Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PERSPECTIVE

View Article Online
View Journal | View Issue

Cite this: *Digital Discovery*, 2024, **3**, 621

The future of self-driving laboratories: from human in the loop interactive AI to gamification

Holland Hysmith,[a] Elham Foadian,[b] Shakti P. Padhy,[b] Sergei V. Kalinin,[bc] Rob G. Moore,[d] Olga S. Ovchinnikova[be] and Mahshid Ahmadi *[b]

Recent developments in artificial intelligence (AI) and machine learning (ML), implemented through self-driving laboratories (SDLs), are rapidly creating unprecedented opportunities for the accelerated discovery and optimization of materials. This paper provides a joint analysis of SDLs from both academic and industry perspectives, highlighting the importance of integrating human intelligence in these systems. It discusses the necessity of careful planning in SDL design across physical, data, and workflow dimensions, including instrumental setup, experimental workflow, data management, and human–SDL interaction. The significance of integrating human input within SDLs, especially as the focus shifts from individual tools and tasks to the creation and management of complex workflows, is emphasized. The paper stresses on the crucial role of reward function design in developing forward-looking workflows and examines the interplay between hardware evolution, ML application across chemical processes, and the influence of reward systems in research. Ultimately, the article advocates for a future where SDLs blend human intuition in hypothesis formulation with AI's precision, speed, and data-handling capabilities.

Received 2nd February 2024
Accepted 14th March 2024

DOI: 10.1039/d4dd00040d

rsc.li/digitaldiscovery

I. Materials matter

Materials discovery has been driving technological evolution since the dawn of time, long predating the formal establishment of materials science in the 1960s.[1] Understanding material properties lies at the cross-section of scientific fields including biology, chemistry, physics, and engineering.[2] Materials discovery and optimization comprises the synergy of synthesis and fabrication with property measurements, whether mechanical, chemical, or electrical.[1,2] Traditionally, all stages of this process were ideated and implemented by human scientists, with automated approaches used only for well-defined and simple operations.

The introduction of machine learning (ML) sparked a wave of curiosity among scientists with a new perspective on the scientific method – both in the theory and computation domains, and in real-world applications. The race to create the next best technological breakthrough became not a matter of human tenacity only, but also of utilization of artificial intelligence (AI).[3,4] For many decades, computational approaches including molecular dynamics simulations or density functional theory (DFT) unlocked molecular structures, elastic constants, electron densities, and vibrational properties.[5] Combining theoretical computations with ML speeds up discovery due to the use of known material properties and computed possibilities to construct virtually indefinite lists of new materials.[6,7] Although this feat has been driving much of the excitement over the past two decades, it has a considerable limitation. Namely, the key for practical applications is synthesizing materials in a traditional laboratory setting, and scaling from the lab to the prototypes and industrial settings. Until very recently, it was the role of the human scientist to bridge the worlds of theoretical and experimental science. The process of creating hypotheses, training computational models, setting realistic parameters, hypothesizing new discoveries, and conceptualizing and operationalizing future innovations can be enhanced by machine learning (ML) and used in laboratory synthesis, but it still requires human involvement.

As demonstrated in Fig. 1, the materials discovery process typically starts with a conceptual idea that is refined into a hypothesis. This hypothesis explores potential experimental routes to achieve specific material properties, functionalities, or physical mechanisms. The experiment process is a harmonious combination of two critical steps – the synthesis of the material and its characterization. This synergy not only offers feedback to theoretical models but also paves the way for unexpected discoveries. A broad gamut of methods, ranging from solution and solid-state synthesis to physical and chemical film deposition, have been developed to design complex materials in

[a]*Bredesen Center for Interdisciplinary Research, University of Tennessee, Knoxville, TN 37996, USA*

[b]*Institute for Advanced Materials and Manufacturing, Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN, 37996, USA. E-mail: mahmadi3@utk.edu*

[c]*Pacific Northwest National Laboratory, Richland, WA 99354, USA*

[d]*Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA*

[e]*Thermo Fisher Scientific, Carlsbad, CA 92008, USA*

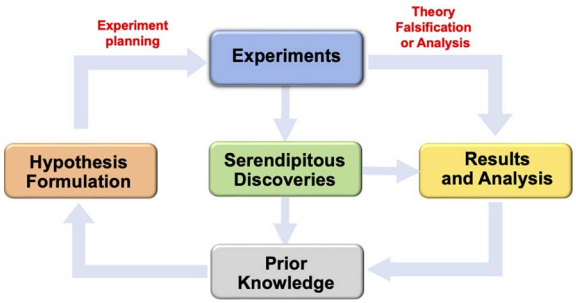

Fig. 1 A schematic that represents the scientific reward system through utilizing prior knowledge, hypothesis formulation, experiments, results and analysis, with serendipitous discoveries. Each component uncovers multiple pathways towards scientific discovery and can be used infinite times within the cycle.

laboratories.[8–10] Apart from these, tools such as nitrogen or argon gloveboxes, wet-lab benches, mass balances, pipettes, *etc.* are essential components in this process. Once the material recipes are refined through trial and error, various characterization instruments are used to measure and quantify the material properties.

Material characterization is a pivotal component of this process, involving a myriad of methods to either confirm or disprove the initial hypothesis. Some examples of characterization techniques are X-ray diffraction (XRD), atomic force microscopy (AFM), photoluminescence (PL), and X-ray photoelectron spectroscopy (XPS) which are used to assess structural properties, surface properties, electronic structure and properties, and surface chemistry of materials.[11–13] Furthermore, there are techniques to assess mechanical, thermal, magnetic, and optical properties and the study of all these various properties is essential in understanding how materials will perform in real-world applications.

This comprehensive characterization process is augmented by a sophisticated software tool to enhance the analysis of data generated. While traditional software such as Origin by OriginLab Corporation and Microsoft Excel have been fundamental for basic data organization and analysis since the late 1980s, the integration of advanced cross-platform software like MATLAB and Python has transformed the landscape of materials data analysis.[14,15] These modern tools are instrumental in materials science research, particularly through their integration of machine learning and advanced optimization algorithms. They enable efficient data analysis, handling large datasets, and offer advanced tools for pattern recognition, property prediction, and optimization of compositions and process conditions. Additionally, both platforms provide powerful visualization features for data interpretation and offer flexibility for research customization. Moreover, their capacity for simulation and predictive modeling aids in predicting material behavior, enhancing research efficiency and innovation in materials science.

Overall, all of this has notably influenced the synthesis, characterization, and analysis landscape of complex materials. With over two centuries of relentless synthetic research, the pursuit for groundbreaking chemical transformations and novel reactivity has grown progressively challenging and crucial. Recent advancements in various areas such as two-dimensional (2D) materials,[16] perovskites,[17] quantum dots,[18] transition metal catalysts,[19] metal–organic frameworks (MOFs),[20] and so on, coupled with rapid technological progress, have unveiled unprecedented possibilities. There is a growing demand for the discovery of new materials rapidly and efficiently with multifunctionality, capable of seamlessly integrating various functionalities to meet specific application requirements. To address this challenge, high-throughput experiments (HTE), like what is featured in SDLs, have been implemented in the past two decades which has helped significantly in increasing the productivity and speed of research and development (R&D) of new materials.[21–32]

The recent advancements in AI[33–35] and the refined efficiency of ML algorithms[36] have significantly transformed the landscape of materials science research, especially with the emergence of SDLs. These SDLs utilize machine learning to autonomously conduct complex experiments, enhancing precision and consistency while reducing human error.[37–40] The integration of AI and ML algorithms enables these labs to rapidly process and interpret large datasets, uncovering patterns and insights that might elude human analysis. This synergy of automated experimentation and analysis facilitates a dynamic, iterative research process. AI-driven systems can adjust experimental parameters in real time, allowing for a more targeted exploration of new materials. This approach accelerates the discovery process, particularly in high-throughput experimentation, and is invaluable in fields like energy storage[41,42] and nanotechnology,[43,44] where finding the right material is crucial.

However, despite the exponential growth over the last several years, limitations such as high initial investment, difficulty in automating human tasks, and rigid manufacturing of instruments pose significant obstacles towards using SDLs.[45,46] A pivotal challenge in constructing SDLs lies in creating a platform that can independently conduct all experiments, spanning from synthesis to characterization. Automated synthesis platforms (ASPs) play a vital role in enabling the capabilities of SDLs.[47–49] Larger institutions have constructed SDLs promising future success, but many are skeptical about whether the overall benefits outweigh the drawbacks for their research applications.[50] Overall, the motivation to integrate SDLs into materials research is the opportunities for high-throughput data collection, enhanced data reproducibility, reduction of time-intensive tasks, faster sample to answer and increased safety.[51]

Thus, it is essential to clearly define the role of scientists in the operation of SDLs. This will be crucial in guiding the continuous advancement and adaptability of SDLs, ensuring that they are optimally leveraged for future scientific applications. We consider what makes SDLs an asset for materials discovery and explore the opportunity to allow for both automation and human intervention to co-exist. As scientists incorporate more artificial intelligence into their laboratories, humans are the key in determining which revelations are worth contributing to the research community. Therefore, we propose

that the human-in-the-loop automated experiments will pave the future for materials discovery.

II. Workflow

In traditional human-dependent laboratories, particularly those focused on materials synthesis, the role of scientists and engineers is central and multifaceted, ranging from forming hypothesis, ideation of complex workflows, and their orchestration and execution. In emerging areas, the human scientist is initially wholly responsible even for the design and construction of the experimental apparatus. The design phase is critical, as it lays the foundation for successful experimentation. Moreover, this stage also serves as a significant opportunity for knowledge transfer, with experienced scientists guiding less experienced team members, thereby fostering a collaborative and educational environment.

During the experimental phase, shown in Fig. 2, the reliance on human expertise is even more pronounced. Scientists are tasked with conducting experiments, a process that involves more than just following protocols. It requires making informed decisions based on their scientific judgment and observations. This phase often involves a lot of trial and error, hypothesis testing, and problem-solving, all of which depends heavily on the scientist's experience and intuition. Additionally, peer interaction plays a crucial role in this phase. Scientists often collaborate, share insights and discuss results, which not only aids in interpreting data but also in shaping future research directions. The lack of advanced computational tools means that these scientists have to rely on their collective wisdom to analyze data and draw conclusions, making the human element indispensable in the realm of materials science research.

Despite their cost-effectiveness compared to robotic labs, traditional human-dependent laboratories have several limitations. The reliance on manual labor makes processes more time-consuming, often resulting in slower research progress. Human error, variability in experimental techniques, and fatigue can also affect the accuracy and repeatability of results. Additionally, these labs cannot simultaneously perform multiple tasks or experiments, limiting their throughput and efficiency. Finally, human-operated labs necessarily limit environmental conditions to human-friendly and impose stringent limitations on safety. The absence of advanced analytics and AI-driven insights means that complex data analysis and pattern recognition were more challenging and slower, potentially overlooking critical insights that automated systems might capture more readily. Consequently, while being cost-effective, these labs face significant constraints in terms of speed, efficiency, and capacity for handling complex and large-scale data analyses. The advances in the physical sciences, achieved through the synergy of theory and physical experiments, have paved the way for the integration of AI and ML methods to accelerate materials discovery.

