Materials Horizons

ROYAL SOCIETY OF CHEMISTRY

REVIEW

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Mater. Horiz.*, 2026, **13**, 4712

Toward self-driving laboratory 2.0 for chemistry and materials discovery

Heeseung Lee,†[ab] Hyuk Jun Yoo,†[a] Hye Su Jang,[ab] Byeongho Park,[a] Yang Jeong Park*[a] and Sang Soo Han*[a]

The convergence of laboratory automation, artificial intelligence (AI), and data-driven science has catalyzed the emergence of self-driving laboratories (SDLs), autonomous platforms capable of designing, executing, and analyzing experiments with minimal human input. While early SDLs (SDL 1.0) demonstrated the feasibility of closed-loop discovery, their impact has been constrained by limited scope, poor interoperability, and reliance on human-curated heuristics. This review outlines the vision of SDL 2.0: a new generation of flexible, scalable, and collaborative discovery engines for chemistry and materials science. We discuss recent advances in modular hardware design, AI-driven decision-making including Bayesian optimization, computer vision, and large language models, and orchestration software that integrate scheduling, data management, and safety protocols. Building on these foundations, we propose six defining characteristics for SDL 2.0: interoperable, collaborative, generalizable, orchestrated, safe, and creative. Together, these features establish SDLs as globally networked platforms, enabling reproducible experimentation, accelerated innovation, and democratized access to advanced research infrastructure. By embedding modularity, AI reasoning, and community-driven standards into their core, SDLs 2.0 promise to transform not only how experiments are conducted, but also who can participate in and benefit from the accelerating pace of scientific discovery.

Received 19th October 2025,
Accepted 3rd March 2026

DOI: 10.1039/d5mh01984b

rsc.li/materials-horizons

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Wider impact

This review synthesizes key developments in self-driving laboratories (SDLs), highlighting the emergence of SDLs that integrate modular automation, robotics, Bayesian optimization, AI-driven computer vision, large language models, and laboratory operating systems to enable fully autonomous, closed-loop experimentation. These advances have demonstrated substantial gains in efficiency, reproducibility, and safety across chemistry and materials science, establishing SDLs as a transformative research paradigm. The broader significance of this field lies in its ability to democratize access to advanced experimental capabilities through interoperable and cloud-connected platforms, enabling global collaboration and accelerating discovery beyond the limits of traditional, human-centered experimentation. Looking forward, the field is poised to transition toward SDL 2.0—intelligent, interoperable, and collaborative ecosystems that seamlessly integrate synthesis, characterization, and theory across laboratories. The conceptual framework articulated in this review provides a roadmap for this evolution, guiding the development of generalizable and orchestrated SDL platforms that will reshape materials science and chemistry by fostering human-AI collaboration, accelerating knowledge generation, and enabling globally networked, data-driven discovery.

1. Introduction

The pursuit of new chemical compounds and functional materials has historically been constrained by the inherently slow and resource-intensive nature of traditional experimentation.[1,2] Conventional laboratory research relies heavily on iterative cycles of human intuition, hypothesis-driven experimentation, and trial-and-error optimization. While this paradigm has enabled major scientific and technological breakthrough, the pace of discovery has often lagged behind the rapidly growing demand for advanced materials and sustainable chemical processes. Accelerating discovery requires a fundamental transformation of how laboratories operate. In response, the integration of laboratory automation, data science, and artificial intelligence (AI) has given rise to the innovative concept of the self-driving laboratory (SDL).[3–28] By coupling robotic

[a] *Computational Science Research Center, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea. E-mail: sangsoo@kist.re.kr*
[b] *Department of Materials Science and Engineering, Korea University, Seoul 02841, Republic of Korea*
† These authors contributed equally.

