THE PROMISE AND PITFALLS OF AI FOR DESIGNING EXPERIMENTS IN A REAL LAB

FELIX STRIETH-KALTHOFF*

School of Mathematics and Natural Sciences, University of Wuppertal, Germany

The fact that Artificial Intelligence (AI) has approached human-level intelligence in several areas has fueled the vision of AI-driven experimentation for chemistry. Despite this promise, integrating AI into day-to-day laboratory research remains at an early stage. Below, I outline my view on the current state of AI-guided experimentation, my recent contributions, and potential future directions.

My view of the present state of research on AI, Robotics, Automation

Over the past decade, machine learning (ML) has entered the chemistry toolbox, though its impact varies by domain. Where large-scale, high-quality data exists – most commonly from simulation – ML shows strong predictive power. For example, learned force fields, trained *via* geometric deep learning on large-scale computational data, are now routine in molecular dynamics.[1] However, the picture changes with experimental data. Large-scale datasets are rare, and existing data are often scarce, noisy, and biased. Powerful ML models for small molecules, macromolecules, or solid-state materials do exist, but they are usually much narrower in scope, and cannot be transferred to new problems without extensive validation. Broad breakthroughs, comparable to the impact of *AlphaFold* in protein structure prediction,[2] remain rare.

These constraints are reflected in AI-driven experimentation, which has remained largely case-specific. Several studies across process optimization, drug discovery and materials chemistry have successfully shown AI-driven discoveries.[3] At the same time, they have largely relied on static, human-designed workflows: a fixed experimental pipeline, heuristic data analysis, and a pre-defined search space. Early uses of large language model (LLM) agents to design and operate these workflows indicate potential, but their limits and failure modes remain unclear, and the field is evolving at a rapid pace.

My recent contributions to AI, Robotics, Automation

My research aims to integrate AI into experimental workflows in molecular discovery and synthesis, while accounting for chemical principles, heuristics, and practical constraints (Fig. 1). At present, inferring these factors from data alone is challenging, and

we recently highlighted the gap between "data-only" models and expert-augmented models for obtaining practically useful predictions.[4]

As an example, we recently showed that a model for selecting candidate molecules with attractive emission properties became predictive only after pre-training on simulated data *and* fine-tuning on experimental measurements. This hybrid model then guided automated synthesis and spectroscopy, and yielded state-of-the-art organic laser emitters.[5] Inspired by these findings, we later found that pre-trained embeddings from domain-specific LLMs are strong baseline representations for molecular property prediction.[6] These findings point towards the importance of pre-trained molecular foundation models for property prediction and experiment planning.

Moreover, our research focuses on tools that steer ML-driven discovery toward practically relevant outcomes. Therefore, we recently introduced *BoTier* to flexibly encode a hierarchy of multiple optimization objectives, such as reaction yield and cost in chemical reaction optimization.[7] Moreover, we proposed a framework for generality-oriented optimization that seeks optima which are transferrable across different substrates, and that includes benchmarks to inspire algorithm development.[8]