Platforms for accelerating materials discovery aim to go beyond traditional human-led science but face several challenges. The first is to create automated hardware platforms capable of executing experiments.[52–54] The second is to build data flow and management systems that will optimize the sharing of data across platforms and AI/ML models.[55–58] The third is to design and operationalize the workflows running the hardware that will not only incorporate data with theory based on known physical laws, but also allow for learning from human intuition and experience in the decision-making process.[53,54,59] The key aspect of the workflow design is the need to define the target, whether it is materials optimization or physics discovery. The lesson of the past decade is that scaling experiments or computations individually by orders of magnitude or accelerating data acquisition is insufficient to expedite materials discovery and operationalization.

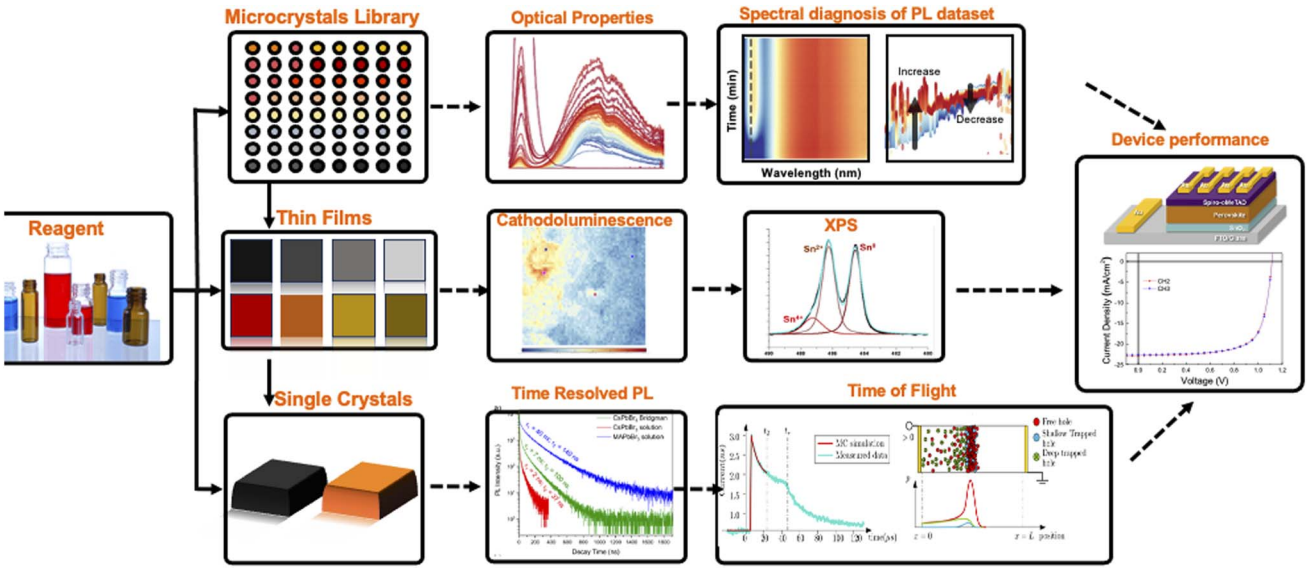

Fig. 2 An example diagram demonstrating the workflow of hybrid perovskite formulation from synthesis to characterization, leading to overall device performance evaluation.

III. Automated labs on the physical plane

HTE in materials science, also known as high-throughput materials discovery or combinatorial materials science, started gaining traction in the late 1990s and early 2000s. This approach was adopted due to the success of high-throughput screening in the pharmaceutical and biotechnology industries, where it was used to speed up drug discovery by allowing for the rapid screening of thousands of compounds to identify potential drug candidates.[60–63] Automation, infused with robotics, was introduced to enhance the screening process. Custom-designed systems, compatible with multi-well plates, were employed to facilitate the simultaneous handling and assessment of numerous samples, thereby significantly ramping up the throughput. By leveraging automation, high-throughput screening could effectively process, for instance, up to 10 000 fermentation broths per week, a dramatic increase in capacity compared to 800 samples per week at the maximum capacity of the preceding methodologies. Since then, automated high-throughput screening has signified a pivotal shift towards accelerated, efficient, and large-scale screening processes in drug discovery in the pharmaceutical industry, overcoming the inadequacies of the previous labour-intensive and time-consuming practices.[60,61,64–66]

Inspired by this, researchers in the realm of materials science have adapted the high-throughput screening concepts in experiments to accelerate the discovery and development of new materials with desirable properties. High-throughput synthesis (HTS) has enabled researchers to rapidly synthesize vast arrays of materials and efficiently explore a broad landscape of materials. At the core of HTS is the use of automated, parallel processing techniques that allow for the simultaneous synthesis of a multitude of different materials under various conditions. Some of the notable HTS methods are combinatorial physical vapour deposition (PVD),[25,67,68] combinatorial chemical vapor deposition (CVD),[25,68] robot-assisted materials synthesis,[23,42,68,69] inkjet printing,[70] additive manufacturing,[30,71–73] high-throughput spark plasma sintering,[31] microfluidic synthesis,[67,68,74] combinatorial flow synthesis,[31,75] laser scanning ablation,[76] combinatorial hydrothermal synthesis,[77] and rapid microwave-assisted synthesis.[78] Among these, few methods are in a true sense combinatorial methods of synthesis, *e.g.*, combinatorial PVD, combinatorial CVD, microfluidic synthesis, combinatorial flow synthesis, and laser engineered net shaping. In the rest of the methods, the compositions of the materials in the library are either directly or indirectly defined by human beings.

Automated characterization techniques are pivotal in complementing high-throughput synthesis by providing fast and efficient analysis of large volumes of samples. These techniques enable rapid assessment of the physical, chemical, and structural properties of materials, facilitating the identification of materials with targeted properties/performance. Some of the key automated characterization techniques include high-throughput X-ray diffraction,[79–81] automated spectroscopy techniques,[81] automated electron microscopy,[82–85] automated scanning probe microscopy techniques,[59,82,83,86] and high-throughput optical, electrical, and magnetic properties measurements.[79,87–89] For example, Kalinin *et al.* mentions how human-in-the-loop can be used to improve the traditional microscopy workflow. Automation of experiments for materials discovery is beneficial in terms of cost cutting per experiment, eliminating human error and repetition of menial tasks, and generating significant amounts of data spread over a vast composition and processing parameter space. It is recommended to integrate this methodology for all lab components: synthesis, characterization, and optimization with a human-in-the-loop workflow.

Featured in Fig. 3, the emerging paradigm in materials discovery is to achieve a fully automated workflow, or "closing the loop", by seamlessly integrating all stages from materials synthesis and characterization to data analysis and decision-making into a continuous, computer-controlled feedback loop with advanced automation.[54,90] Some proof-of-concept SDLs demonstrated by different laboratories around the world are the Hitosugi–Shimizu lab in Japan,[91] Cronin[92] and Cooper[93,94] labs in the United Kingdom, Swiss CAT+ in Switzerland,[95] Ada[96] in Canada, and the Hippalgaonkar[97] lab in Singapore, as well as the A-Lab,[98,102] Abolhasani,[37] Ahmadi,[100] Buonassisi,[90] Fenning,[101] Amassian,[98,99] Brown,[73] and Coley[75] labs in the United States. These labs showcase the integration of robotics, AI, and machine learning in materials science to automate and optimize the process of materials synthesis, property evaluation, and discovery. This illustrates a future where much of the labor-intensive and repetitive tasks in scientific research are handled by intelligent, autonomous systems and how these technologies are transforming traditional research methodologies, enabling rapid, high-throughput experimentation with minimal human intervention.

The evolution of automation platforms and the growing accessibility of ML techniques have indeed been pivotal in the emergence and advancement of SDLs and near-complete automation in scientific research. This trend is reflected in the research works of MacLeod,[96,103] Seifrid,[104] Steiner,[92] and Chatterjee,[105] wherein they demonstrate the impressive capabilities of automated systems in optimizing material properties, synthesizing complex compounds, and enabling multistep chemical processes with minimal manual intervention. MacLeod's work in autonomously optimizing the optical and electronic properties of thin-film materials, through adjusting film composition and processing conditions, exemplifies the effectiveness of SDLs in material property refinement for specific applications.[96] This kind of optimization, driven by machine learning and automated processes, represents a significant leap in the field of materials science, especially for clean energy technologies. Seifrid's research, focusing on the balance between manual and automated synthesis, highlights the necessity of human oversight as automation evolves in materials science.[104] This balance is crucial, especially in overcoming the limitations of fully automated systems. However, there is difficulty in automating a streamline of instrumental analysis due to the specific operational requirements of each instrument.[45] Integration of ML needed in SDLs is crucial for

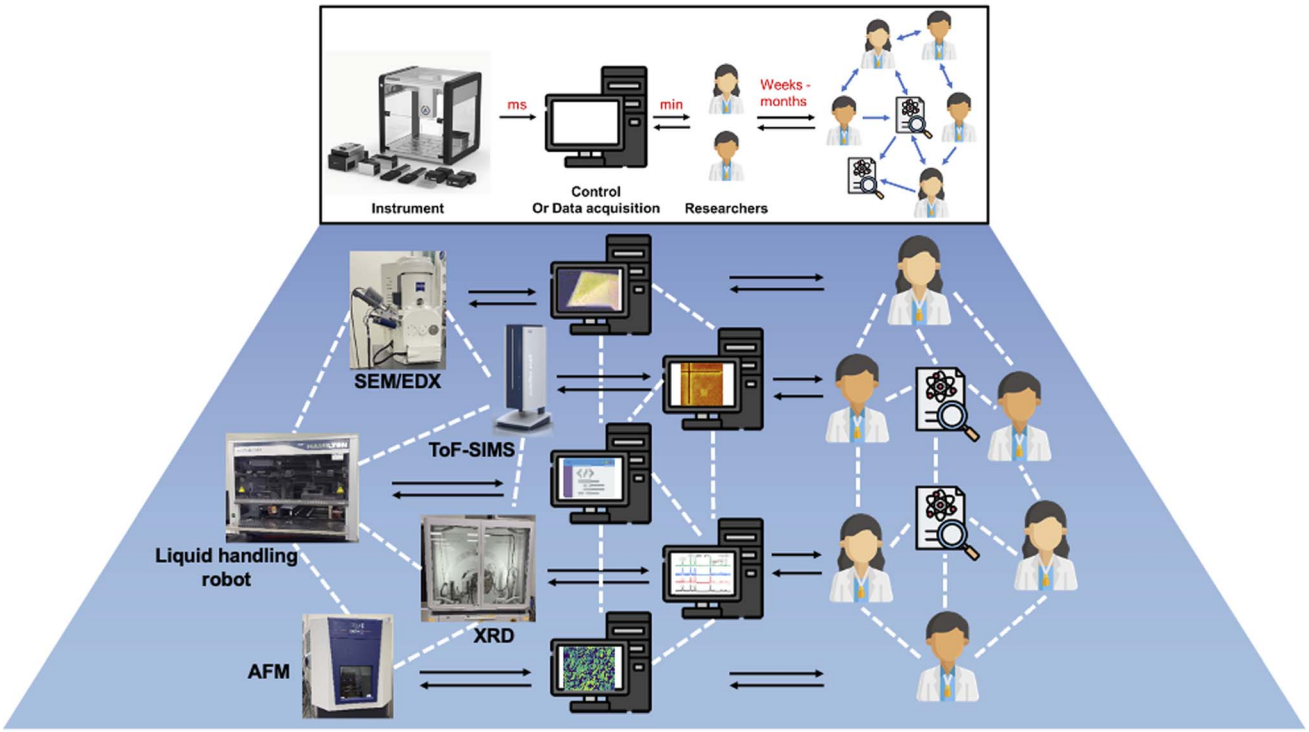

Fig. 3 The diagram on top represents classical research. The diagram at the bottom shows the interconnecting relationship between different instruments and computers for data acquisition (cloudification) and researchers for experiment workflows. Instrumentation and data shown: OT2 liquid handling robot, X-ray diffractometer, scanning electron microscope, atomic force microscope, time-of-flight secondary ionization mass spectrometer.

interpreting data trends across multiple platforms, enhancing the analytical capabilities of these systems.

