasets, human-AI collaboration, robust tools and methods for validating AI outputs, and trustworthiness. By tackling these issues, we can unlock AI's transformative potential in research. In this workshop, themed **Agentic AI for Science**, we will explore these critical topics and welcome diverse perspectives. We will focus on integrating agentic AI systems to enhance scientific discovery while upholding rigorous standards. For AI to contribute effectively, it must **generate novel hypotheses, comprehend**

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WWW Companion '25, April 28-May 2, 2025, Sydney, NSW, Australia

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1331-6/2025/04

<https://doi.org/10.1145/3701716.3717754>

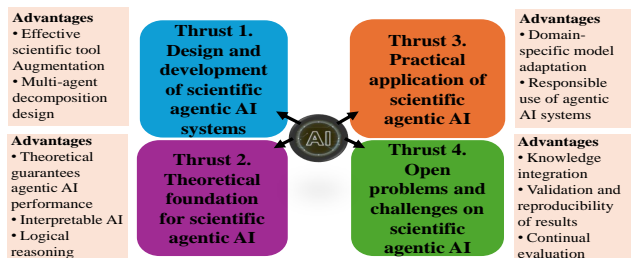


Figure 1: Illustrative figure showcasing the research thrusts.

their applications, quantify testing resources, and validate feasibility through well-designed experiments. This workshop serves as a vital forum for collaboration and knowledge-sharing aimed at redefining the landscape of scientific discovery. This workshop aims to address four main research thrusts to propel future research, including (non-exclusively):

Thrust 1. Design and development of agentic AI systems for scientific discovery. The emergence of agentic AI, particularly through foundation and generative models, presents significant opportunities for scientific discovery by revolutionizing processes like hypothesis generation and validation. Notable projects such as ChemCrow [3] in chemistry, Crispr-GPT [12] in genetic engineering, and SciAgents [9] in multi-agent systems highlight the potential of agentic AI to accelerate scientific progress. Key research areas include developing tailored scientific foundation models, enhancing existing scientific tools with AI capabilities, designing frameworks for multi-agent hypothesis generation, and creating human-in-the-loop systems to improve the reliability and interpretability of AI-driven discoveries.

Thrust 2. Theoretical foundation for scientific agentic AI. Developing agentic scientific AI requires robust methods for quantifying predictions, assessing performance, and validating generated hypotheses, supported by solid theoretical foundations. This includes statistical models addressing in-context learning, multi-agent communication, game theory, and optimization constraints. Key topics encompass various reasoning approaches such as inductive, deductive, and abductive as well as Bayesian reasoning and neural-symbolic methods. Additionally, we will focus on developing theory-driven metrics for model quantification and validation, creating benchmarks for evaluating AI-generated hypotheses, and enhancing interpretability to ensure trust, safety, and transparency in agentic AI systems.

Thrust 3. Practical application of scientific agentic AI. The deployment of agentic AI systems across scientific fields faces significant challenges due to the necessity for domain-specific adaptations, such as unique data formats and model constraints. Key concerns include addressing bias in training data, especially in sensitive areas like medicine, as well as ensuring the trustworthiness and explainability of AI-generated hypotheses. Ethical considerations are crucial when automating research decisions that impact public health or policy, underscoring the importance of responsible AI deployment. Key focus areas include adapting models for specific domains, detecting and mitigating bias, enhancing transparency and interpretability, and developing governance frameworks to ensure accountability in AI-assisted research.

Thrust 4. Open problems and challenges on scientific agentic AI. Despite the significant promise of agentic AI for scientific discovery, several challenges remain unresolved. Key issues include the automatic curation of domain-specific knowledge and its integration into agentic AI systems, as well as developing advanced mechanisms for scalable and efficient multi-agent collaboration. Additionally, there is a need for continual evolution of these systems, focusing on effective model updates and performance improvements based on new data and discoveries. Finally, ensuring the validation and reproducibility of results generated by agentic AI systems remains a critical challenge.

