

Conference Report

Open Access



Unlocking the future of materials science: key insights from the DCTMD workshop

Rika Kobayashi^{1,*} , Roger D. Amos², T. Daniel Crawford^{3,4}, Hongxia Hao⁵, Yi Liu^{6,7}, Turab Lookman⁸, Rampi Ramprasad⁹, Matthias Scheffler¹⁰, Hong Wang¹¹, Tong-Yi Zhang⁶

¹Supercomputer Facility, Australian National University, Canberra ACT 2601, Australia.

²University of NSW, Canberra ACT 2600, Australia.

³Department of Chemistry, Virginia Tech, Blacksburg, VA 24061, USA.

⁴Molecular Sciences Software Institute, Blacksburg, VA 24060, USA.

⁵Microsoft Research AI for Science, Shanghai, China.

⁶Materials Genome Institute, Shanghai University, Shanghai 200444, China.

⁷Department of Physics, Shanghai University, Shanghai 200444, China.

⁸AiMaterials Research LLC, USA.

⁹School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA.

¹⁰The NOMAD Laboratory at the Fritz Haber Institute of the Max Planck Society, Berlin 14195, Germany.

¹¹Materials Genome Initiative Center & School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

*Correspondence to: Dr. Rika Kobayashi, Supercomputer Facility, Australian National University, Canberra ACT 2601, Australia.
E-mail: Rika.Kobayashi@anu.edu.au

How to cite this article: Kobayashi, R.; Amos, R. D.; Crawford, T. D.; Hao, H.; Liu, Y.; Lookman, T.; Ramprasad, R.; Scheffler, M.; Wang, H.; Zhang, T. Y. Unlocking the future of materials science: key insights from the DCTMD workshop. *J. Mater. Inf.* **2025**, *5*, 50. <https://dx.doi.org/10.20517/jmi.2025.44>

Received: 1 Jun 2025 **First Decision:** 7 Jul 2025 **Revised:** 14 Jul 2025 **Accepted:** 22 Jul 2025 **Published:** 4 Nov 2025

Academic Editor: William Yi Wang **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

The International Workshop on Data-Driven Computational and Theoretical Materials Design was held between October 9-13, 2024, in Shanghai, gathering leading scientists and researchers from around the world, representing various aspects of data-driven AI methodologies and applications in materials design. The topics covered over 46 talks and 29 posters spanned a wide range of the latest advancements, including Machine Learning for Materials Design, Method Development, Machine Learning Interatomic Potentials, Advanced Computing, Infrastructure and Standards, Large Language Models, and Autonomous Labs. As part of the workshop, a panel discussion titled "Unlocking the AI Future of Materials Science" was held to disseminate the state-of-the-art of AI/ML in materials science and consider directions for the future. This report is a synthesis, for this Special Issue, of the panel discussion - drawing on insights gained from the workshop as a whole and surrounding conversations, in particular, the question of what constitutes success.



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



Keywords: Machine learning, state-of-the-art, materials design, autonomous labs, data management

INTRODUCTION

The International Workshop on Data-Driven Computational and Theoretical Materials Design (DCTMD) was held from October 9 to 13, 2024, in Shanghai, amidst the buzz surrounding the announcements of the Nobel Prizes in Physics and Chemistry. The workshop aimed to gather leading scientists and researchers from around the world, representing various aspects of data-driven artificial intelligence (AI) methodologies and applications in materials design, to facilitate the exchange of the latest research and stimulate discussion. There were 191 participants from 11 different countries, representing academia and industry, at various career levels, providing a diverse range of perspectives. The focus areas were chosen to highlight innovative approaches and technologies in materials research today, including:

- Data management and stewardship for materials
- AI for materials design
- AI/autonomous/self-driving/automatic materials lab
- High-throughput computational and experimental materials design
- Advanced computing for materials design

As part of the workshop, a panel discussion titled “Unlocking the AI Future of Materials Science” was held - available in full on Koushare^[1] - to disseminate the state-of-the-art of AI/ML in materials science and explore directions for the future. The following seed questions were posed to the panel to stimulate discussion:

- What has been the greatest success of AI/ML in the sciences?
- With growing skepticism about the validity of AI claims and the issue of hallucinations in large language models (LLMs), should we be putting standards in place to determine when AI claims can be taken seriously? Especially considering that robust scientific theories and models have clearly defined domains of validity (e.g., classical mechanics vs. quantum mechanics), and data come with error bars. What, then, can we say about machine learning models?
- AI/ML in materials science is undoubtedly data-driven, but compared to some other disciplines, we are not yet truly doing Big Data. What should the community be doing to ensure the integrity and accessibility of data?
- Looking to the future, what will be the next breakthrough area for AI in materials science - or, failing that, what problem would you most like to see AI solve?