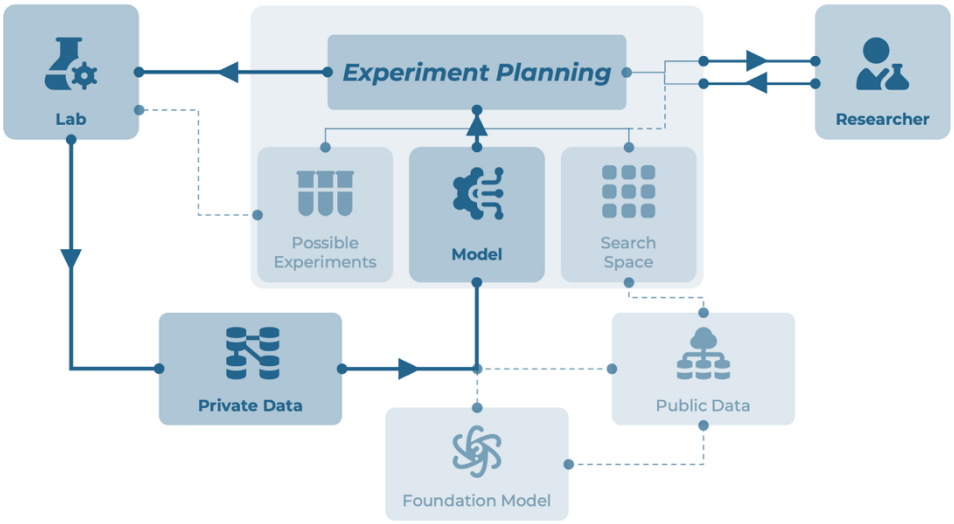

Fig. 1. Workflows and challenges for AI-driven experimentation. Optimal planning involves continuous refinements based on obtained data, literature knowledge, and researcher input. *[Icons: flaticon.com]*

Outlook to future developments on AI, Robotics, Automation

With the growing number of successful demonstrations of AI-driven experimentation, I foresee that the field is entering a transition from *proof-of-concept* studies to practical

tools that provide measurable value in the lab. I see four major directions for AI-driven experiment planning:

1. ***Foundation Models***. Historically, ML has followed "laws of scale",[9] with models trained on large datasets outperforming knowledge-derived systems. Recent results in chemistry indicate similar trends.[10] This suggests that foundation models trained on large computational corpora, and then fine-tuned to experiments (Fig. 1), will be central to property prediction and experiment planning. How reliably can such models capture the trends required for their effective adaption to downstream tasks?
2. ***Benchmarks***: To assess model utility, we need better benchmarks. Rather than relying on a small number of overused and biased datasets, the field needs tasks that reflect practical reality. Can such benchmarks – similar to those in computer science – steer the field towards models that matter in practice?
3. ***Large Language Models***: As LLM capabilities grow, they are likely to play a key role in AI-driven experiment planning:[11,12] searching the literature through retrieval-augmented generation (RAG), defining search spaces, or automating repetitive workflows in experiment and computation. To what extent can they match and outperform human decision-making capacities, and where should they be complemented by human researchers?
4. ***Hybrid, Human-in-the-Loop Systems***: Eventually, data, robotics, and LLM agents will have limits that human adaptability can offset. What is the right balance between model flexibility and inductive biases, between robotic and human experimentation, between agent independence and human oversight?

Overall, I believe that integrated AI–human co-design – rather than fully autonomous workflows – is the most direct route to impact in real-world laboratories.

References

1. I. Batatia, B. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, *et al.*, *arXiv* (2023).
2. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, *Nature* **596**, 583–589 (2021).
3. G. Tom, S. P. Schmid, S. G. Baird, Y. Cao, K. Darvish *et al.*, *Chem. Rev.* **124**, 9633 (2024).
4. F. Strieth-Kalthoff, S. Szymkuc, K. Molga, A. Aspuru-Guzik, F. Glorius *et al.*, *J. Am. Chem. Soc.* **146**, 11005 (2024).
5. F. Strieth-Kalthoff, H. Hao, V. Rathore, J. Derasp, T. Gaudin *et al.*, *Science* **384**, adk9227 (2024).
6. A. Kristiadi, F. Strieth-Kalthoff, M. Skreta, P. Poupart, A. Aspuru-Guzik *et al.*, *Int. Conf. Mach. Learn.* **41** (2024).
7. M. Haddadnia, L. Grashoff, F. Strieth-Kalthoff, *Digit. Discov.* **4**, 1417 (2025).

8. S. P. Schmid, E. M. Rajaonson, C. T. Ser, M. Haddadnia, S. X. Leong *et al.*, *arXiv* (2025).
9. Y. Bahri, E. Dyer, J. Kaplan, J. Lee, U. Sharma, *Proc. Natl. Acad. Sci. USA* **121**, e2311878121 (2024).
10. A. H. Cheng, C. T. Ser, M. Skreta, A. Guzmán-Cordero, L. Thiede, *et al.*, *Faraday Discuss.* **256**, 10–60 (2025).
11. A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White, *et al., Nat. Mach. Intell.* **6**, 525–535 (2023).
12. D. A. Boiko, R. MacKnight, B. Kline, G. Gomes, *Nature* **624**, 570–578 (2023).