This workshop aims to bring together researchers and practitioners from a range of domains to address challenges in Agentic AI for scientific discovery. It will feature experts in computer science, particularly those specializing in web technologies, machine learning, and natural language processing, to enhance AI’s capability to analyze data and simulate human reasoning. Additionally, researchers in human-computer interaction will focus on creating user-friendly software platforms and effective collaboration frameworks. By fostering interdisciplinary collaboration among these diverse fields, the workshop seeks to identify gaps and opportunities in the current landscape of Agentic AI research, promoting advancements that can benefit both academia and industry.

Related Prior Symposiums and Workshops To our knowledge, there have been a few remotely related symposiums and workshops, including “Integrated Approaches to Computational Scientific Discovery” scheduled for the 2024 Fall Symposium Series¹ and the “LLMs as Research Tools” workshop held alongside CHI 2024 [1]. Additionally, the AI4Science series has hosted several workshops, including the 1st AI4Science workshop at NeurIPS 2021, the 2nd at ICML 2022, the 3rd at NeurIPS 2022, the 4th at NeurIPS 2023, and the 5th at ICML 2024². While existing workshops and symposiums have contributed to AI in scientific discovery, our proposed workshop uniquely focuses on *agentic AI* for scientific discovery. The *agentic* nature introduces important considerations, such as interactive and autonomous decision-making, integration of diverse AI capabilities, human-AI agency, and multi-agent collaboration. By exploring these facets, our workshop aims to advance AI-driven scientific discovery beyond the boundaries set by previous events.

Contributions of the Workshop to The Web Conference 2025 The “Agentic AI for Science” workshop aligns perfectly with *The Web Conference 2025* by offering a timely platform for exploring the intersection of AI technologies and scientific discovery. By focusing on the design and application of agentic AI systems, the workshop will facilitate interdisciplinary collaboration among AI researchers, computer scientists, and ethicists, essential for addressing the complexities of web-driven research. This engaging environment will encourage the exchange of innovative ideas and practical solutions, promoting discussions on automating scientific processes while tackling critical technical and ethical challenges. The workshop will highlight diverse research contributions, showcasing AI’s transformative potential in scientific inquiry and reinforcing the relevance of web technologies in this field. Ultimately, it aims to enhance community engagement and foster a commitment to AI-driven

¹<http://cogsys.org/symposium/discovery-2024/>

²<https://ai4sciencecommunity.github.io/>

Table 1: Tentative Schedule

Sydney NSW Time (GMT+11)	Event
9:00-9:10	Opening Remarks
9:10-10:40	Two Keynote Speeches (40 min and 5 min QA each)
10:40-11:55	Oral paper session (12 min talk + 3 min QA)
11:55-12:10	Coffee Break
12:10-13:00	Speed Dating Workshop
13:00-14:30	Two Keynote Speeches (40 min and 5 min QA each)
14:30-15:30	Brainstorming and interaction session with speakers and organizers
15:30-15:45	Coffee Break
15:45-16:30	Poster Session (10-12 poster boards are required)
16:30-17:15	Panel Discussion
17:15-17:30	Closing remarks with awards

advancements, making it an ideal fit for the themes of The Web Conference 2025.

2 Proposed Schedule

Workshop logistics We propose a one-day workshop featuring at least four keynote speeches, along with presentations and poster sessions for accepted papers formatted to the Web Conference 2025 guidelines. The event will culminate in an awards ceremony. A tentative schedule is outlined in Table 1.

Encouraging Discussions among Attendees The workshop schedule is thoughtfully crafted to foster impactful discussions and collaborative opportunities. Strategic coffee breaks will encourage attendees to connect, share insights, and explore potential partnerships. With inspiring keynote speeches and a dedicated poster session, participants will have ample opportunities to spark new ideas. The interactive brainstorming session with speakers and organizers will facilitate deeper dialogues, while oral paper sessions will feature engaging presentations followed by dynamic Q&A segments. Additionally, the innovative speed dating workshop format will ensure that participants make meaningful connections. Through these diverse activities, attendees will not only advance the field of agentic AI for scientific discovery but also strengthen community ties and generate a vibrant exchange of ideas.