This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

experiments with intelligent decision-making algorithms, SDLs are envisioned as autonomous platforms that can design, execute, and analyze experiments with minimal human intervention. Early exemplars demonstrated that the feasibility of closed-loop experimentation in various fields from organic synthesis to thin-film deposition and catalysis.[3–7,10] These systems showed the potential to reduce the experimental search space dramatically, optimize material properties, and uncover unexpected phenomena that may elude traditional approaches. However, these first-generation SDLs—what we may term self-driving laboratory 1.0—remain limited in scope. They often operate within narrow experimental domains, rely on handcrafted features or human-curated heuristics, and lack the interoperability needed to generalize across diverse chemical and materials systems.

An SDL is typically composed of three key elements: automated instruments (Automated Hardware), AI,[9–18,29–33] and an orchestration software (OS)[34–39] that orchestrates and manages them. AI is responsible for hypothesis generation and experimental design, automation executes these tasks with precision, and the OS coordinates the entire process while accumulating and managing data. Crucially, these components should not operate independently but rather function as a unified system that creates synergy and accelerates scientific discovery. To achieve this vision, SDLs must systematically advance along three interdependent axes: the number of experiments per unit time (n_{exp}),[40] the information gained per experiment (I),[40] and the efficiency of converting data into scientific knowledge (η).[40] Maximizing n_{exp} requires not only high-throughput hardware but also OS-level resource management strategies such as optimizing robotic arm trajectories, coordinating parallel tasks, and minimizes idle time between experiments. Enhancing I depends on systematic collection, standardization, and FAIR-compliant management[41–43] of diverse data streams, enabling multimodal AI analysis, noise reduction, and feature extraction. Improving η necessitates advanced AI techniques—including Bayesian optimization,[9,15,20,29,30,33] reinforcement learning,[18,44,45] and foundation models—that establish self-improving loops through iterative cycles of experimentation and feedback.

Although n_{exp}, I, and η may appear to be independent metrics, in practice they are tightly interconnected. Improvements in hardware throughput enable richer datasets, which in turn support more efficient AI-driven decision-making. Conversely, intelligent data management and AI strategies maximize hardware utilization and further increase experimental efficiency. Considering these three elements in an integrated manner is therefore essential not only for optimizing individual laboratories but also for building distributed SDL networks. These principles mark the transition towards what we describe as self-driving laboratory 2.0. Unlike their predecessors, SDL 2.0 platforms aspire to function as flexible discovery engines capable of transitioning between different classes of chemical

Heeseung Lee

Mr Heeseung Lee received his BS degree in Materials Science & Engineering from Pukyong National University, South Korea. He is currently pursuing an integrated MS/PhD program in Materials Science and Engineering at Korea University. His research focuses on self-driving laboratories for accelerated materials discovery, encompassing laboratory automation, data-driven materials design, and the application of large language models to scientific workflows.

Hyuk Jun Yoo

Dr Hyuk Jun Yoo is a postdoctoral researcher in autonomous laboratories at Lawrence Berkeley National Laboratory (LBNL) and University of California, Berkeley under the guidance of Prof. Gerbrand Ceder. He earned his PhD in Chemical and Biomolecular Engineering from Korea University in collaboration with KIST, where he built autonomous systems for noble metal nanoparticle synthesis. His research interests include lab automation, AI for science, orchestration software for self-driving lab and adaptive in situ XRD experimental design for discovering hidden intermediate states.

Sang Soo Han

Dr Sang Soo Han received his PhD from the Korea Advanced Institute of Science and Technology (KAIST) in 2005. After postdoc training at California Institute of Technology, he worked as a senior research scientist at the Korea Research Institute of Standards and Science. In June 2013, he joined the Korea Institute of Science and Technology (KIST), where he has been serving as the Head of the Computational Science Research Center since September 2020. Dr Han specializes in multiscale simulations for materials design. His recent research interests have expanded to data-driven materials discovery based on AI and self-driving laboratories.

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

reactions, material systems, and experimental modalities with minimal reconfiguration. They harness modular hardware, advanced AI reasoning, and interoperable systems to create scalable infrastructures that bridges computation, automation, and theory.

In this review, we will discuss (i) modularization strategies for hardware design, (ii) computer vision as both a decision-making aid and the eyes of robotic systems, (iii) large language models (LLMs) as cognitive partners for AI-driven reasoning, and (iv) orchestration software that orchestrate hardware and AI while handling scheduling and data management. Finally, the Perspective section will address how these technologies can be optimized in an integrated fashion and outline the future directions for SDL 2.0.