These questions underpinned the many talks of the conference, especially in judging metrics for success and how to achieve them. The insights gained from the panel discussion, talks and accompanying conversations are given below.

DISCUSSION

The DCTMD workshop indeed managed to cover a wide range of topics and themes in materials research as outlined by the focus areas above, and these could be roughly categorized across the 7 plenary talks, 25 invited talks, 14 contributed talks, and 29 posters as follows:

- Machine learning for materials design
- Method development
- Machine learning interatomic potentials (MLIP)

- Advanced computing
- Infrastructure and standards
- LLMs
- Autonomous labs

and represented in [Figure 1](#) as a pie chart. This is only a rough breakdown, as some talks covered multiple topics - though to varying extents - but it provides a good representation of where efforts are currently focused in AI/ML in materials science. Additionally, the categories above were chosen to better highlight certain points. By far, the majority of presentations dealt with the application of ML techniques - analyzing data to make predictions through supervised learning. There was also significant work in Method Development, focusing on building better models to more accurately represent materials and predict their properties. This was particularly evident in the plenary and invited talks, which were skewed by the selection of speakers recognized as pioneers in advancing the field. Furthermore, the MLIP category encompasses these two areas - method development and application to materials design - but has been separated to highlight the popularity of this approach. Indeed, this is the methodology underpinning the AlphaFold work, which won the Nobel Prize in Chemistry. There were a few talks on robotic labs, providing impressive evidence that robotic synthesis can be more systematically reproducible than when done by humans. Surprisingly, there was not a larger representation of talks on LLMs, especially given the internet's perception that ChatGPT and DeepSeek are taking over the world. While LLMs were frequently mentioned in many talks, few concrete results were presented, reflecting the hype and showing that the use of LLMs in materials science is still in its early days. Notably, however, is the category labeled Advanced Computing. These talks focused on conventional computational materials science - without ML - and even in many of the ML talks, conventional computational materials science played a significant role. In the context of this audience, the findings of the panel questions are summarized in detail as follows.

Successes of AI/ML in materials science

Shortly before the workshop, the announcement of the Nobel Prizes in physics and chemistry was made. The prize for physics went to John Hopfield and Geoffrey Hinton for “foundational discoveries and inventions that enable machine learning with artificial neural networks” (<https://www.nobelprize.org/prizes/physics/>). The chemistry prize was awarded to David Baker, Demis Hassabis and John Jumper for “computational protein design” and “protein structure prediction” (<https://www.nobelprize.org/prizes/chemistry/>). For materials science the closest equivalent to breakthrough science was generally speculated to be the rise of high throughput materials discovery and autonomous labs (to be discussed in the next section). However, further consideration raised the question, “What does success look like?” The consensus was that there has yet to be an AlphaFold-equivalent breakthrough moment in materials science, but numerous small successes have demonstrated that AI/ML can be a useful tool when used in conjunction with other methods. In particular, as attested by the prevalence of studies, machine learning force fields are popular - indeed, they were part of the AlphaFold breakthrough. AI-driven materials design success stories are beginning to emerge in many areas of materials science, such as the design of application-specific practical polymeric materials^[2] or the development of a tolerance factor to predict the stability of not yet synthesized perovskites^[3]. Feature engineering, integrating digital materials representations, has provided insight into determining the capability and accuracy of material property prediction.

Autonomous labs

One of the biggest talking points in recent years was spurred by a pair of Nature papers^[4,5] on the AI discovery of novel materials and the use of autonomous labs. In the first paper^[4], machine learning techniques claim to have “discovered” 2.2 million novel structures, heralded as a breakthrough in materials discovery. Coupled with a workflow where these novel compounds can go straight to synthesis through