3 Workshop Attendees

Our workshop promises to attract a diverse and engaged audience, including our organizing team and a lineup of four speakers—two of whom are women—who bring invaluable expertise in “AI for Science.” We expect around **150 participants**, creating a dynamic environment for collaboration and innovation.

Audience Engagement Plan To ensure strong attendance, we have crafted a robust outreach strategy that leverages social media and established academic networks. We will promote the workshop on popular platforms like X (formerly Twitter), r/MachineLearning, and Facebook, as well as through networks like WikiCFP. Our efforts will extend to affinity groups such as LXAI, Black in AI, MusML, Queer in AI, and the Women in Machine Learning Group, ensuring we attract a diverse and inclusive audience. With our

program committee collectively reaching around 10,000 followers, we will maximize our visibility and engagement.

4 Diversity Commitment

Our workshop embraces a rich diversity of topics within Agentic AI research, spanning areas such as natural language processing and human-computer interaction, which fosters innovation and equips attendees with a holistic understanding of the field. The organizing committee, comprised of talented members from six prestigious institutions including Stanford University, University of Michigan, UC Davis, Virginia Tech, Penn State, and Dartmouth College brings a global perspective, enriched by varied experiences across four continents and multiple disciplines. This collaboration cultivates an environment where diverse ideas can thrive and lead to groundbreaking advancements. Additionally, our impressive lineup of invited speakers from top-tier institutions, including Stanford University, UC Berkeley, the University of Washington, the Kempner Institute, represents a range of backgrounds, inspiring participants to pursue their passions and champion equality in academia. We are also dedicated to engaging a diverse participant base, fostering a vibrant exchange of ideas and nurturing the next generation of leaders and innovators in AI. Together, we aim to ignite curiosity, inspire collaboration, and empower every participant to make a significant impact in their fields.

5 Keynote Speakers

Prof. Markus J. Buehler (**confirmed**): McAfee Professor of Engineering at MIT, Dr. Buehler is renowned for pioneering advanced biomaterials research. He has authored over 450 peer-reviewed publications, with an H-index of 89 and approximately 30,000 citations. His book “Biomateriomics” explores the intersection of materials science with music and linguistics. Dr. Buehler has received prestigious awards such as the NSF CAREER Award, DARPA Young Faculty Award, and PECASE. He serves as the editor-in-chief of the “Journal of the Mechanical Behavior of Biomedical Materials” and has delivered over 400 invited talks, chaired numerous conferences, and organizes an annual materials research camp at MIT.

Prof. Michael W. Mahoney (**tentative**): Dr. Mahoney is a prominent researcher at UC Berkeley, specializing in algorithmic and statistical methods for large-scale data analysis in machine learning. His expertise covers randomized matrix algorithms, numerical linear algebra, and geometric network analysis, with applications across genetics, astronomy, and more. He holds leadership roles at the International Computer Science Institute and Lawrence Berkeley National Laboratory and directs the NSF-funded Foundations of Data Analysis Institute. Additionally, he organizes the biennial MMDS Workshops on data analysis algorithms.

Prof. Sanmi Koyejo (**tentative**): Dr. Koyejo, Assistant Professor at Stanford University and adjunct at the University of Illinois, leads the Stanford Trustworthy AI Research (STAIR) group. His work focuses on applying machine learning to neuroscience and healthcare. He holds a Ph.D. from the University of Texas at Austin and completed a postdoctoral fellowship at Stanford. Notable accolades include a Best Paper Award at the Conference on Uncertainty in Artificial Intelligence and a Sloan Fellowship. He collaborates with Google’s Brain team and is involved with the Neural Information

Processing Systems Foundation. As president of Black in AI, he advocates for diversity in AI, impacting both science and societal applications.

Prof. Marinka Zitnik (tentative): Assistant Professor at Harvard Medical School and Associate Faculty at the Kempner Institute, Dr. Zitnik's work in Biomedical Informatics spans affiliations with the Broad Institute and Harvard Data Science Initiative. Her research focuses on leveraging AI for scientific innovation and personalized medicine, with her algorithms used by leading institutions like Baylor College of Medicine and Stanford Medical School. She has received the Bayer Early Excellence in Science and NSF CAREER awards, among others. Dr. Zitnik, an ELLIS Scholar and Therapeutics Data Commons co-founder, also leads the AI4Science initiative and received Harvard's 2022 Young Mentor Award.

