https://doi.org/10.1038/s43588-025-00769-x

Balancing autonomy and expertise in autonomous synthesis laboratories

Xiaozhao Liu, Bin Ouyang & Yan Zeng

Autonomous synthesis laboratories promise to streamline the plan–make–measure–analyze iteration loop. Here, we comment on the barriers in the field, the promise of a human on-the-loop approach, and strategies for optimizing accessibility, accuracy, and efficiency of autonomous laboratories.

Motivations of autonomous laboratories

An autonomous laboratory is a system that leverages data, computation, expert knowledge, and artificial intelligence (AI) to predict and interpret experimental outcomes generated from an automated robotic platform. Once a research objective is set, an autonomous laboratory is expected to plan, execute, and iterate with minimal human intervention[1]. Autonomous laboratories hold substantial potential to accelerate research and have recently drawn particular attention in chemical and materials synthesis. These systems aim at making high-performance molecules or materials that meet the demands of advanced technologies, areas where computational design has set a substantially faster rate of progress[2]. The advantages of such an approach include increased throughput, improved reproducibility and replicability, standardized data management, self-learning capability, and a faster turnaround due to streamlining the iterative plan–make–measure–analyze loop.

Barriers faced by autonomous laboratories

To date, notable barriers still hinder the broader development and implementation of autonomous laboratories. Common challenges include sourcing off-the-shelf instruments, building customized instruments and tools from scratch, integrating multiple types of analytical instruments with various functionalities, handling special conditions such as high temperature and controlled atmosphere, interpreting complicated analytical results, and predicting synthesis outcomes when there is a lack of mechanistic understanding[3]. The last point is especially true for inorganic predictive synthesis compared to organic retrosynthetic analysis. The complexity and large capital requirement for building and maintaining an autonomous laboratory depends on the research scope, as a wide variety of materials, processes, and methods may be involved. These challenges might be partially solved by adopting semi-automation with certain manual intervention[4] or by implementing flexible and modular automation designs[5].

In order to effectively manage resources and data, monitor status, and control their operations, a sophisticated and robust workflow management software combined with programmable logic controllers and microcontrollers are required. This software orchestrates samples, equipment, robots, experiments, and data throughout the stages of planning, execution, and updating in an autonomous workflow. Notable examples include AlabOS[6] implemented in the A-Lab[2] and ChemOS 2.0[7]. However, the current workflows still face challenges related to increased throughput and greater complexity when integrating a large variety of devices in autonomous laboratories. Conflicts may arise when handling multiple tasks and devices simultaneously. This presents a substantial challenge, as it is also desired for these systems to go beyond basic functions and integrate advanced capabilities such as error detection, error reporting, error handling, recalibration, and ideally, the ability to reconfigure workflows and even equipment to adapt to the dynamic nature of research goals and environments. Error handling and workflow reconfigurations heavily rely on domain knowledge and the researcher's experience, which hinders the implementation of these advanced functions in an autonomous laboratory in the early stage of development. The efficiency of the central workflow management software also relies on the computation and AI's capabilities for rapid data processing, analysis, and decision making.

The accuracy and reliability of AI-driven data interpretation and the subsequent decision-making have raised concerns due to the trade-off between efficiency and quality and the impact of biases. These biases may originate from imperfect or inappropriate data sources, including biased sampling, measurements, labeling, and annotations. Additionally, human experts might perpetuate these biases by using the biased outputs from the AI models to further refine them. This cycle of biased-data-in and biased-data-out can ultimately worsen the accuracies and consistencies in autonomous laboratories[8]. Therefore, a well-designed strategy for human intervention in the process is critical to mitigating these biases, thereby enhancing accuracy, trustworthiness, and overall efficiency.

A human on-the-loop strategy

Although a 'human-out-of-the-loop' (fully autonomous) approach is being pursued in many autonomous research efforts, it may be impractical for most scientific fields and research groups due to the previously noted barriers and limitations. Human involvement in autonomous laboratories is not necessarily a disadvantage and should not be thought of as such. When strategically integrated, this involvement can enhance both the implementation and operation of these AI and robotics-driven systems. Properly designed human oversight can help to overcome many of the current barriers faced by fully autonomous systems. While some researchers have considered a 'human-in-the-loop' workflow that requires continuous human intervention, we propose a 'human-on-the-loop' approach. This approach offers a more effective and balanced way to lower the cost while boosting the efficiency of the experiment, allowing humans to oversee and intermittently interact with the automated processes when requested by the system (Fig. 1).