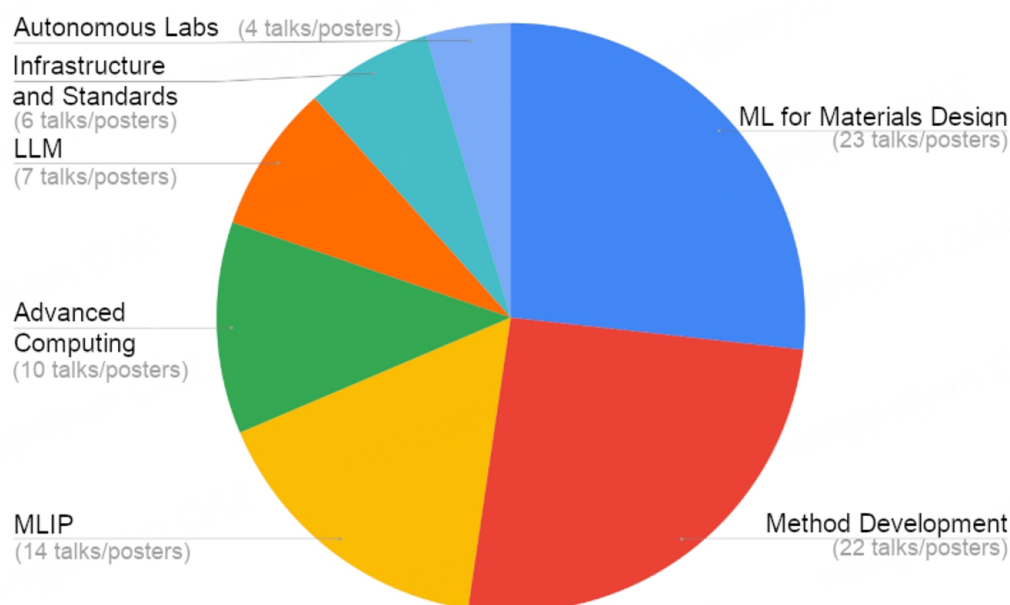


Figure 1. Breakdown of the topics and themes covered, to some extent, in the talks and posters.

ML-recommended processes via autonomous (robot-driven) laboratories^[5], this heralded another breakthrough in the high-throughput creation of new materials. However, these papers were swiftly followed by disputes, notably Refs.^[6,7], on the analyses questioning the claims of novelty - that some of the materials were already known, not all material classes had been included and a lot of the proposed new materials were not stable. There were questions about the value of novelty without functionality; although some of the structures were new, their utility was unclear, and whether proceeding to synthesize them without human quality control was cost-cutting or simply wasting money. Nevertheless, as Leeman *et al.* admitted, there are impressive aspects to autonomous labs, including AI's ability to develop working recipes for synthesis and the simplification of procedures by removing labor-intensive steps from humans^[7]. This view was certainly reflected in the workshop where rather than taking humans out of the loop, there was recognition of a need still for human intervention, and the focus was on integrating theory and experiment^[8-10]. As Jiang pointed out, robots can improve efficiency as they do not need to rest and it is generally accepted that they can carry out experiments reproducibly and accurately^[11]. However, data from experiment is still scarce and sparse and often incompletely characterized by metadata and so is mainly augmented by theoretical data to guide the next experiment in a feedback mechanism referred to as "inverse" or "adaptive design". In this workflow, the data, obtained from theory or increasingly from LLMs, which may contain errors, is replaced by confirmed experimental data, and fed back into, thus improving, the training model. Following this protocol, robotic labs have been shown to be successful in tackling a variety of problems, e.g., the automated synthesis of oxygen-producing catalysts^[12] or the design of chiroptical films^[13]. With recent advances in the use of efficient algorithms and reinforcement learning to mimic reasoning, such as in "Chain of Thought" in next-generation LLMs^[14] there is promise that data requirements and computations can be kept to a minimum, while achieving performance comparable to or exceeding the state of the art. Algorithms such as proximal policy optimization (PPO)^[15] will undoubtedly play an increasingly important role in controlling autonomous workflows in these labs.

Sharing data and setting standards

From the core themes of the workshop, it was evident that data plays a very important role. AI/ML is very much dependent on data and to work well needs good quality data probably through well-curated

databases. As Trunschke found from her experience in the field of catalysis, there is currently not enough data for standard ML techniques to be effectively used^[16]. Such data are hard to find, often only available through papers, sometimes in the form of figures, and negative results often go unreported. She showed examples of how the SISO method can work well with sparse data provided it is “clean”, i.e., well-characterized data^[17]. Consequently, her recent efforts have been concentrated on strategies for data acquisition, storage and use^[18]. The importance of data sharing was found to be paramount, though there was recognition of possible limitations due to proprietary concerns, especially in industry. However, even with freely shared data there were still questions of trustworthiness and making sense of it. This immediately raised the thorny issue of setting standards, especially for interoperability, and brought out a lot of differing opinions, essentially who, what and why. There is an inherent belief that for data to be shareable it needs to be in a certain format following certain rules, such as in the philosophy of FAIR data^[19]. However, who decides what rules to set? In reality, it is hard to get people to agree to which standard to adopt. A dominant publisher such as Materials Project^[20] or the Protein Data Bank^[21] has enough driving force for people to follow their lead, but, on the whole, most people want to do their own thing. The analogy was given of electrical plugs around the world. However, as with electrical plugs, why should people need to be forced to adhere to the one standard when it may be less troublesome to just work with converters.