The expanded vision of materials science research involves not just automating individual laboratories but also linking them in a networked manner to share data, resources, and insights. This could potentially elevate the efficiency and effectiveness of research by leveraging shared technical resources (both experimental and computational) and datasets, thus accelerating the discovery and development of new materials. Such interconnected labs, as envisioned in the HELAO-async framework implemented *via* the Python *asyncio* package,[106] PyLabRobot open-source framework,[107,108] and ORCESTRA based on Pachyderm for data orchestration,[109] would not only handle labor-intensive and repetitive tasks but also foster a more collaborative and integrated approach to scientific research. The HALEO-async framework is designed to facilitate coordination among multiple research workflows, particularly in materials science and chemistry experiments. It utilizes asynchronous programming to manage and orchestrate laboratory automation, enabling adaptive experiments that are more efficient and flexible. PyLabRobot is an open-source framework, based on Python, designed to provide a hardware-agnostic interface for programming a diverse range of liquid-handling robots. It enables seamless and flexible control over robotic laboratory equipment, facilitating automation in various scientific experiments and research settings. ORCESTRA is a cloud-based platform that facilitates the automated and reproducible processing of biomedical data, providing tools for integrating and analyzing diverse molecular and clinical datasets. It leverages Pachyderm orchestration tools to manage data workflows, ensuring transparent and traceable data processing and analysis.

These research works represent the cutting-edge efforts in orchestrating complex research operations across multiple laboratories, highlighting the potential for increased efficiency, collaboration, and innovation in materials science research. Through the integration of automation, advanced computing resources, AI, and data management tools, these frameworks aim to facilitate more seamless, collaborative, and efficient research processes, paving the way for a new era of interconnected scientific exploration for new materials.

IV. Automated labs of the data plane

While the independent SDLs outlined above demonstrate potential scientific impact for autonomous workflows, there is a forward-looking vision for coupling labs and resources across labs and institutions. This approach allows for collaborative efforts with more sophisticated toolsets to interrogate materials and enables more theory guided approaches. However, the desire to unify SDLs into scientific ecosystems poses new sets of challenges.

For SDLs, researchers grapple with diverse formats and metadata associated with various theoretical and experimental generators. The challenge lies in envisioning multi-modal workflows that seamlessly integrate data from diverse sources.

These challenges are compounded by the fact that state-of-the-art data sources were developed in isolation and many tool vendors do not provide adequate application access for autonomous control and data extraction. This necessitates the development of instrument wrappers and containerized data packages, sets of datasets from different generators, containing comprehensive metadata for reproducibility. Such metadata extends beyond provenance and must encompass data reduction and analysis methods encountered at different stages of the workflow. These data packages should be comprehensive as to not rely on human intuition gained from years of experience to fill gaps but accommodate independently developed early-stage machine agents with complementary capabilities.

Managing the flow of data and command-and-control messages poses a challenge in a landscape where unique facilities operate independently. Developing microservices that can interoperate with different institutional entities is essential for ensuring standardized communication across varied hardware and infrastructure. For example, there currently exists over 300 workflow solutions, highlighting the need for a holistic approach within a federated or hybrid onprem/cloud scientific framework.[110] These microservices must harmonize goals for security and intellectual property while respecting individual institutional policies and protocols.

A vision for a unified data flow revolves around a seamlessly connected scientific data ecosystem, where scientific instruments, robot-controlled laboratories, and computing resources intertwine. This vision is being pursued at national laboratories as exemplified by the Interconnected Science Ecosystem (INTERSECT) at Oak Ridge National Laboratory,[111] Superfacility at Lawrence Berkeley National Laboratory[112] and Globus at the University of Chicago and Argonne National Laboratory,[113] but there is a need to extend this vision to connect with university and industrial SDLs. This interconnected future envisions autonomous experiments and self-driving laboratories guided by AI algorithms, reducing human intervention through machine-in-the-loop intelligence. To realize such a vision, a comprehensive federated hardware/software standard, inspired by System of Systems (SoS) design patterns and microservice architectures, is needed to orchestrate autonomous processes and foster standardized communication.[114–116] Such a SoS ecosystem must embody five key characteristics: operational independence of systems, managerial independence of systems, geographical distribution, emergent behavior, and evolutionary development.[115]

The deployment of AI/ML workflows in experimental campaigns will create a seismic shift in data collection. The massive datasets required for training AI/ML agents result in data that human agents may typically discard becoming valuable datapoints, which must be cataloged and maintained. As data volumes surge, laboratories globally must explore innovative approaches to scientific data management within the broader context of scientific ecosystems. Such scientific data ecosystems must adhere to FAIR (findable, accessible, interoperable, and reusable) principles,[117,118] ensuring the findability, accessibility, interoperability, and reusability of data, fostering scientific reproducibility and simplifying experiment replication.[119] The FAIR-principled scientific data ecosystem introduces a paradigm shift from system-centric to data-centric perspectives. Automating this approach streamlines data management tasks, allowing scientists to concentrate on analysis and discovery within an open scientific ecosystem. The broad vision of a scientific data ecosystem makes it an attractive solution compatible across various facilities and domains. Its adherence to FAIR principles ensures the development of uniform APIs for cross-facility implementation and autonomous workflow use cases spanning edge to exascale computing platforms.

A federated hardware/software framework and the scientific data ecosystem represent significant strides toward the laboratories of the future. The former, propelled by intelligent systems and autonomous experiments, aims to redefine scientific exploration, while the latter, through FAIR data principles, seeks to transform how data is managed and shared. Together, they propel scientific discovery and innovation, ushering in a new era of collaborative and intelligent scientific ecosystems. The convergence of intelligent systems paints a promising picture of a future where scientific exploration is not only advanced but also accessible and reproducible across diverse domains.

V. ML for automated labs

The utilization of ML in automated laboratories is revolutionizing the approach to materials discovery, yet this integration is not without its challenges. Traditionally, molecular and materials discovery follows a linear process – beginning with theoretical screening, followed by synthesis and characterization, often without a feedback mechanism.[120,121] This method, however, is evolving with the integration of ML techniques in automated labs, facilitating a more dynamic and responsive approach. The primary challenge in this domain is navigating the high-dimensional and often intractable search spaces corresponding to compositions, processing parameters and histories, or molecular structures. These spaces are defined by their dimensionality, the properties of the parameter space, and the characteristics of the functions targeted for discovery. The dimensionality, completeness, and differentiability of these spaces offer a systematic guide to discerning problems that are suitable for classical ML methods and those that require more innovative strategies. Furthermore, the practical aspect of accessing specific points in the search space is critical in the context of chemical optimization.

The first key parameter is the dimensionality, or the number of independent degrees of freedom, of the search space. For example, a ternary phase diagram represents a two-dimensional search space, while a synthesis process that varies temperature over time would possess an infinite-dimensional parameter space. Another essential element is the completeness of the parameter space, referring to whether all combinations of parameters represent physically realizable scenarios. In natural search spaces like concentrations and temperatures, parameter spaces are typically complete. However, in contexts like phase compositions, the space can be incomplete, as not all compositions are physically possible. This aspect is particularly critical

when applying machine learning to large dimensional spaces, necessitating dimensionality compression techniques such as variational autoencoders (VAEs).[122,123] The use of molecular encodings like SMILES and SELFIES exemplifies this, with SMILES providing an incomplete representation where not all strings correspond to real molecules, while SELFIES offers a complete mapping.[124]

Similarly, the differentiability of the function to be optimized is a crucial consideration. Physical properties like phase transition temperatures or band gaps might be differentiable over certain compositional spaces but become discontinuous across phase boundaries.[125] Even in simple systems, the numerous phases and boundaries present significant challenges for optimizing material properties.

Additionally, practical factors such as experimental budget and synthesizability must be considered. It's often infeasible to experimentally realize a physically possible point in the search space, due to constraints in synthesis steps or costs. Moreover, the interplay of theory and experiment in materials and molecular discovery is pivotal. Computational methods, while leveraging prior knowledge to predict specific material properties, must integrate past experimental and theoretical insights. This iterative balance between theory and experiment influences the predictability of models and guides new compound discovery. From a machine learning standpoint, fully realizing this co-navigation of theory and experiment is an aspirational goal, given the immense dimensionality and the non-differentiable nature of the chemical space, which complicates the application of traditional optimization strategies.

VI. Human in the loop

In the evolving landscape of self-driving labs, especially in materials synthesis, the collaboration between human researchers and robotic systems is vital. Human researchers contribute their skilled expertise and nuanced understanding to laboratory, excelling in tasks that require meticulous attention and precision, such as solving unexpected reactions, separation of materials, precise layering of materials to create advanced nanocomposites, and so on. Their role is particularly crucial in processes that demand a depth of knowledge, intuition, and adaptability.[52]

On the other hand, robots are increasingly playing a crucial role in performing repetitive and data-intensive tasks. Their involvement is particularly noteworthy in elementary steps where performance can be significantly enhanced through automation. For instance, operations such as pipetting[42] and material depositions[91] are areas where robots now demonstrate superior performance due to their precision and consistency. These tasks, often repetitive and requiring exactness, benefit from the robotic systems' ability to execute with minimal error over extended periods. However, this robotic automation is not yet universal across all laboratory processes. Many complex tasks still heavily rely on the expertise and adaptability of human researchers which includes processes that need nuanced decision-making, real-time adjustments based on sensory feedback, and intricate handling of materials.

Moreover, humans excel in innovative problem-solving, especially in areas like materials synthesis, which presents its own set of unique challenges. Where a machine might be stumped by an unexpected issue, a researcher can think laterally, drawing from diverse experiences to explore new methodologies or craft custom solutions allowing them venture into uncharted territories of material development and unlock potential breakthroughs. In essence, while machines and systems might be limited to the data and algorithms they've been provided, human ingenuity can introduce fresh perspectives, ensuring continual advancement in materials development. A human researcher, with their adaptive problem-solving skills, can quickly identify anomalies, adjust the synthesis parameters, or modify the process to mitigate these issues. Additionally, if an instrument malfunctions during an experiment, a human can often troubleshoot and fix the issue on the spot or find alternative ways to continue the work.

This adaptability extends to the data analysis post-experimentation. Human researchers can provide a depth of judgment and perspective to enable AI to adeptly interpret complex data, identifying subtle patterns or anomalies, and making strategic decisions along with human intervention that align the scientific process with broader research goals and real-world applications. Translational AI aims to blend the strengths of human scientists with ML solutions to enhance the entire experimental cycle, from synthesis to imaging.[126] This approach involves not just leveraging AI for data processing but also for strategic planning and decision-making, guided by human insights. By capitalizing on recent investments in educational systems and infrastructure, such as collaborations with national labs and industrial partnerships, the aim is to forge a new paradigm in AI-driven experimental science that will propel fields like materials science into a new era of innovation and discovery.