Why we invited these speakers for the workshop We are proud to invite an exceptional lineup of keynote speakers who exemplify our commitment to advancing Agentic AI for scientific discovery. **Dr. Markus J. Buehler** specializes in biomaterials, showcasing how AI can revolutionize materials science. **Dr. Michael W. Mahoney** offers insights into algorithmic and statistical analysis, linking machine learning with various scientific disciplines. **Dr. Sanmi Koyejo** will explore the implications of trustworthy AI in neuroscience and healthcare, emphasizing ethical aspects. Lastly, **Dr. Marinka Zitnik** will discuss her pioneering work in using algorithms to enhance medical diagnoses in personalized medicine. Together, these speakers will ignite transformative discussions on the future of Agentic AI in scientific discovery.

6 Organizers of the workshop

Our organizing committee comprises experts with strong technical knowledge, excellent organizational skills, and extensive connections across relevant communities. Each member brings valuable experience in organizing workshops, ensuring a successful event. **Prof. Lifu Huang** [UC Davis] (lfuhuang@ucdavis.edu) Google Scholar | Website: Known for his role in "WISE-Supervision" and AI4Research workshops at top AI conferences. **Prof. Danai Koutra** [University of Michigan] (dkoutra@umich.edu) Google Scholar | Website: Experienced in leading interdisciplinary workshops at major data mining events. **Adithya Kulkarni** [Virginia Tech] (aditkulk@vt.edu) Google Scholar | Website: Focuses on AI-driven hypothesis generation and validation. **Prof. Temiloluwa Prioleau** [Dartmouth College] (tprioleau@dartmouth.edu) Google Scholar | Website: Specializes in digital health and technology workshop organization. **Prof. Qingyun Wu:** [Pennsylvania State University] (qingyun.wu@psu.edu) Google Scholar | Website: Organizes workshops combining machine learning and AI. **Prof. Yujun Yan** [Dartmouth College] (yujun.yan@dartmouth.edu) Google Scholar | Website: Active in workshops and panels on network science and machine learning. **Prof. Yaoqing Yang** [Dartmouth College] (yaoqing.yang@dartmouth.edu) Google Scholar | Website: Served as area chair at NeurIPS and ICLR, promoting collaborative knowledge sharing. **Prof. James Zou** [Stanford University] (jamesz@stanford.edu) Google Scholar | Website: Organizes biomedical data science workshops. **Prof. Dawei Zhou** [Virginia Tech] (zhoud@vt.edu) Google Scholar | Website: Has organized various workshops, including (TrustLOG at WWW'24 and CIKM'22,

DLG@AAAI'21, and AI4FS@SDM'20), contributing to machine learning and AI fields.

Program Committee We expect about 20 submissions for our workshop and have prepared a thorough review process with a dedicated team of 13 Program Committee (PC) members from global institutions. Each member will handle up to three submissions, ensuring detailed feedback and accommodating any unforeseen changes. This diversity will enhance the quality of our evaluations. The confirmed PC members are Adithya Kulkarni (VT), James Zou (Stanford), Barry Menglong Yao (VT), Dawei Zhou (VT), Fatimah Alotaibi (VT), Lifu Huang (UC Davis), Longfeng Wu (VT), Mohna Chakraborty (U Mich.), Qingyun Wu (PSU), Temiloluwa Prioleau (Dartmouth), Tong Zeng (VT), Yaoqing Yang (Dartmouth), Yujun Yan (Dartmouth). The program committee will collaborate to accept high-quality papers that meet *Web Conference* standards for the workshop.