2. Modular hardware strategies for SDLs: balancing flexibility and functionality

Selecting an appropriate hardware development strategy is a fundamental step in designing an SDL. The chosen approach can differ substantially depending on research objectives, budget constraints, and the targeted degree of automation.[48–51] Nevertheless, the central question is whether the hardware design can achieve both operational efficiency and long-term sustainability, which in turn requires a modular architecture.[48] Rather than merely assembling individual instruments, SDL hardware must form an organically interconnected structure that is adaptable to evolving experimental demands and scalable with technological progress. In this context, modularity becomes a defining principle of SDL design.

The core challenge, however, is to balance functionality with flexibility.[51] A robust modularization strategy enables this balance by treating each experimental component as an independent, interchangeable unit. For example, robotic arms, sample-handling devices, and analytical instruments can all function as discrete modules that may be added, replaced, or upgraded with minimal disruption. Crucially, an interoperable environment that integrates such modules seamlessly allows the SDL to maximize its technological potential. Modularity should therefore be understood not only at the device level, but as a system-wide design philosophy that underpins both performance and sustainability. To ensure practical functionality, modular hardware must meet four key attributes, collectively described by the RAST framework: reusability (the ability to repeatedly utilize hardware), Adaptability (responsiveness to evolving research needs), Scalability (the ease of expanding functions or incorporating new experimental types), and Tunability (the precision to control experimental parameters). Only when these attributes are realized does modularity evolve from a conceptual guideline to the operational foundation of SDL hardware.

The specific way in which modularity is defined and implemented strongly influences an SDL's flexibility, scalability and adaptability. Broadly, two primary strategies can be distinguished: unit-based compact systems[4–8,10,12,13,18,22,28,52] and station-based distributed systems.[19,24,27,46,47,53] In unit-based compact systems such as AlphaFlow,[18] Chemputer,[6] and Ada,[5,10] modules generally refer to internal components (*e.g.*, pumps, valves, or heating units) designed for straightforward replacement. By contrast, in station-based platforms such as A-Lab[27] or MARK,[20,36,46] a "module" typically denotes an entire equipment station dedicated to a specific task. Here, robotic systems transport samples between stations, enabling modularity at the workflow level. In summary, unit-based systems excel in ease of control and component replaceability but face scalability challenges when experimental instruments must be integrated. Station-based systems, meanwhile, offer greater workflow flexibility and scalability but require sophisticated software coordination, standardized communication protocols, and precise robotic control to maintain reliability.[54,55]

2.1. Unit-based modular systems

A unit-based modular system[4–8,10,12,13,18,22,28,52] integrates core laboratory functions—such as sample injection, reaction, analysis, and cleaning—into compact modular units within a single device (Fig. 1a–d). Direct connectivity between modules minimizes the need for robotic arms, streamlining both installation and operation. These systems are especially well-suited for specialized domains such as flow chemistry, thin-film fabrication, and polymer studies. Their key strengths lie in the integration of core functions and reconfigurability through standardized interfaces, which allow modules (*e.g.*, photoreactors, chillers, or online analyzers) to be easily added, exchanged, or combined. This reconfigurability supports rapid prototyping, lowers the entry barriers to automation, and enhances flexibility at the laboratory scale.

For instance, AlphaFlow[18] integrates modular units for reagent injection, reaction, cleaning, measurement, and separation into a microfluidic-based platform (Fig. 1a). Similarly, Coley *et al.*[6] developed a robotic-arm-based flow platform that incorporates reagent reservoirs, pumps, automated valve blocks, interchangeable reactor cartridges, cleaning loops, and robotic transport (Fig. 1b). The Ada platform,[5] optimized for thin-film synthesis, integrates a spin-coater, plate heater, and analytical tools around a robotic handler to achieve high precision (Fig. 1c). Rooney *et al.*[56] also developed a unit-based system for adhesive research.