Machine learning, particularly reinforcement learning (RL), has shown significant promise in controlled simulations but often encounters challenges in real-world scenarios, where reward functions are complex and not easily defined.[127,128] This is particularly evident in fields where experiments are guided by long-term objectives, such as combating climate change or developing new energy technologies, rather than immediate outcomes. In these contexts, human expertise becomes crucial for designing reward functions that align with such nuanced real-world goals, such as combating climate change or developing new energy technologies, where the true impact of research may only be discernible over extended periods.[129]

In Fig. 4, the dynamic interaction between scientists and the experimental process is key to refining ML applications, especially when leveraging the adaptability of large language models (LLMs) like ChatGPT,[130] Copilot,[131] Tabnine,[132] and Gemini,[133] to enhance these workflows. These LLMs can assist in optimizing experiment designs by suggesting viable parameters and material compositions, thus navigating vast databases to pinpoint promising research avenues.[134] Additionally, they are instrumental in analyzing complex data, identifying patterns and providing key insights, thereby focusing researchers' attention on the most pertinent findings.[135] Moreover, the iterative learning process inherent in scientific research is

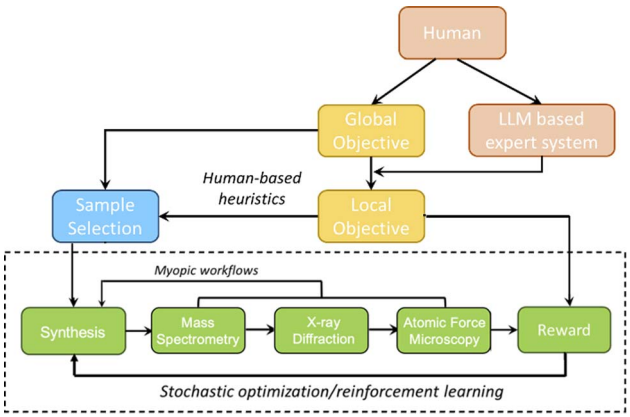

Fig. 4 A diagram highlighting the process of human-in-the-loop discovery with the reward system interacting on materials synthesis and characterization methods. Characterization methods listed: mass spectrometry, X-ray diffraction, and atomic force microscopy. Each aspect plays a role in the myopic workflow with synthesis that finally leads to a reward.

enhanced by LLMs, which adapt and refine their outputs based on the feedback from experimental results. This continuous interaction facilitates the improvement of experimental designs and theoretical models, making them more aligned with the intricacies of material behaviors and properties.

Ultimately, the integration of human intelligence with domain-specific ML models is proposed as a pathway to develop robust workflows that can navigate the complexities of real-world scientific challenges. This approach not only harnesses the computational power of ML but also incorporates the critical, nuanced understanding and creativity of human expertise, essential for tackling long-term, complex scientific objectives.

Hypothesis generation

Large Language Models (LLMs) like GPT-4, while not fully replicating the enigmatic nature of human creativity, have emerged as powerful tools in guiding and augmenting human creative processes, particularly in hypothesis generation. The exact nature of human creativity remains a complex and largely uncharted territory; however, LLMs, with their vast repository of knowledge and advanced pattern recognition capabilities, can significantly aid in navigating this realm.[136] They excel in quickly assimilating and synthesizing information from a multitude of sources, enabling them to suggest diverse and innovative hypotheses.[137] This is particularly useful in rapidly evolving or interdisciplinary fields, where new insights are constantly emerging. By analyzing existing data and trends, LLMs can identify potential areas of exploration that might not be immediately apparent to human researchers. Their ability to process and integrate cross-disciplinary knowledge allows them to propose hypotheses that bridge different areas of study, thereby fostering novel perspectives and approaches. While they do not replace the intrinsic creativity of the human mind, LLMs serve as a complementary tool, expanding the horizons of human thought and enabling a more rapid exploration of new fields and ideas. Their role in hypothesis generation is thus not just as an automated generator of ideas, but also as a catalyst that enhances and directs human creativity towards unexplored possibilities.

Reward engineering

Machine learning methods are progressively considered as a part of real-world technological solutions, including workflows for materials synthesis and optimization, computation, imaging, and characterization. However, methods such as reinforcement learning (RL) that had been shown to be highly effective in simulated environments such as computer games or simulations are often inadequate for real-world applications. One of the key elements of RL is a reward function that is made available for the algorithm during the training.[138–141] As demonstrated in Snapp *et al.*, methods like Bayesian optimization can be easily integrated into SDL environments. However, for many real-world problems the reward functions available at the end of the experimental campaign (or after several steps) are absent; rather the experiments are motivated by the long-term objectives. Designing a reward function that adequately represents real-world objectives and does not lead to reward hacking is a challenge. Similarly, very often experimental results can contribute to multiple objectives, with fundamental scientific research being the most notable example of such activity.

As an example of such a problem, consider climate change, the problem motivating multi-billion-dollar investments over the globe. Minimizing climate change is a very long-term objective. The lower rank objectives are the development of solar and wind energy and associated grid-level storage and effective energy transport methods, along with the technologies for direct carbon capture. The even lower rank objectives are the development of cheap, environmentally friendly, and stable chemistries for grid storage. None of these objectives can be translated into a reward for an experimental campaign. Rather, these objectives serve as a motivation for experiment planning – and the reward is often a short-term battery performance or observation of a specific mechanism under a microscope[138,139] that can suggest potential ways to improve the battery materials.

We believe that the discovery of short-term rewards can be used for hypothesis making to guide experimental research. Additionally, reward functions can guide and ascertain the success of experimental campaigns as the missing link required to connect ML to real-world applications. As potential pathways to address this challenge, we can consider:

(1) Literature mining towards building directed acyclic graphs (DAGs) connecting experimental outcomes (rewards) and objectives (motivation).

(2) Technoeconomic analysis of past publication outcomes.

(3) Crowdsourcing to the community of experts (a.k.a. "what would be the potential of high temperature superconductivity to change the world" to "how does the phase separation in cuprates affect peak-effect and losses").

With this, we aim to create the capability to separate the specific objective into the probabilistic graph of short-term reward functions that can guide experiment planning and establish measures of success. Naturally, these reward functions will be probabilistic, and the value of the real-world

experiment can affect (much) more than one objective. For example, mechanisms of metal–air interactions can be used both for corrosion mitigation and for design of metal–air batteries. The important element of this approach is that humans are part of the theory–experiment loop – and hence the structure of the rewards can be amended *via* human feedback on the observations (much like science works now).

Notably, the LLMs are often capable of making the connection between high- and lower-level objectives (*e.g.* prompts "what should I study with a microscope to understand plasticity" gives very plausible answers). This action provides a starting point for a school of thought that may not already be obvious to the person posing the question or from conversations with their peers. LLMs can be utilized to create smaller stepping stones towards larger, complex issues that may require multi-year efforts to resolve. For example, prompting ChatGPT with "I want to solve the climate crisis, what should I study" can provide a pathway to discuss renewable energy sources such as solar, and result in determining what kind of experiments should be conducted for hybrid perovskites.[142] Presumably, complementing LLMs with models trained on domain-specific literature can both allow systematic developments of such workflows and their integration across multiple domains following common rewards.[137]

Homo ludens

We believe that the future of materials synthesis lies in a harmonious blend of human expertise and technological advancement. While ML and automation are transforming this field, the indispensable role of human insight, creativity, and strategic thinking cannot be replaced. The most effective and innovative outcomes in materials synthesis will emerge from workflows that skillfully combine the strengths of both humans and machines. This integration is imperative because the reward functions, which drive workflow development, must originate from human objectives and goals. Additionally, humans can collaborate with AI in refining hypotheses and recognizing serendipitous experimental findings, a partnership that helps mitigate potential misalignments between ML-generated workflows and human intentions. This collaborative dynamic raises a compelling question: what is the optimal form of human-in-the-loop technology, tailored to human preferences and capabilities?

The response to this question varies with the complexity of the workflow. In the case of automated microscopy, for instance, it involves fine-tuning reward functions and exploration policies.[83] Given the limited range of operations in microscopy, the resulting meta-controls are relatively straightforward, albeit

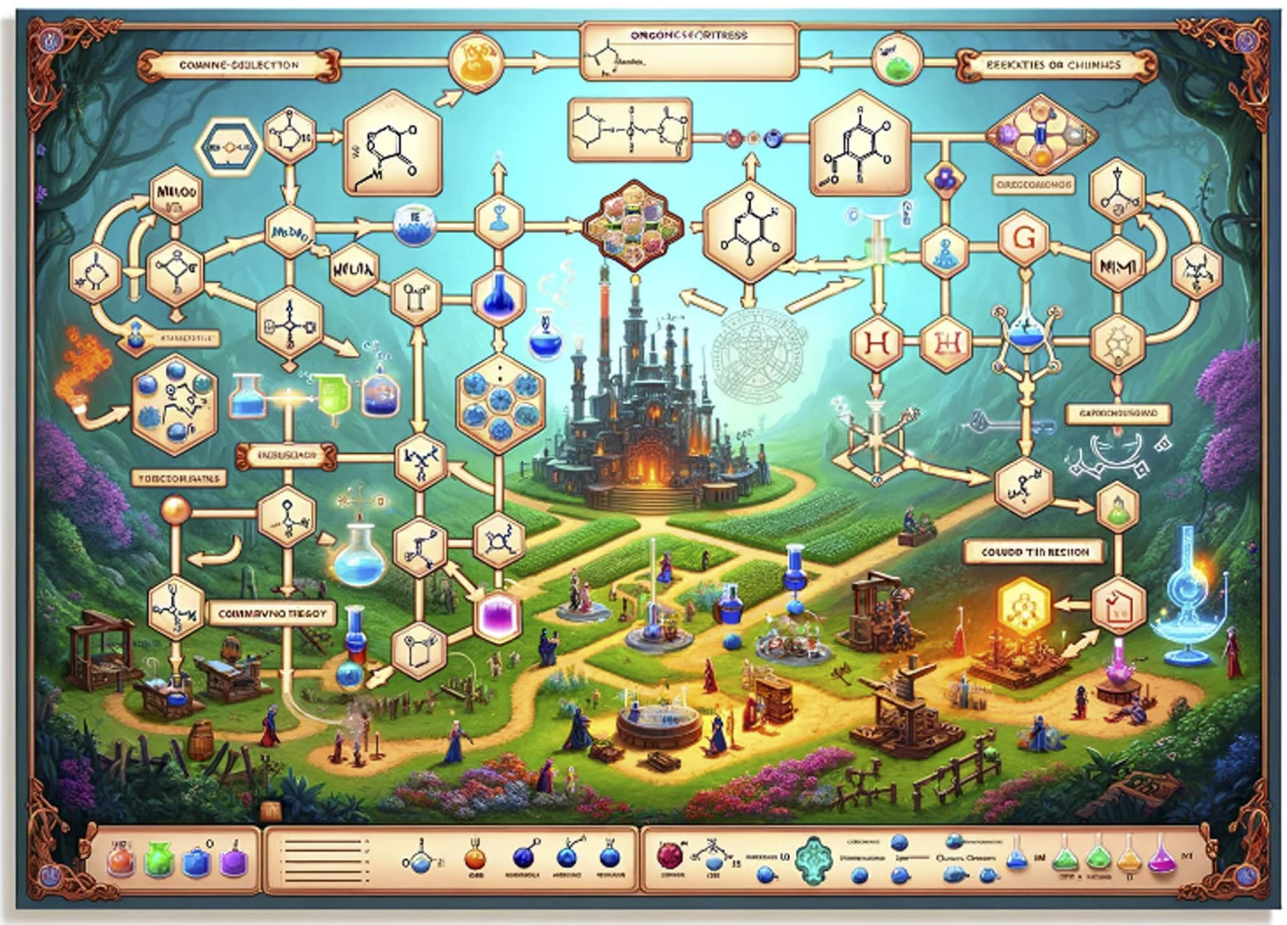

Fig. 5 An example of gamification used for designing scientific workflows based on role-playing games.

unfamiliar to traditional microscopists, indicating a novice user as the target audience. In contrast, materials discovery presents a more formidable challenge due to the complexity, branching, and interconnectivity of workflows.