Building community

As part of the discussion of setting standards, the primary question was “who sets the rules?” and it was agreed that it has to be done by community consensus rather than be imposed by some governing authority as happens all too often. It became clear that the materials science community needed to talk, for which the Workshop provided a good forum, but not how to start the conversation. There are the beginnings of communities being built in materials science through data and tools platforms, such as Materials Project^[20], Materials Cloud^[22] and DP Technology^[23]. Also, there are community efforts to establish ontology for sharing data, such as NFDI4Cat^[24]. However, these are not as established as the Molecular Sciences Software Institute (MolSSI)^[25] nexus for science, education and co-operation for the global computational molecular sciences community^[26]. The success of MolSSI is rooted in community engagement. Its origins were in identifying common needs and letting standards grow naturally. Notably, unlike the aforementioned materials science platforms, it encompasses a range of software packages and tools and works with the developers as being the drivers of what people will end up using.

CONCLUSION

The International Workshop on DCTMD demonstrated that work in the area of AI/ML in materials science is still going strong and producing new insights. Though there has yet to be an equivalent AlphaFold breakthrough moment there have been many small successes or achievements. AI/ML has improved greatly the success rate, saving time and reducing cost, by guiding iterative high-throughput experiments along the whole process of materials development, though humans are still needed in the loop. And there is still a lot of work for humans to do. There remains a strong feeling that AI/ML is not yet at a stage to be trusted in isolation and theory and modelling are still the way forward. Even so, AI/ML is proving a useful addition to the toolkit augmenting fundamental theory and experiment. AI can accelerate existing computational prediction, bridge the gaps across multiple length/time scales, and even map the relationships between structure-property relationships without an explicitly defined underlying theory. The rational design of physically meaningful ML features to do this is essential to materials science. More materials science-adapted ML algorithms need to be developed to tackle the challenge of scarce materials data. For AI/ML to work well, high-quality data that are well-curated and fully characterized by metadata - so that they are accessible, shareable, and reusable - are essential. The community still needs to come together to achieve this, and conferences such as these provide a good way to move forward.

DECLARATIONS

Acknowledgments

We thank Prof. Jincang Zhang and Dr. Runhai Ouyang for their efforts in the organization of the DCTMD workshop. Thanks also go to Dr Mohammad Khatamirad of Technische Universität Berlin for his contribution to the panel and related discussions.

Authors' contributions

Mainly responsible for writing the manuscript: Kobayashi, R.; Amos, R. D.

Reading and contributing to the ideas presented in this manuscript: Kobayashi, R.; Amos, R. D.; Crawford, T. D.; Hao, H.; Liu, Y.; Lookman, T.; Ramprasad, R.; Scheffler, M.; Wang, H.; Zhang, T. Y.

Availability of data and materials

The talks from the DCTMD conference for which permission has been granted can be found at <https://www.koushare.com/live/details/33143>.

Financial support and sponsorship

This work was supported by the mobility program of Sino-German Center, National Natural Science Foundation of China (No. M-0209). The authors are grateful to all our sponsors (<https://dctmd2024.scievent.com/sponsors.html>) for their generous support of the Workshop. TDC was supported by the U.S. National Science Foundation via grant CHE-2136142.

Conflicts of interest

This manuscript was processed under a double-blind peer review system. Kobayashi, R. and Wang, H. serve as Associate Editors of *Journal of Materials Informatics*, while Liu, Y.; Scheffler, M. and Kobayashi, R. are Guest Editors of the Special Issue “Unlocking the AI Future of Materials Science”: Selected Papers from the International Workshop on Data-driven Computational and Theoretical Materials Design (DCTMD). Zhang, T. Y. is the Editor-in-Chief of *Journal of Materials Informatics*. None of the above individuals were involved in any part of the editorial process, including reviewer selection, manuscript handling, or decision making. The other authors declare no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Panel Discussion: Unlocking the AI future of Materials Science. 2024. <https://www.koushare.com/live/details/33143?vid=150539>. (accessed 23 Jul 2025).
2. Tran, H.; Gurnani, R.; Kim, C.; et al. Design of functional and sustainable polymers assisted by artificial intelligence. *Nat. Rev. Mater.* **2024**, *9*, 866-86. DOI PubMed PMC
3. Bartel, C. J.; Sutton, C.; Goldsmith, B. R.; et al. New tolerance factor to predict the stability of perovskite oxides and halides. *Sci. Adv.* **2019**, *5*, eaav0693. DOI PubMed PMC
4. Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. Scaling deep learning for materials discovery. *Nature* **2023**, *624*, 80-5. DOI PubMed PMC
5. Szymanski, N. J.; Rendy, B.; Fei, Y.; et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **2023**, *624*, 86-91. DOI PubMed PMC