Despite these advantages, unit-based systems are usually tailored to specific applications, which limits their general interoperability, scalability in physical space, and adaptability to diverse experimental workflows. Nevertheless, they remain highly powerful hardware strategies in environments where reproducibility and rapid iteration are critical, serving as foundational entry points into laboratory automation.

2.2. Station-based distributed systems

Station-based distributed systems[19,24,27,46,47,53] distributed key experimental functions—such as synthesis, pretreatment, and analysis—across multiple specialized stations (Fig. 1e–h). Each

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

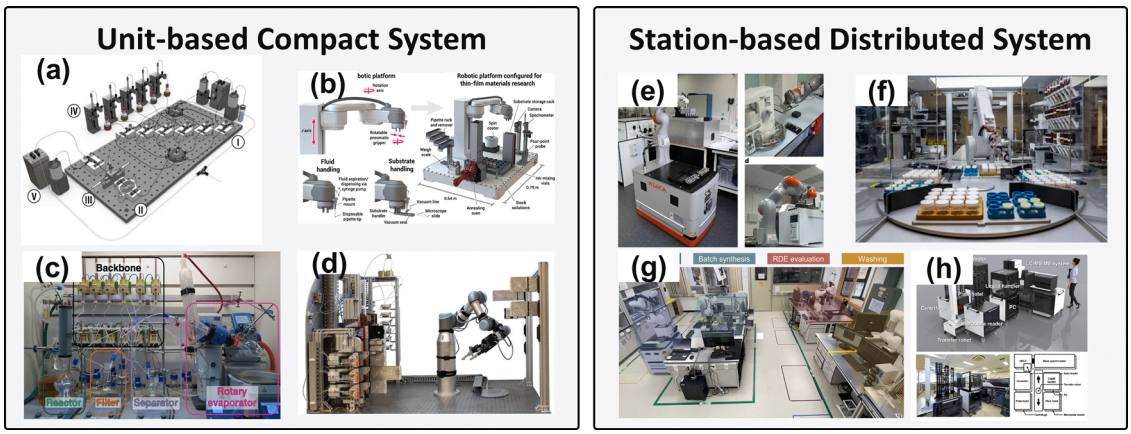

Fig. 1 (a)–(d) Unit-based compact systems where modular components are integrated into a single compact platform for automated synthesis and analysis. (a) Reproduced from ref. 18. Copyright 2023, The Authors. (b) Reproduced with permission. Ref. 5 Copyright 2020, American Association for the Advancement of Science. (c) Reproduced with permission. Ref. 6 Copyright 2019, American Association for the Advancement of Science. (d) Reproduced with permission. Ref. 4 Copyright 2019, American Association for the Advancement of Science. (e)–(h) Station-based distributed systems, where independent robotic stations are interconnected *via* mobile robots or conveyor systems. (e) Reproduced with permission. Ref. 24 Copyright 2020, Springer Nature. (f) Reproduced from ref. 27. Copyright 2023, Springer Nature. (g) Reproduced from ref. 46. Copyright 2024, Wiley-VCH. (h) Reproduced with permission. Ref. 47 Copyright 2025, Springer Nature.

station operates independently as a module, while robotic handlers (*e.g.*, robotic arms or mobile robots) transport samples between them. Although stations are physically separate, integrated software coordinates workflow to ensure automation operates as a coherent whole. This architecture is particularly advantageous for complex experimental pipelines that require diverse processing and analytical instruments, such as solid-state synthesis,[27] nanoparticle catalyst development,[46] and biotechnology research.[47]

The strength of distributed platforms lies in their scalability and adaptability. New instruments can be readily integrated, and existing workflows can be expanded or parallelized to increase throughput. Robotic handlers provide additional versatility by accommodating diverse equipment layouts and experimental conditions. Examples include the platform developed by Szymanski *et al.*[27] (Fig. 1f), which automates solid-state synthesis by linking independent modules such as an automated dispenser, a tube furnace, and an X-ray diffractometer. Burger *et al.*[24] (Fig. 1e) designed a rack-based platform where a robotic arm sequentially execute dispensing, reactions, analysis. The MARK system (Fig. 1g),[46] developed by KIST, combines nanoparticle synthesis, preprocessing, and an electrochemical measurement stations connected by a mobile robot to achieve both throughput and flexibility. Fushimi *et al.*[47] (Fig. 1h) created a biotechnology-oriented SDL, where modules for cultivation, pretreatment, and analysis are mounted on movable carts, allowing spatial reconfiguration as needed.