Looking ahead 10–20 years, as workflows intertwine across multiple fields, the situation grows even more complex. Human actors will likely struggle to fully grasp these workflows, which span diverse domains and physical locations. Similarly, ML alone may not suffice, as these workflows amalgamate elements of the physical world that are partially unknown, and sustainable systems will require not just prediction and experimentation, but also tool-making – with the tools themselves possessing partially unknown properties.

As shown in Fig. 5, a paradoxical yet intriguing possibility is conceptualizing human-in-the-loop R&D as a Massive Multiplayer Online Role-Playing Game (MMORPG). In this framework, activities like magic, forging, and potion making symbolize real-world processes in physics, engineering, and chemistry, with AI acting as an advanced intermediary, connecting humans to remote-controlled and automated research and manufacturing tools.

In the interim, and on a more pragmatic note, gamifying human-in-the-loop ML presents an engaging user experience (UX) concept. Gather.Town has already experimented with this idea for virtual meetings, but the potential for further development is vast. Thus, the future might well embrace the concept of Homo Ludens – humans engaged in playful yet productive interaction with technology.

VII. Summary

In conclusion, advancing towards complete automation in SDLs is a complex process that requires a collaborative approach by combining the expertise of various professionals from diverse fields such as materials science, instrumentation, mechatronics, software development, and data science.[143] This interdisciplinary teamwork is critical to overcome the substantial challenges with achieving seamless operational flows in automation, which are often impeded due to procedural errors and system stoppages.[93,144] While SDLs offer remarkable efficiency and data generation capabilities, the essential role of human researchers remains irreplaceable, underscoring the need for their continued involvement in guiding and overseeing these automated systems. Collaboration between academia and industry, combined with the integration of AI algorithms, holds promise to overcome challenges in achieving seamless operational flow in automation. By leveraging diverse expertise and advanced technologies, such a partnership could drive innovation, streamline processes, and enhance the efficacy and reliability of automated systems.

We believe that the future of materials synthesis lies in a harmonious blend of human expertise and technological advancement. While ML and automation are transforming the field, the value of human insight, creativity, and strategic thinking remains irreplaceable. The most effective and innovative outcomes in materials synthesis will emerge from workflows that skillfully combine the strengths of both humans and machines. This integration is imperative because the reward functions, which drive workflow development, must originate from human objectives and goals. Additionally, humans can collaborate with AI in refining hypotheses and recognizing serendipitous experimental findings, a partnership that helps mitigate potential misalignments between ML-generated workflows and human intentions. This collaborative dynamic raises a compelling question: what is the optimal form of human-in-the-loop technology, tailored to human preferences and capabilities.

Data availability

As this is a Perspective article, no primary research results, data, software or code has been included.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

HH was supported by the DOE Office of Science Research Program for Microelectronics Codesign (sponsored by ASCR, BES, HEP, NP, and FES) through the Abisko Project with program managers Robinson Pino (ASCR), Hal Finkel (ASCR), and Andrew Schwartz (BES). EF and MA acknowledge support from the National Science Foundation (NSF), award number no. 2043205 and Alfred P. Sloan Foundation, award no. FG-2022-18275. RGM acknowledges support from the INTERSECT Initiative as part of the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725. This effort (SVK; ML and human in the loop) was supported as part of the center for 3D Ferroelectric Microelectronics (3DFeM), an Energy Frontier Research Center funded by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences under award number DE-SC0021118.

References

1 J. D. Martin and C. C. M. Mody, *Materials Science, Encyclopedia of the History of Science*, 2020.
2 C. K. N. Patel, D. S. Kukich, R. L. McCullough, L. A. Girifalco, R. E. Marchant and J. D. Venables, *Materials science, Encyclopedia Britannica*, 2024, **https://www.britannica.com/technology/materials-science**.
3 J. Cai, *et al.*, Machine learning-driven new material discovery, *Nanoscale Adv.*, 2020, **2**(8), 3115–3130.
4 M. L. Green, B. Maruyama and J. Schrier, Autonomous (AI-driven) materials science, *Appl. Phys. Rev.*, 2022, **9**(3), 030401, DOI: **10.1063/5.0118872**.
5 G. R. Schleder, A. C. Padilha, C. M. Acosta, M. Costa and A. Fazzio, From DFT to machine learning: recent approaches to materials science-a review, *J. Phys.: Mater.*, 2019, **2**(3), 032001.

6 S. P. Ong, *et al.*, Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis, *Comput. Mater. Sci.*, 2013, **68**, 314–319.
7 E. Stach, B. DeCost, A. G. Kusne, J. Hattrick-Simpers, K. A. Brown, K. G. Reyes, J. Schrier, S. Billinge, T. Buonassisi, I. Foster, *et al.*, Autonomous experimentation systems for materials development: a community perspective, *Matter*, 2021, **4**(9), 2702–2726, DOI: **10.1016/j.matt.2021.06.036**.
8 N. J. Szymanski, P. Nevatia, C. J. Bartel, Y. Zeng and G. Ceder, Autonomous and dynamic precursor selection for solid-state materials synthesis, *Nat. Commun.*, 2023, **14**(1), 6956, DOI: **10.1038/s41467-023-42329-9**.
9 H. Hysmith, S. Y. Park, J. Yang, A. V. Ievlev, Y. Liu, K. Zhu, B. G. Sumpter, J. Berry, M. Ahmadi and O. S. Ovchinnikova, The Role of SnO2 Processing on Ionic Distribution in Double-Cation–Double Halide Perovskites, *ACS Appl. Mater. Interfaces*, 2023, **15**(30), 36856–36865, DOI: **10.1021/acsami.3c03520**.
10 O. Er-raji, L. Rustam, B. P. Kore, S. W. Glunz and P. S. C. Schulze, Insights into Perovskite Film Formation Using the Hybrid Evaporation/Spin-Coating Route: An In Situ XRD Study, *ACS Appl. Energy Mater.*, 2023, **6**(11), 6183–6193, DOI: **10.1021/acsaem.3c00698**.
11 A. Magazzù and C. Marcuello, Investigation of Soft Matter Nanomechanics by Atomic Force Microscopy and Optical Tweezers: A Comprehensive Review, *Nanomaterials*, 2023, **13**(6), 963.
12 J. Nie, B. Zhou, S. Fang, Y. Wang, Y. Wang, B. Tian, H. Hu, H. Zhong, H. Li and Y. Shi, Chemical doping of lead-free metal-halide-perovskite related materials for efficient white-light photoluminescence, *Mater. Today Phys.*, 2023, **31**, 100992, DOI: **10.1016/j.mtphys.2023.100992**.
13 S. M. Park, M. Wei, J. Xu, H. R. Atapattu, F. T. Eickemeyer, K. Darabi, L. Grater, Y. Yang, C. Liu and S. Teale, Engineering ligand reactivity enables high-temperature operation of stable perovskite solar cells, *Science*, 2023, **381**(6654), 209–215.
14 W. McKinney, *Python for data analysis: data wrangling with Pandas, NumPy, and IPython*, O'Reilly Media, Inc., 2012.
15 D. T. Valentine and B. H. Hahn, *Essential MATLAB for engineers and scientists*, Academic Press, 2022.
16 G. R. Bhimanapati, Z. Lin, V. Meunier, Y. Jung, J. Cha, S. Das, D. Xiao, Y. Son, M. S. Strano, V. R. Cooper, *et al.*, Recent Advances in Two-Dimensional Materials beyond Graphene, *ACS Nano*, 2015, **9**(12), 11509–11539, DOI: **10.1021/acsnano.5b05556**.
17 S. Yang, W. Fu, Z. Zhang, H. Chen and C.-Z. Li, Recent advances in perovskite solar cells: efficiency, stability and lead-free perovskite, *J. Mater. Chem. A*, 2017, **5**(23), 11462–11482, DOI: **10.1039/C7TA00366H**.
18 Y. Yan, J. Gong, J. Chen, Z. Zeng, W. Huang, K. Pu, J. Liu and P. Chen, Recent Advances on Graphene Quantum Dots: From Chemistry and Physics to Applications, *Adv. Mater.*, 2019, **31**(21), 1808283, DOI: **10.1002/adma.201808283**.
19 Z. Pu, T. Liu, I. S. Amiinu, R. Cheng, P. Wang, C. Zhang, P. Ji, W. Hu, J. Liu and S. Mu, Transition-Metal Phosphides: Activity Origin, Energy-Related Electrocatalysis Applications, and Synthetic Strategies, *Adv. Funct. Mater.*, 2020, **30**(45), 2004009, DOI: **10.1002/adfm.202004009**.
20 M. Safaei, M. M. Foroughi, N. Ebrahimpoor, S. Jahani, A. Omidi and M. Khatami, A review on metal-organic frameworks: Synthesis and applications, *TrAC, Trends Anal. Chem.*, 2019, **118**, 401–425, DOI: **10.1016/j.trac.2019.06.007**.
21 Y. L. Dar, High-Throughput Experimentation: A Powerful Enabling Technology for the Chemicals and Materials Industry, *Macromol. Rapid Commun.*, 2004, **25**(1), 34–47, DOI: **10.1002/marc.200300166**.
22 S. A. Biyani, Y. W. Moriuchi and D. H. Thompson, Advancement in Organic Synthesis Through High Throughput Experimentation, *Chem.: Methods*, 2021, **1**(7), 323–339, DOI: **10.1002/cmtd.202100023**.
23 R. Hoogenboom, M. A. R. Meier and U. S. Schubert, Combinatorial Methods, Automated Synthesis and High-Throughput Screening in Polymer Research: Past and Present, *Macromol. Rapid Commun.*, 2003, **24**(1), 15–32, DOI: **10.1002/marc.200390013**.
24 S. Baudis and M. Behl, High-Throughput and Combinatorial Approaches for the Development of Multifunctional Polymers, *Macromol. Rapid Commun.*, 2022, **43**(12), 2100400, DOI: **10.1002/marc.202100400**.
25 R. Potyrailo, K. Rajan, K. Stoewe, I. Takeuchi, B. Chisholm and H. Lam, Combinatorial and High-Throughput Screening of Materials Libraries: Review of State of the Art, *ACS Comb. Sci.*, 2011, **13**(6), 579–633, DOI: **10.1021/co200007w**.
26 W. F. Maier, Early Years of High-Throughput Experimentation and Combinatorial Approaches in Catalysis and Materials Science, *ACS Comb. Sci.*, 2019, **21**(6), 437–444, DOI: **10.1021/acscombsci.8b00189**.
27 L. Buglioni, F. Raymenants, A. Slattery, S. D. A. Zondag and T. Noël, Technological Innovations in Photochemistry for Organic Synthesis: Flow Chemistry, High-Throughput Experimentation, Scale-up, and Photoelectrochemistry, *Chem. Rev.*, 2022, **122**(2), 2752–2906, DOI: **10.1021/acs.chemrev.1c00332**.
28 I. G. Clayson, D. Hewitt, M. Hutereau, T. Pope and B. Slater, High Throughput Methods in the Synthesis, Characterization, and Optimization of Porous Materials, *Adv. Mater.*, 2020, **32**(44), 2002780, DOI: **10.1002/adma.202002780**.
29 S. Langner, F. Häse, J. D. Perea, T. Stubhan, J. Hauch, L. M. Roch, T. Heumueller, A. Aspuru-Guzik and C. J. Brabec, Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize Multicomponent Systems, *Adv. Mater.*, 2020, **32**(14), 1907801, DOI: **10.1002/adma.201907801**.
30 W. H. Teh, V. Chaudhary, S. Chen, S. H. Lim, F. Wei, J. Y. Lee, P. Wang, S. P. Padhy, C. C. Tan and R. V. Ramanujan, High throughput multi-property evaluation of additively manufactured Co-Fe-Ni materials