References

- [1] Marianne Aubin Le Quéré and et al. 2024. LLMs as Research Tools: Applications and Evaluations in HCI Data Work. In *Ext. Abstr. 2024 CHI Conf. on Human Factors in Comput. Syst. (CHI EA '24)*. ACM, New York, NY, USA, Article 479, 7 pages. <https://doi.org/10.1145/3613905.3636301>
- [2] Alexandre Blanco-Gonzalez and et al. 2023. The role of AI in drug discovery: challenges, opportunities, and strategies. *Pharmaceuticals* 16, 6 (2023), 891.
- [3] Andres M Bran, Sam Cox, and et al. 2023. ChemCrow: Augmenting large-language models with chemistry tools. *arXiv preprint arXiv:2304.05376* (2023).
- [4] Sathish Kumar Chintala. 2022. AI in public health: modelling disease spread and management strategies. *NeuroQuantology* 20, 8 (2022), 10830.
- [5] Cristina Cornelio and et al. 2023. Combining data and theory for derivable scientific discovery with AI-Descartes. *Nature Communications* 14, 1 (2023), 1777.
- [6] Jiafei Duan, Samson Yu, and et al. 2022. A survey of embodied ai: From simulators to research tasks. *IEEE Trans. Emerg. Top. Comput. Intell.* 6, 2 (2022), 230–244.
- [7] Alhussein Fawzi, Matej Balog, et al. 2022. Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature* 610, 7930 (2022), 47–53.
- [8] Raymond Fok and Daniel S Weld. 2023. In search of verifiability: Explanations rarely enable complementary performance in AI-advised decision making. *AI Magazine* (2023).
- [9] Alireza Ghafarollahi and Markus J Buehler. 2024. SciAgents: Automating scientific discovery through multi-agent intelligent graph reasoning. *arXiv preprint arXiv:2409.05556* (2024).
- [10] Aritra Ghosh, Maria M Larrondo-Petrie, and Mirjana Pavlovic. 2023. Revolutionizing vaccine development for COVID-19: a review of AI-based approaches. *Information* 14, 12 (2023), 665.
- [11] John E Hallsworth, Zulema Udaondo, et al. 2023. Scientific novelty beyond the experiment. *Microbial Biotechnology* 16, 6 (2023), 1131–1173.
- [12] Kaixuan Huang, Yuanhao Qu, et al. 2024. Crispr-GPT: An LLM agent for automated design of gene-editing experiments. *arXiv preprint arXiv:2404.18021* (2024).
- [13] John Jumper, Richard Evans, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *nature* 596, 7873 (2021), 583–589.
- [14] Chris Lu, Cong Lu, Robert Tjarko Lange, Jakob Foerster, Jeff Clune, and David Ha. 2024. The ai scientist: Towards fully automated open-ended scientific discovery. *arXiv preprint arXiv:2408.06292* (2024).
- [15] Ioannis Papadimitriou et al. 2024. AI methods in materials design, discovery and manufacturing: A review. *Comput. Mater. Sci.* 235 (2024), 112793.
- [16] Aditya Paul et al. 2024. AgentPeerTalk: Empowering Students through Agentic-AI-Driven Discernment of Bullying and Joking in Peer Interactions in Schools. arXiv:2408.01459 [cs.CY] <https://arxiv.org/abs/2408.01459>
- [17] Henri Poincaré. 2022. *The foundations of science: Science and hypothesis, the value of science, science and method*. DigiCat.
- [18] Huachuan Qiu and Zhenzhong Lan. 2024. Interactive Agents: Simulating Counselor-Client Psychological Counseling via Role-Playing LLM-to-LLM Interactions. *arXiv preprint arXiv:2408.15787* (2024).
- [19] Antonin Sulc et al. 2024. Towards Agentic AI on Particle Accelerators. arXiv:2409.06336 [physics.acc-ph] <https://arxiv.org/abs/2409.06336>
- [20] Jules White. 2024. Building Living Software Systems with Generative & Agentic AI. arXiv:2408.01768 [cs.SE] <https://arxiv.org/abs/2408.01768>
- [21] Wan-Ying Zhang et al. 2024. Revolutionizing adjuvant development: harnessing AI for next-generation cancer vaccines. *Front. Immunol.* 15 (2024), 1438030.