In short, station-based distributed systems excel at handling complex workflows and scaling with new equipment, but they demand advanced robotics, robust error-handling, and integrated software infrastructures.[48,57] These requirements raise both technical complexity and cost, which can limit accessibility.

3. Recent advances in AI for SDLs

AI has rapidly become the central driving force behind the development of SDLs. Unlike traditional automation systems, which primarily emphasize high-throughput experimentation, AI enables SDLs to strategically design experiments, dynamically control complex processes, and interpret outcomes to generate new knowledge. In this sense, AI functions not only as the operational engine but also a cognitive partner in autonomous scientific discovery. Broadly, its role can be understood across three complementary domains: experimental design AI, vision AI, and language AI.

First, machine learning, deep learning, and reinforcement learning have emerged as key enabler in materials science and chemistry, driving accelerated discovery and optimization of material properties[1,2,7,9,18,19,21,25,29,32,44–49,52,53,56,58–61] Among these, Bayesian optimization[5,12,20,29,30,33,52,56,61–63] has gained particular prominence as a strategy for efficiently identifying optimal compositions and discovering materials with targeted functionalities. Second, computer vision[64–73] systems provide SDLs with real-time situational awareness. By monitoring and quantifying experimental conditions, vision-based AI ensures that robotic platforms can execute precise manipulations and safely operate in dynamic, unpredictable laboratory environments. Third, large language models (LLMs)[60,74–80] are beginning to serve as cognitive interfaces between humans and machines. By translating natural language into experimental actions, generating hypotheses, and facilitating interdisciplinary collaboration, LLMs position themselves as essential partners for scientific reasoning and communicating within SDLs. Taken together, these three AI paradigms transform SDLs from rigid, pre-programmed automation pipelines into adaptive, reasoning-driven scientific agents. The following sections

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

highlight recent advances and representative applications in each domain.

3.1. Expanding the frontiers of autonomous discovery: recent advances in Bayesian frameworks

The core of SDLs lies not only in automated hardware but also in algorithms that guide experimental decision-making.[3–10,12,13,81] A range of approaches has been proposed—such as evolutionary algorithms,[82] particle swarm optimization,[83] genetic algorithms,[84] and reinforcement learning[85]—but Bayesian optimization (BO)[33] has emerged as the most widely adopted. BO has gained attention because it can identify optimal conditions for maximizing target properties within vast design spaces while requiring relatively few experiments.[20,24,86–88] Numerous studies have demonstrated the ability of Bayesian optimization to efficiently explore high-dimensional parameter spaces.[9,89–95] The principle of BO lies in approximating the experimental space with a probabilistic surrogate model, then using an acquisition function to determine the next experiment.[96] In chemistry and materials science, however, experimental search spaces are often constrained by chemical feasibility, synthesis limitations, or measurement costs.[97,98] This making it necessary to adapt BO frameworks so that surrogate models not only predict outcomes but also embed domain knowledge, thereby reducing wasted efforts and steering the search toward viable solutions.[99–103]

A comprehensive overview of BO methodologies and applications in chemical and materials research has been provided by Aspuru-Guzik and colleagues in their recent *Chemical Reviews* article.[48] Building on this foundation, the present review emphasizes how BO is being integrated into the SDL paradigm and how it has evolved in recent years. Recent advances focus on several directions: efficiently navigating high-dimensional design spaces,[9,89–94] employing multi-objective optimization[10,104–106] to balance competing experimental goals, and extending beyond parameter optimization to include procedural aspects of experimental protocols (*i.e.*, process constraints).[29,97] These developments are particularly important for the complexity of real-world chemical and materials workflows.[9,20,24,90–93] Thus, this review discusses extending BO frameworks that have been proposed to address such requirements in the Section 3.1.1. Moreover, purely statistical BO approaches are often insufficient in chemistry and materials sciences.[62,95,107] To enhance both efficiency and scientific validity, there is a growing emphasis on domain-informed BO, which integrates chemical and physical principles directly into the optimization framework,[99–103,107–111] which is discussed in detail in the Section 3.1.2.