libraries, *Addit. Manuf.*, 2022, **58**, 102983, DOI: **10.1016/j.addma.2022.102983**.

31 S. P. Padhy, L. P. Tan, V. B. Varma, V. Chaudhary, Z. Tsakadze and R. V. Ramanujan, Accelerated multi-property screening of Fe–Co–Ni alloy libraries by hyper-heuristic combinatorial flow synthesis and high-throughput spark plasma sintering, *J. Mater. Res. Technol.*, 2023, **27**, 2976–2988, DOI: **10.1016/j.jmrt.2023.10.124**.

32 S. P. Padhy, Z. Tsakadze, V. Chaudhary, G. J. Lim, X. Tan, W. S. Lew and R. V. Ramanujan, Rapid multi-property assessment of compositionally modulated Fe-Co-Ni thin film material libraries, *Results Mater.*, 2022, **14**, 100283, DOI: **10.1016/j.rinma.2022.100283**.

33 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek and A. Potapenko, Highly accurate protein structure prediction with AlphaFold, *Nature*, 2021, **596**(7873), 583–589.

34 D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai and A. Bolton, Mastering the game of go without human knowledge, *Nature*, 2017, **550**(7676), 354–359.

35 E. O. Pyzer-Knapp, J. W. Pitera, P. W. Staar, S. Takeda, T. Laino, D. P. Sanders, J. Sexton, J. R. Smith and A. Curioni, Accelerating materials discovery using artificial intelligence, high performance computing and robotics, *npj Comput. Mater.*, 2022, **8**(1), 84.

36 H. Tao, T. Wu, S. Kheiri, M. Aldeghi, A. Aspuru-Guzik and E. Kumacheva, Self-driving platform for metal nanoparticle synthesis: combining microfluidics and machine learning, *Adv. Funct. Mater.*, 2021, **31**(51), 2106725.

37 R. W. Epps, A. A. Volk, M. Y. Ibrahim and M. Abolhasani, Universal self-driving laboratory for accelerated discovery of materials and molecules, *Chem*, 2021, **7**(10), 2541–2545.

38 F. Häse, L. M. Roch and A. Aspuru-Guzik, Next-generation experimentation with self-driving laboratories, *Trends Chem.*, 2019, **1**(3), 282–291.

39 M. Abolhasani and E. Kumacheva, The rise of self-driving labs in chemical and materials sciences, *Nat. Synth.*, 2023, 1–10.

40 B. P. MacLeod, F. G. Parlane, C. C. Rupnow, K. E. Dettelbach, M. S. Elliott, T. D. Morrissey, T. H. Haley, O. Proskurin, M. B. Rooney and N. Taherimakhsousi, A self-driving laboratory advances the Pareto front for material properties, *Nat. Commun.*, 2022, **13**(1), 995.

41 S. Matsuda, G. Lambard and K. Sodeyama, Data-driven automated robotic experiments accelerate discovery of multi-component electrolyte for rechargeable Li–O2 batteries, *Cell Rep. Phys. Sci.*, 2022, **3**(4), 100832, DOI: **10.1016/j.xcrp.2022.100832**.

42 S. L. Sanchez, Y. Tang, B. Hu, J. Yang and M. Ahmadi, Understanding the ligand-assisted reprecipitation of $CsPbBr_3$ nanocrystals via high-throughput robotic synthesis approach, *Matter*, 2023, **6**(9), 2900–2918, DOI: **10.1016/j.matt.2023.05.023**.

43 A. Ghosh, B. G. Sumpter, O. Dyck, S. V. Kalinin and M. Ziatdinov, Ensemble learning-iterative training machine learning for uncertainty quantification and automated experiment in atom-resolved microscopy, *npj Comput. Mater.*, 2021, **7**(1), 100, DOI: **10.1038/s41524-021-00569-7**.

44 R. K. Vasudevan, K. P. Kelley, J. Hinkle, H. Funakubo, S. Jesse, S. V. Kalinin and M. Ziatdinov, Autonomous Experiments in Scanning Probe Microscopy and Spectroscopy: Choosing Where to Explore Polarization Dynamics in Ferroelectrics, *ACS Nano*, 2021, **15**(7), 11253–11262, DOI: **10.1021/acsnano.0c10239**.

45 M. Seifrid, R. Pollice, A. Aguilar-Granda, Z. Morgan Chan, K. Hotta, C. T. Ser, J. Vestfrid, T. C. Wu and A. Aspuru-Guzik, Autonomous Chemical Experiments: Challenges and Perspectives on Establishing a Self-Driving Lab, *Acc. Chem. Res.*, 2022, **55**(17), 2454–2466, DOI: **10.1021/acs.accounts.2c00220**.

46 F. Delgado-Licona and M. Abolhasani, Research Acceleration in Self-Driving Labs: Technological Roadmap toward Accelerated Materials and Molecular Discovery, *Adv. Intell. Syst.*, 2023, **5**(4), 2200331, DOI: **10.1002/aisy.202200331**.

47 D. P. Tabor, L. M. Roch, S. K. Saikin, C. Kreisbeck, D. Sheberla, J. H. Montoya, S. Dwaraknath, M. Aykol, C. Ortiz and H. Tribukait, Accelerating the discovery of materials for clean energy in the era of smart automation, *Nat. Rev. Mater.*, 2018, **3**(5), 5–20.

48 A. Aspuru-Guzik and K. Persson, *Materials Acceleration Platform: Accelerating Advanced Energy Materials Discovery by Integrating High-Throughput Methods and Artificial Intelligence*, Mission Innovation, 2018.

49 M. M. Flores-Leonar, L. M. Mejía-Mendoza, A. Aguilar-Granda, B. Sanchez-Lengeling, H. Tribukait, C. Amador-Bedolla and A. Aspuru-Guzik, Materials acceleration platforms: on the way to autonomous experimentation, *Curr. Opin. Green Sustainable Chem.*, 2020, **25**, 100370.

50 A. Vriza, H. Chan and J. Xu, Self-Driving Laboratory for Polymer Electronics, *Chem. Mater.*, 2023, **35**(8), 3046–3056, DOI: **10.1021/acs.chemmater.2c03593**.

51 R. J. Hickman, P. Bannigan, Z. Bao, A. Aspuru-Guzik and C. Allen, Self-driving laboratories: a paradigm shift in nanomedicine development, *Matter*, 2023, **6**(4), 1071–1081, DOI: **10.1016/j.matt.2023.02.007**.

52 M. Abolhasani, K. A. Brown and E. Guest, Role of AI in experimental materials science, *MRS Bull.*, 2023, **48**(2), 134–141, DOI: **10.1557/s43577-023-00482-y**.

53 P. M. Maffettone, P. Friederich, S. G. Baird, B. Blaiszik, K. A. Brown, S. I. Campbell, O. A. Cohen, R. L. Davis, I. T. Foster, N. Haghmoradi, *et al.*, What is missing in autonomous discovery: open challenges for the community, *Digital Discovery*, 2023, **2**(6), 1644–1659, DOI: **10.1039/D3DD00143A**.

54 H. S. Stein and J. M. Gregoire, Progress and prospects for accelerating materials science with automated and autonomous workflows, *Chem. Sci.*, 2019, **10**(42), 9640–9649, DOI: **10.1039/C9SC03766G**.

55 G. Cattabriga, E. M. Chan, S. A. Friedler, Z. Li, M. A. Najeeb, A. J. Norquist, I. M. Pendleton and J. Schrier, Experiment Specification, Capture and Laboratory Automation Technology (ESCALATE): a software pipeline for automated chemical experimentation and data management, *MRS Commun.*, 2019, **9**(3), 846–859, DOI: 10.1557/mrc.2019.72.

56 E. Soedarmadji, H. S. Stein, S. K. Suram, D. Guevarra and J. M. Gregoire, Tracking materials science data lineage to manage millions of materials experiments and analyses, *npj Comput. Mater.*, 2019, **5**(1), 79, DOI: 10.1038/s41524-019-0216-x.

57 J. O'Mara, B. Meredig and K. Michel, Materials Data Infrastructure: A Case Study of the Citrination Platform to Examine Data Import, Storage, and Access, *JOM*, 2016, **68**(8), 2031–2034, DOI: 10.1007/s11837-016-1984-0.

58 B. Blaiszik, K. Chard, J. Pruyne, R. Ananthakrishnan, S. Tuecke and I. Foster, The Materials Data Facility: Data Services to Advance Materials Science Research, *JOM*, 2016, **68**(8), 2045–2052, DOI: 10.1007/s11837-016-2001-3.

59 Y. Liu, K. P. Kelley, R. K. Vasudevan, W. Zhu, J. Hayden, J.-P. Maria, H. Funakubo, M. A. Ziatdinov, S. Trolier-McKinstry and S. V. Kalinin, Automated Experiments of Local Non-Linear Behavior in Ferroelectric Materials, *Small*, 2022, **18**(48), 2204130, DOI: 10.1002/smll.202204130.

60 S. M. Mennen, C. Alhambra, C. L. Allen, M. Barberis, S. Berritt, T. A. Brandt, A. D. Campbell, J. Castañón, A. H. Cherney, M. Christensen, *et al.*, The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future, *Org. Process Res. Dev.*, 2019, **23**(6), 1213–1242, DOI: 10.1021/acs.oprd.9b00140.

61 D. A. Pereira and J. A. Williams, Origin and evolution of high throughput screening, *Br. J. Pharmacol.*, 2007, **152**(1), 53–61, DOI: 10.1038/sj.bjp.0707373.