The human-on-the-loop approach takes advantage of the human's flexibility, adaptability, and expertise to fulfill tasks beyond the current capabilities of robots and AI. In addition to human expert quality control based on expertise, human experts excel in tasks where robots

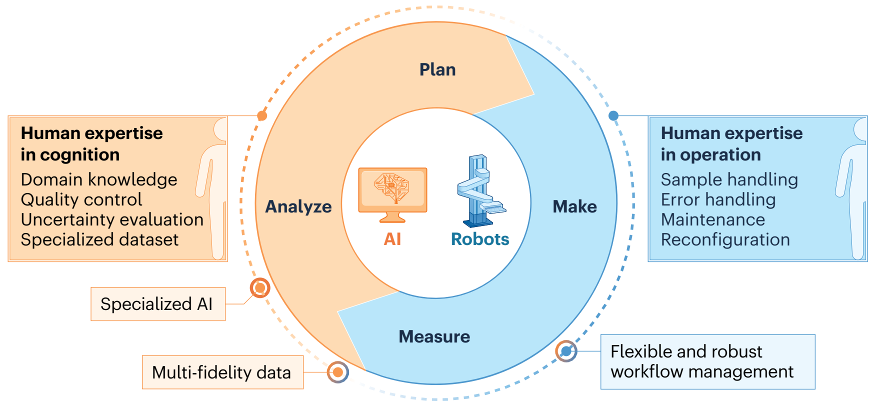

Fig. 1 | A schematic demonstration of a 'human-on-the-loop' strategy in autonomous synthesis laboratories where flexible robots, specialized AI, and human experts interact synergistically. This strategy emphasizes strategic human supervision with minimal intervention upon request, thereby enhancing the AI accuracy while maximizing overall research efficiency. Orange colored region represents the cognition part (planning and analysis), and blue represents the operation part (synthesis and measurement).

and AI may struggle, such as identifying and resolving errors on-the-fly, repairing or reconfiguring equipment, maintaining and cleaning devices, and handling apparatus/samples that require transporting or transferring that is easily performed by humans but more challenging for robots. Human experts may also be able to make real-time decisions based on intuition and expertise and evaluate uncertainties and outliers. All these aspects are invaluable in managing and executing complex tasks in material synthesis processes.

However, incorporating human oversight and assistance into autonomous laboratories without disrupting the workflow still faces some potential issues: how to avoid human precognitive bias, how to effectively enable the interoperability between human operators and robotic systems, how to standardize collaborative environments, and how to develop and implement advanced workflow management software that allows for seamless and timely augmentation of human contributions. By addressing these challenges, we can establish a more robust framework for integrating data and algorithms that further enhance AI accuracy in autonomous laboratories.

Enhancing the accuracy of AI using improved data and algorithms

Real-time data interpretation, insight extraction, and decision-making using AI-based models are desired capabilities for autonomous laboratories. However, data scarcity still limits the power of AI models in autonomous synthesis laboratories. Proxy or low-resolution data may be utilized to streamline the iterative plan–make–measure–analyze loop. This approach allows for quick but preliminary insights, which are useful to move the exploration forward in early stages.

To achieve a balance between accuracy and efficiency, one promising direction is to collect and use multi-source, multi-fidelity data to train machine learning (ML) models. In the context of materials synthesis, this data can include a range of sources: computationally obtained data such as energies, compositions, structures, and properties, as well as experimentally obtained data. Incorporating experimental data, including both successful and null results, helps to capture the complexities and nuances of real-world conditions that purely computational data often overlooks. Relevant experimental synthesis details can be extracted from published literature using natural language processing techniques[9]. Additionally, data generated from autonomous laboratories would be particularly useful in further enhancing AI accuracy by providing specialized, standardized, and diverse results including both 'successful' and 'null' outcomes[10].

For materials characterization, pre-trained ML models with simulated characterization data can efficiently generate usable ML models that can be further finetuned with experimental data[11]. Moreover, incorporating multiple characterization data, both structural and compositional, can effectively improve the efficiency and accuracy of the ML models[12]. To fully exploit multi-fidelity data, data standardization and automated data cleaning workflows are crucial, as they contribute to the development of more robust and meaningful ML models. An autonomous synthesis laboratory may leverage data across a spectrum of fidelity to balance quality, quantity and efficiency. This ranges from generic computational datasets, low-cost and fast proxy measurement datasets of certain materials characteristics, generic experimental datasets, and all the way to specialized, standardized, and carefully evaluated datasets.

Another critical area of focus is the customization of the training processes to better utilize multi-fidelity data. One emerging protocol is to develop foundation models trained on extensive, generic datasets[13]. Finetuning these foundation models using experimental data has the potential to reveal the transferability of global models to domain-specific science. Additionally, a rising trend is to create specialized experimental or computational datasets that encapsulate domain-specific knowledge more accurately[12,14]. Models developed using these datasets are better suited for adaptation or fine-tuning with data generated for specific research domains.