3.1.1. Extending the BO framework to complex problems. Beyond its fundamental formulation, recent advances in BO have shifted toward expanding the framework itself to accommodate more complex and realistic problem settings.[112–115] These extensions do not merely refine BO's accuracy within its traditional scope but instead broadens the spectrum of optimization tasks it can address.[61,116,117] While classical BO was largely confined to single-objective optimization over continuous variables, recent approaches now handle multi-objective problems, categorical and ordinal variables, and optimization under practical constraints. Such developments significantly expand the applicability of BO across scientific and engineering domains.

For instance, Low *et al.*[112] introduced the EGBO (Evolution-Guided Bayesian Optimization) framework, which combines evolutionary algorithms with multi-objective BO (q-NEHVI) to efficiently identify Pareto fronts while avoiding infeasible regions (Fig. 2a).

Scientific discovery in materials science typically requires the simultaneous optimization of continuous process parameters (*e.g.*, temperature) and categorical variables (*e.g.*, the selection of solvents, catalysts, or synthesis sequences), which cannot be targeted efficiently with standard continuous methods. To address this challenge, Häse *et al.* developed GRYFFIN,[115] which extends Bayesian optimization (BO) to categorical variables by leveraging physicochemical descriptors to define inter-category similarity. This descriptor-based approach enables the robust and efficient optimization of mixed continuous–categorical spaces. Similarly, addressing the complexity of synthesis sequences in nanoparticle discovery, Yoo *et al.* developed NanoChef[113] (Fig. 2b). This method extends BO to categorical variables—specifically reagent injection orders—by embedding sequential information using positional encoding and capturing chemical semantics through MatBERT,[118] thereby enabling the simultaneous optimization of synthesis protocols and continuous parameters. At the same time, efforts have been made to render the BO process itself more adaptive and intelligent. Mottafegh *et al.*[61] introduced a meta-optimization strategy in which multiple surrogate models are run in parallel, with the best-performing model selected in real time to guide experiments. Regarding adaptive representations, while the dynamic formulation of GRYFFIN allows for the refinement of descriptor weights to capture relevant trends, Rajabi-Kochi *et al.* proposed FABO[117] (Feature Adaptive Bayesian Optimization). Distinct from the reweighting approach, FABO dynamically selects explicit feature subsets at each iteration to mitigate the curse of dimensionality, particularly in data-scarce regimes.

Furthermore, Cissé *et al.*[114] developed BORA (Fig. 2c), a hybrid framework in which a LLM intervenes when BO becomes trapped in a local minimum. By analyzing accumulated data and generating new hypotheses, the LLM augments the mathematical efficiency of BO with broader reasoning capacities—offering a powerful step toward cognitively enhanced optimization.

3.1.2. Domain-informed BO in chemistry and materials science. BO relies on probabilistic modeling and has demonstrated potential in chemistry and materials science.[62,86,95] However, when applied in a purely data-driven manner, its dependence on statistical correlations introduces inherent limitations. For example, in alloy composition optimization problems, where variables (x^1, x^2, x^3, …, x^n) are treated independently, BO may generate infeasible compositions during