62 D. G. Karp, D. Cuda, D. Tandel, K. Danh, P. V. Robinson, D. Seftel, H. Tian, M. Pandori, K. W. P. Miller and C.-T. Tsai, Sensitive and Specific Detection of SARS-CoV-2 Antibodies Using a High-Throughput, Fully Automated Liquid-Handling Robotic System, *SLAS Technol.*, 2020, **25**(6), 545–552, DOI: 10.1177/2472630320950663.

63 L. M. Mayr and P. Fuerst, The Future of High-Throughput Screening, *SLAS Discovery*, 2008, **13**(6), 443–448, DOI: 10.1177/1087057108319644.

64 S. W. Krska, D. A. DiRocco, S. D. Dreher and M. Shevlin, The Evolution of Chemical High-Throughput Experimentation To Address Challenging Problems in Pharmaceutical Synthesis, *Acc. Chem. Res.*, 2017, **50**(12), 2976–2985, DOI: 10.1021/acs.accounts.7b00428.

65 F. Kong, L. Yuan, Y. F. Zheng and W. Chen, Automatic Liquid Handling for Life Science: A Critical Review of the Current State of the Art, *J. Lab. Autom.*, 2012, **17**(3), 169–185, DOI: 10.1177/2211068211435302.

66 B. M. Everhart, R. Rao, P. Nikolaev, T.-W. Liu, D. A. Gómez-Gualdrón, B. Maruyama and P. B. Amama, High-Throughput Experimentation for Selective Growth of Small-Diameter Single-Wall Carbon Nanotubes Using Ru-Promoted Co Catalysts, *Chem. Mater.*, 2022, **34**(10), 4548–4559, DOI: 10.1021/acs.chemmater.2c00347.

67 H. Koinuma and I. Takeuchi, Combinatorial solid-state chemistry of inorganic materials, *Nat. Mater.*, 2004, **3**(7), 429–438.

68 I. Takeuchi, J. Lauterbach and M. J. Fasolka, Combinatorial materials synthesis, *Mater. Today*, 2005, **8**(10), 18–26, DOI: 10.1016/S1369-7021(05)71121-4.

69 J. Yang, B. J. Lawrie, S. V. Kalinin and M. Ahmadi, High-Throughput Automated Exploration of Phase Growth Kinetics in Quasi-2D Formamidinium Metal Halide Perovskites, *Adv. Energy Mater.*, 2023, **13**(43), 2302337.

70 F. Kotz, P. Risch, D. Helmer and B. E. Rapp, High-Performance Materials for 3D Printing in Chemical Synthesis Applications, *Adv. Mater.*, 2019, **31**(26), 1805982, DOI: 10.1002/adma.201805982.

71 V. Chaudhary, S. A. Mantri, R. V. Ramanujan and R. Banerjee, Additive manufacturing of magnetic materials, *Prog. Mater. Sci.*, 2020, **114**, 100688, DOI: 10.1016/j.pmatsci.2020.100688.

72 A. E. Gongora, B. Xu, W. Perry, C. Okoye, P. Riley, K. G. Reyes, E. F. Morgan and K. A. Brown, A Bayesian experimental autonomous researcher for mechanical design, *Sci. Adv.*, 2020, **6**(15), eaaz1708, DOI: 10.1126/sciadv.aaz1708.

73 A. E. Gongora, K. L. Snapp, E. Whiting, P. Riley, K. G. Reyes, E. F. Morgan and K. A. Brown, Using simulation to accelerate autonomous experimentation: a case study using mechanics, *iScience*, 2021, **24**(4), DOI: 10.1016/j.isci.2021.102262.

74 S. Marre and K. F. Jensen, Synthesis of micro and nanostructures in microfluidic systems, *Chem. Soc. Rev.*, 2010, **39**(3), 1183–1202, DOI: 10.1039/B821324K.

75 C. W. Coley, D. A. Thomas III, J. A. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers and H. Gao, A robotic platform for flow synthesis of organic compounds informed by AI planning, *Science*, 2019, **365**(6453), eaax1566.

76 Z. Zou, B. Wang, C. Wang, Y. Xiwen, Y. Cao, L.-F. Gao, W. Luo, C. Wu, Y. Yao and Z. Lin, General synthesis of high-entropy alloy and ceramic nanoparticles in nanoseconds, *Nat. Synth.*, 2022, **1**, 138–146.

77 R. Wendelbo, D. E. Akporiaye, A. Karlsson, M. Plassen and A. Olafsen, Combinatorial hydrothermal synthesis and characterisation of perovskites, *J. Eur. Ceram. Soc.*, 2006, **26**(6), 849–859, DOI: 10.1016/j.jeurceramsoc.2004.12.031.

78 C. O. Kappe, High-speed combinatorial synthesis utilizing microwave irradiation, *Curr. Opin. Chem. Biol.*, 2002, **6**(3), 314–320, DOI: 10.1016/S1367-5931(02)00306-X.

79 A. Ludwig, Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods, *npj Comput. Mater.*, 2019, **5**(1), 70, DOI: 10.1038/s41524-019-0205-0.

80 F. Oviedo, Z. Ren, S. Sun, C. Settens, Z. Liu, N. T. P. Hartono, S. Ramasamy, B. L. DeCost, S. I. P. Tian, G. Romano, *et al.*, Fast and interpretable classification of small X-ray

diffraction datasets using data augmentation and deep neural networks, *npj Comput. Mater.*, 2019, **5**(1), 60, DOI: 10.1038/s41524-019-0196-x.

81 S. Sun, N. T. P. Hartono, Z. D. Ren, F. Oviedo, A. M. Buscemi, M. Layurova, D. X. Chen, T. Ogunfunmi, J. Thapa, S. Ramasamy, *et al.*, Accelerated Development of Perovskite-Inspired Materials via High-Throughput Synthesis and Machine-Learning Diagnosis, *Joule*, 2019, **3**(6), 1437–1451, DOI: 10.1016/j.joule.2019.05.014.

82 Y. Liu, K. P. Kelley, R. K. Vasudevan, M. Ziatdinov and S. V. Kalinin, *Machine Learning-driven Autonomous Microscopy for Materials and Physics Discovery*, Oxford University Press, US, 2023.

83 S. V. Kalinin, M. Ziatdinov, J. Hinkle, S. Jesse, A. Ghosh, K. P. Kelley, A. R. Lupini, B. G. Sumpter and R. K. Vasudevan, Automated and Autonomous Experiments in Electron and Scanning Probe Microscopy, *ACS Nano*, 2021, **15**(8), 12604–12627, DOI: 10.1021/acsnano.1c02104.

84 N. Creange, O. Dyck, R. K. Vasudevan, M. Ziatdinov and S. V. Kalinin, Towards automating structural discovery in scanning transmission electron microscopy, *Mach. Learn.: Sci. Technol.*, 2022, **3**(1), 015024.

85 S. V. Kalinin, Y. Liu, A. Biswas, G. Duscher, U. Pratiush, K. Roccapriore, M. Ziatdinov and R. Vasudevan, Human-in-the-Loop: The Future of Machine Learning in Automated Electron Microscopy, *Microsc. Today*, 2024, **32**(1), 35–41.

86 Y. Liu, K. P. Kelley, R. K. Vasudevan, H. Funakubo, S. S. Fields, T. Mimura, S. Trolier-McKinstry, J. F. Ihlefeld, M. Ziatdinov and S. V. Kalinin, Machine Learning-Driven Automated Scanning Probe Microscopy for Ferroelectrics, *Microsc. Microanal.*, 2022, **28**(S1), 2924–2926, DOI: 10.1017/s1431927622010972.

87 H. Joress, M. L. Green, I. Takeuchi and J. R. Hattrick-Simpers, Applications of High Throughput (Combinatorial) Methodologies to Electronic, Magnetic, Structural, and Energy-Related Materials, in *Encyclopedia of Materials: Metals and Alloys*, ed. F. G. Caballero, Elsevier, 2022, pp. 353–371.

88 D. Bash, Y. Cai, V. Chellappan, S. L. Wong, X. Yang, P. Kumar, J. D. Tan, A. Abutaha, J. J. W. Cheng, Y.-F. Lim, *et al.*, Multi-Fidelity High-Throughput Optimization of Electrical Conductivity in P3HT-CNT Composites, *Adv. Funct. Mater.*, 2021, **31**(36), 2102606, DOI: 10.1002/adfm.202102606.

89 J. R. Hattrick-Simpers, C. Jun, M. Murakami, A. Orozco, L. Knauss, R. J. Booth, E. W. Greve, S. E. Lofland, M. Wuttig and I. Takeuchi, High-throughput screening of magnetic properties of quenched metallic-alloy thin-film composition spreads, *Appl. Surf. Sci.*, 2007, **254**(3), 734–737, DOI: 10.1016/j.apsusc.2007.07.104.

90 K. Hippalgaonkar, Q. Li, X. Wang, J. W. Fisher III, J. Kirkpatrick and T. Buonassisi, Knowledge-integrated machine learning for materials: lessons from gameplaying and robotics, *Nat. Rev. Mater.*, 2023, **8**(4), 241–260.

91 R. Shimizu, S. Kobayashi, Y. Watanabe, Y. Ando and T. Hitosugi, Autonomous materials synthesis by machine learning and robotics, *APL Mater.*, 2020, **8**, 11.

92 S. Steiner, J. Wolf, S. Glatzel, A. Andreou, J. M. Granda, G. Keenan, T. Hinkley, G. Aragon-Camarasa, P. J. Kitson, D. Angelone and L. Cronin, Organic synthesis in a modular robotic system driven by a chemical programming language, *Science*, 2019, **363**(6423), eaav2211.

93 B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, *et al.*, A mobile robotic chemist, *Nature*, 2020, **583**(7815), 237–241, DOI: 10.1038/s41586-020-2442-2.

94 A. M. Lunt, H. Fakhruldeen, G. Pizzuto, L. Longley, A. White, N. Rankin, R. Clowes, B. M. Alston, A. I. Cooper and S. Y. Chong, Powder-Bot: A Modular Autonomous Multi-Robot Workflow for Powder X-Ray Diffraction, *Chem. Sci.*, 2024, **15**, 2456–2463.

95 P. Laveille, P. Miéville, S. Chatterjee, E. Clerc, J.-C. Cousty, F. de Nanteuil, E. Lam, E. Mariano, A. Ramirez and U. Randrianarisoa, Modular, multi-robot integration of laboratories: an autonomous workflow for solid-state chemistry, *Chimia*, 2023, **77**(3), 154–158.

96 B. P. MacLeod, F. G. Parlane, T. D. Morrissey, F. Häse, L. M. Roch, K. E. Dettelbach, R. Moreira, L. P. Yunker, M. B. Rooney and J. R. Deeth, Self-driving laboratory for accelerated discovery of thin-film materials, *Sci. Adv.*, 2020, **6**(20), eaaz8867.

97 D. Bash, F. H. Chenardy, Z. Ren, J. J. Cheng, T. Buonassisi, R. Oliveira, J. N. Kumar and K. Hippalgaonkar, Accelerated automated screening of viscous graphene suspensions with various surfactants for optimal electrical conductivity, *Digital Discovery*, 2022, **1**(2), 139–146, DOI: 10.1039/D1DD00008J.

98 R. W. Epps, M. S. Bowen, A. A. Volk, K. Abdel-Latif, S. Han, K. G. Reyes, A. Amassian and M. Abolhasani, Artificial chemist: an autonomous quantum dot synthesis bot, *Adv. Mater.*, 2020, **32**(30), 2001626.