The integration of human expertise remains indispensable for the efficient training of interpretable ML models. For instance, human-supervised approaches can enhance the sampling process and adapt models into specific research domains. Symbolic ML, an emerging technique, shows promise in developing phenomenological models that more accurately describe data[14,15]. The adaptation of such methods into featurization and sampling processes could prove beneficial. When combined with the state-of-the-art active learning algorithms, the infusion of interpretable domain knowledge through phenomenological theories is likely to greatly enhance both the efficiency of training processes and the interpretability of final models. However, we need to be cautious when incorporating domain expertise into the featurization and sampling processes, as not all domain knowledge has been thoroughly validated, and this can lead to the introduction of biases in AI predictions.

Future perspectives

The augmentation of AI and laboratory automation in the form of autonomous laboratories sets a new data-driven high-throughput paradigm to advance scientific discovery. Automated experimentation is the most controlled, precise, and efficient way to validate the rapid predictions from AI and computations. In the meantime, the large volume of data generated by automated experiments requires the use of AI to quickly, and seamlessly, analyze results and make decisions.

The presence of biases and unsatisfactory accuracy in autonomous materials synthesis laboratories urges the need for improvements. Enhancement could be achieved through several approaches: 1) improving data quality by obtaining higher quality, better annotations by experts, focusing on more specific information, and incorporating diverse sources combined with physics-informed data augmentation; 2) reducing human bias and improving uncertainty quantification by strategically involving human intervention; 3) minimizing AI bias by appropriately incorporating human supervision that provides expert evaluation and data quality control when AI predictions lack confidence; and 4) standardizing data creation and making it machine-readable and unbiased through automated workflows. More specifically, human researchers need to pre-screen the data for various ML training purposes based on established phenomenological theories. Additionally, there should be a fine-tuning interface that allows smooth human intervention, enabling real-time monitoring of the training process and timely action when bias occur.

To make autonomous laboratories and workflows more accessible and practical, several key improvements are necessary. Open-sourcing data, software, and hardware is important to lower the barriers and costs associated with their adoption and implementation by a broader community. Improving the application programming interfaces and graphical user interfaces are essential for supporting effective 'human-on-the-loop' human–machine collaboration. Further development in synthesis-by-design methodologies is needed so that the efficiency and predictability of experiments can be enhanced. Standardizing programmable instruments can promote compatibility across different research domains, enhancing the ease of use and facilitating interdisciplinary research. Implementing in situ and operando experiments that can reveal dynamic changes during materials synthesis will enable a deeper mechanistic understanding of the process involved. These enhancements will not only improve the functionality, affordability, and reliability of autonomous laboratories but also empower more researchers to leverage the benefits of autonomous research and collaboratively push this technology forward.

Xiaozhao Liu, Bin Ouyang & Yan Zeng
Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL, USA.
e-mail: bouyang@fsu.edu; zeng@chem.fsu.edu

Published online: 05 February 2025

References

1. Wang, H. et al. *Nature* **620**, 47–60 (2023).
2. Szymanski, N. J. et al. *Nature* **624**, 86–91 (2023).
3. Szymanski, N. J. et al. *Mater. Horiz.* **8**, 2169–2198 (2021).
4. McCalla, E. *ACS Eng. Au* **3**, 391–402 (2023).
5. MacLeod, B. P., Parlane, F. G. L., Brown, A. K., Hein, J. E. & Berlinguette, C. P. *Nat. Mater.* **21**, 722–726 (2022).
6. Fei, Y. X. et al. *Digit. Discov.* **3**, 2275–2288 (2024).
7. Sim, M. et al. *Matter* **7**, 2959–2977 (2024).
8. Sun, W., Nasraoui, O. & Shafto, P. *PLoS One* **15**, e0235502 (2020).
9. Kononova, O. et al. *Sci. Data* **6**, 203 (2019).
10. Raccuglia, P. et al. *Nature* **533**, 73–76 (2016).
11. Szymanski, N. J. et al. *npj Comput. Mater.* **9**, 31 (2023).
12. Wang, L., He, T. & Ouyang, B. Preprint at *ChemRxiv* https://doi.org/10.26434/chemrxiv-2024-fmq8p (2024).
13. Deng, B. W. et al. *Nat. Mach. Intell.* **5**, 1031–1041 (2023).
14. Wang, L. & Ouyang, B. *Adv. Mater.* **36**, e2307860 (2024).
15. Davies, A. et al. *Nature* **600**, 70–74 (2021).

Acknowledgements

This work is supported by the startup fundings for Y.Z. and B.O., respectively, from Florida State University.

Competing interests

The authors declare no competing interests.