6. Cheetham, A. K.; Seshadri, R. Artificial intelligence driving materials discovery? Perspective on the article: scaling deep learning for materials discovery. *Chem. Mater.* **2024**, *36*, 3490-5. DOI PubMed PMC
7. Leeman, J.; Liu, Y.; Stiles, J.; et al. Challenges in high-throughput inorganic materials prediction and autonomous synthesis. *PRX. Energy*. **2024**, *3*, 011002. DOI
8. Zhang, B.; Zhu, Z.; Li, H.; Cao, J.; Jiang, J. Revolutionizing chemistry and material innovation: an iterative theoretical-experimental paradigm leveraged by robotic AI chemists. *CCS. Chem.* **2025**, *7*, 345-60. DOI
9. MacLeod, B. P.; Parlane, F. G. L.; Morrissey, T. D. Self-driving laboratory for accelerated discovery of thin-film materials. *Sci. Adv.* **2020**, *6*, eaaz8867. DOI PubMed PMC
10. Xue, D.; Balachandran, P. V.; Hogden, J.; Theiler, J.; Xue, D.; Lookman, T. Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* **2016**, *7*, 11241. DOI PubMed PMC
11. Jiang, J. A data-driven robotic AI-chemist. 2024. <https://www.koushare.com/live/details/33143?vid=150642>. (accessed 23 Jul 2025).
12. Zhu, Q.; Huang, Y.; Zhou, D.; et al. Automated synthesis of oxygen-producing catalysts from Martian meteorites by a robotic AI chemist. *Nat. Synth.* **2024**, *3*, 319-28. DOI
13. Xie, Y.; Feng, S.; Deng, L.; et al. Inverse design of chiral functional films by a robotic AI-guided system. *Nat. Commun.* **2023**, *14*, 6177. DOI PubMed PMC
14. Wei, J.; Wang, X.; Schuurmans, D.; et al. Chain-of-thought prompting elicits reasoning in large language models. *arXiv* **2022**, arXiv:2201.11903. <https://doi.org/10.48550/arXiv.2201.11903>. (accessed 23 Jul 2025)
15. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347. <https://doi.org/10.48550/arXiv.1707.06347>. (accessed 23 Jul 2025)
16. Trunschke, A. Creating synergies between experimental and computational approaches in advanced materials design. 2024. <https://www.koushare.com/live/details/33143?vid=150646>. (accessed 23 Jul 2025).
17. Foppa, L.; Ghiringhelli, L. M.; Girgsdies, F.; et al. Materials genes of heterogeneous catalysis from clean experiments and artificial intelligence. *MRS. Bull.* **2021**, *46*, 1016-26. DOI PubMed PMC
18. Marshall, C. P.; Schumann, J.; Trunschke, A. Achieving digital catalysis: strategies for data acquisition, storage and use. *Angew. Chem. Int. Ed. Engl.* **2023**, *62*, e202302971. DOI PubMed
19. Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data.* **2016**, *3*, 160018. DOI PubMed PMC
20. The Materials Project. <https://next-gen.materialsproject.org/>. (accessed 23 Jul 2025).
21. RCSB Protein Data Bank. <https://www.rcsb.org/>. (accessed 23 Jul 2025).
22. Materials Cloud. <https://www.materialscloud.org/home>. (accessed 23 Jul 2025).
23. DP Technology. <https://www.dp.tech/en>. (accessed 23 Jul 2025).
24. NFDI4Cat. <https://github.com/nfdi4cat/voc4cat/>. (accessed 23 Jul 2025).
25. MolSSI- The Molecular Sciences Software Institute. <https://molssi.org/>. (accessed 23 Jul 2025).
26. Crawford, D. The Molecular Sciences Software Institute. 2024. <https://www.koushare.com/live/details/33143?vid=150526>. (accessed 23 Jul 2025).