99 D. Corzo, T. Wang, M. Gedda, E. Yengel, J. I. Khan, R. Li, M. R. Niazi, Z. Huang, T. Kim and D. Baran, A universal cosolvent evaporation strategy enables direct printing of perovskite single crystals for optoelectronic device applications, *Adv. Mater.*, 2022, **34**(9), 2109862.

100 K. Higgins, S. M. Valleti, M. Ziatdinov, S. V. Kalinin and M. Ahmadi, Chemical Robotics Enabled Exploration of Stability in Multicomponent Lead Halide Perovskites via Machine Learning, *ACS Energy Lett.*, 2020, **5**(11), 3426–3436, DOI: 10.1021/acsenergylett.0c01749.

101 C. J. Dolan, D. N. Cakan, R. E. Kumar, M. Kodur, J. R. Palmer, Y. Luo, B. Lai and D. P. Fenning, Scanning x-ray excited optical luminescence of heterogeneity in halide perovskite alloys, *J. Phys. D: Appl. Phys.*, 2022, **56**(3), 034002.

102 N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk and A. Merchant, An autonomous laboratory for the accelerated synthesis of novel materials, *Nature*, 2023, 1–6.

103 B. P. MacLeod, F. G. L. Parlane and C. P. Berlinguette, How to build an effective self-driving laboratory, *MRS Bull.*, 2023, **48**(2), 173–178, DOI: 10.1557/s43577-023-00476-w.
104 M. Seifrid, R. J. Hickman, A. Aguilar-Granda, C. Lavigne, J. Vestfrid, T. C. Wu, T. Gaudin, E. J. Hopkins and A. Aspuru-Guzik, Routescore: punching the ticket to more efficient materials development, *ACS Cent. Sci.*, 2022, **8**(1), 122–131.
105 S. Chatterjee, M. Guidi, P. H. Seeberger and K. Gilmore, Automated radial synthesis of organic molecules, *Nature*, 2020, **579**(7799), 379–384.
106 A. Mammoliti, P. Smirnov, M. Nakano, Z. Safikhani, C. Eeles, H. Seo, S. K. Nair, A. S. Mer, I. Smith, C. Ho, *et al.*, Orchestrating and sharing large multimodal data for transparent and reproducible research, *Nat. Commun.*, 2021, **12**(1), 5797, DOI: 10.1038/s41467-021-25974-w.
107 R. P. Wierenga, S. M. Golas, W. Ho, C. W. Coley and K. M. Esvelt, PyLabRobot: an open-source, hardware-agnostic interface for liquid-handling robots and accessories, *Device*, 2023, **1**(4), 100111, DOI: 10.1016/j.device.2023.100111.
108 S. P. Padhy and S. V. Kalinin, Domain hyper-languages bring robots together and enable the machine learning community, *Device*, 2023, **1**(4), 100115, DOI: 10.1016/j.device.2023.100115.
109 D. Guevarra, K. Kan, Y. Lai, R. J. Jones, L. Zhou, P. Donnelly, M. Richter, H. S. Stein and J. M. Gregoire, Orchestrating nimble experiments across interconnected labs, *Digital Discovery*, 2023, **2**(6), 1806–1812.
110 B. Chapman, J. Chilton, M. Heuer, A. Kartashov, D. Leehr, H. Ménager, M. Nedeljkovich, M. Scales, S. Soiland-Reyes and L. Stojanovic, *Common Workflow Language, v1.0*, Specification, Common Workflow Language working group, ed. P. Amstutz, M. R. Crusoe and N. Tijanić, 2016, https://w3id.org/cwl/v1.0/, DOI: 10.6084/m9.figshare.3115156.v2.
111 Intersect, https://www.ornl.gov/intersect, accessed, 2023, December 25.
112 Superfacility, https://www.nersc.gov/research-and-development/superfacility/, accessed, 2023, December 25.
113 Globus, https://www.globus.org/, accessed, 2023, December 25.
114 W. H. J. Manthorpe, *The Emerging Joint System of Systems: A Systems Engineering Challenge and Opportunity for APL*, Johns Hopkins APL Technical Digest, 1996, vol. 17(13).
115 M. W. Maier, Architecting principles for systems-of-systems, *Syst. Eng.*, 1998, **1**(4), 267–284, DOI: 10.1002/(SICI)1520-6858(1998)1:4<267::AID-SYS3>3.0.CO;2-D.
116 System of systems (SoS) considerations in life cycle stages of a system, in *ISO/IEC/IEEE*, ISO, 2019.
117 M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, *et al.*, The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*, 2016, **3**(1), 160018, DOI: 10.1038/sdata.2016.18.
118 L. C. Brinson, L. M. Bartolo, B. Blaiszik, D. Elbert, I. Foster, A. Strachan and P. W. Voorhees, Community action on FAIR data will fuel a revolution in materials research, *MRS Bull.*, 2023, 1–5.
119 *FAIR Data Principles*, GO FAIR, https://www.go-fair.org, accessed.
120 D. Morgan and R. Jacobs, Opportunities and Challenges for Machine Learning in Materials Science, *Annu. Rev. Mater. Res.*, 2020, **50**(1), 71–103, DOI: 10.1146/annurev-matsci-070218-010015.
121 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, Machine learning for molecular and materials science, *Nature*, 2018, **559**(7715), 547–555, DOI: 10.1038/s41586-018-0337-2.
122 S. V. Kalinin, O. Dyck, S. Jesse and M. Ziatdinov, Exploring order parameters and dynamic processes in disordered systems via variational autoencoders, *Sci. Adv.*, 2021, **7**(17), eabd5084, DOI: 10.1126/sciadv.abd5084.
123 S. Sanchez, Y. Liu, J. Yang, S. V. Kalinin, M. Ziatdinov and M. Ahmadi, Exploring the Evolution of Metal Halide Perovskites via Latent Representations of the Photoluminescent Spectra, *Adv. Intell. Syst.*, 2023, **5**(5), 2200340, DOI: 10.1002/aisy.202200340.
124 M. Krenn, Q. Ai, S. Barthel, N. Carson, A. Frei, N. C. Frey, P. Friederich, T. Gaudin, A. A. Gayle, K. M. Jablonka, *et al.*, SELFIES and the future of molecular string representations, *Patterns*, 2022, **3**(10), DOI: 10.1016/j.patter.2022.100588.
125 T. J. Smart, E. F. O'Bannon, M. R. Diamond, S. Stackhouse, B. K. Godwal, Q. Williams and R. Jeanloz, Equation of state, phase transitions, and band-gap closure in PbCl2 and SnCl2., *Phys. Rev. B*, 2023, **107**(13), 134113, DOI: 10.1103/PhysRevB.107.134113.
126 E. O. Pyzer-Knapp, J. W. Pitera, P. W. J. Staar, S. Takeda, T. Laino, D. P. Sanders, J. Sexton, J. R. Smith and A. Curioni, Accelerating materials discovery using artificial intelligence, high performance computing and robotics, *npj Comput. Mater.*, 2022, **8**(1), 84, DOI: 10.1038/s41524-022-00765-z.
127 G. Dulac-Arnold, N. Levine, D. J. Mankowitz, J. Li, C. Paduraru, S. Gowal and T. Hester, Challenges of real-world reinforcement learning: definitions, benchmarks and analysis, *Mach. Learn.*, 2021, **110**(9), 2419–2468.
128 G. Dulac-Arnold, N. Levine, D. J. Mankowitz, J. Li, C. Paduraru, S. Gowal and T. Hester, An empirical investigation of the challenges of real-world reinforcement learning, *arXiv*, 2020, preprint, arXiv:2003.11881, DOI: 10.48550/arXiv.2003.11881.
129 L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama and A. Ray, Training language models to follow instructions with human feedback, *Adv. Neural Inf. Process. Syst.*, 2022, **35**, 27730–27744.
130 ChatGPT, https://openai.com/blog/chatgpt/, accessed, 2023, 26 December.
131 Copilot, https://github.com/features/copilot, accessed, 2023, 26 December.
132 Tabnine, https://www.tabnine.com/, accessed, 2023, 26 December.

133 Gemini, **https://deepmind.google/technologies/gemini/#introduction**, accessed, 2023, 26 December.
134 D. A. Boiko, R. MacKnight and G. Gomes, Emergent autonomous scientific research capabilities of large language models, *arXiv*, 2023, preprint, arXiv:2304.05332, DOI: **10.48550/arXiv.2304.05332**.
135 A. Birhane, A. Kasirzadeh, D. Leslie and S. Wachter, Science in the age of large language models, *Nat. Rev. Phys.*, 2023, **5**(5), 277–280, DOI: **10.1038/s42254-023-00581-4**.
136 G. M. Hocky and A. D. White, Natural language processing models that automate programming will transform chemistry research and teaching, *Digital Discovery*, 2022, **1**(2), 79–83, DOI: **10.1039/D1DD00009H**.
137 A. D. White, G. M. Hocky, H. A. Gandhi, M. Ansari, S. Cox, G. P. Wellawatte, S. Sasmal, Z. Yang, K. Liu, Y. Singh, *et al.*, Assessment of chemistry knowledge in large language models that generate code, *Digital Discovery*, 2023, **2**(2), 368–376, DOI: **10.1039/D2DD00087C**.
138 M. A. Ziatdinov, Y. Liu, A. N. Morozovska, E. A. Eliseev, X. Zhang, I. Takeuchi and S. V. Kalinin, Hypothesis Learning in Automated Experiment: Application to Combinatorial Materials Libraries, *Adv. Mater.*, 2022, **34**(20), 2201345, DOI: **10.1002/adma.202201345**.
139 M. Ziatdinov, Y. Liu, K. Kelley, R. Vasudevan and S. V. Kalinin, Bayesian Active Learning for Scanning Probe Microscopy: From Gaussian Processes to Hypothesis Learning, *ACS Nano*, 2022, **16**(9), 13492–13512, DOI: **10.1021/acsnano.2c05303**.
140 R. K. Vasudevan, E. Orozco and S. V. Kalinin, Discovering mechanisms for materials microstructure optimization via reinforcement learning of a generative model, *Mach. Learn.: Sci. Technol.*, 2022, **3**(4), 04LT03.
141 K. L. Snapp and K. A. Brown, Driving school for self-driving labs, *Digital Discovery*, 2023, **2**(5), 1620–1629, DOI: **10.1039/D3DD00150D**.
142 ChatGPT, **https://chat.openai.com/share/e09963ff-9f5b-4ab2-8adb-15ddcabec007**, accessed, 2024, March 11.
143 B. P. MacLeod, F. G. L. Parlane, A. K. Brown, J. E. Hein and C. P. Berlinguette, Flexible automation accelerates materials discovery, *Nat. Mater.*, 2022, **21**(7), 722–726, DOI: **10.1038/s41563-021-01156-3**.
144 B. P. MacLeod, F. G. L. Parlane, T. D. Morrissey, F. Häse, L. M. Roch, K. E. Dettelbach, R. Moreira, L. P. E. Yunker, M. B. Rooney, J. R. Deeth, *et al.*, Self-driving laboratory for accelerated discovery of thin-film materials, *Sci. Adv.*, 2020, **6**(20), eaaz8867, DOI: **10.1126/sciadv.aaz8867**.