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

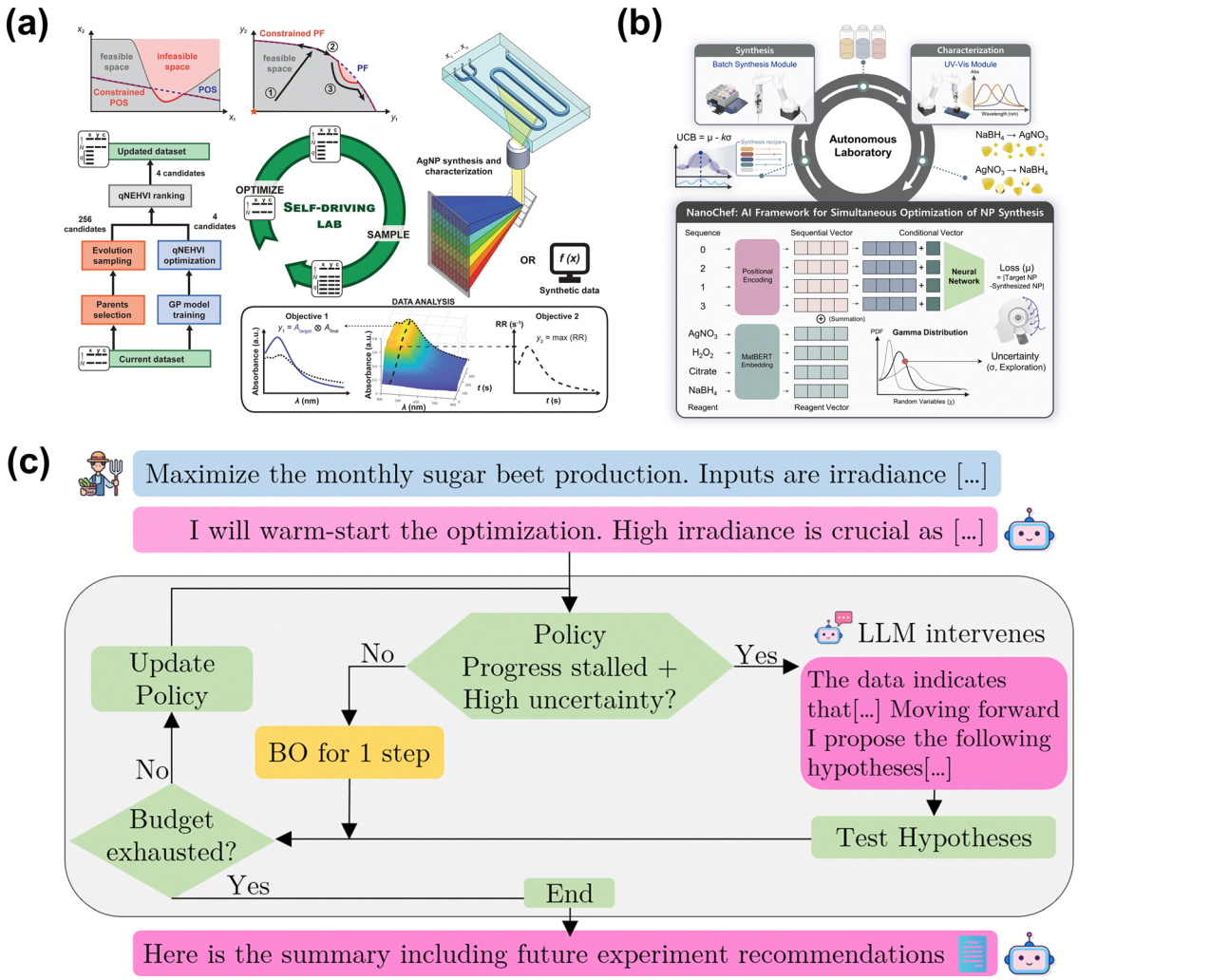

Fig. 2 Representative extensions of BO frameworks. (a) EGBO combining evolutionary search with multi-objective BO. Reproduced with permission. Ref. 112 Copyright 2024, Springer Nature. (b) NanoChef enabling simultaneous optimization over categorical and continuous variables for nanoparticle synthesis. Reproduced with permission from the authors. Ref. 113 (c) BORA coupling BO with LLM reasoning to escape local minima. Reproduced with permission from the authors. Ref. 114.

global exploration. Such suggestions reduce physical consistency, leading to inefficiencies in both sampling and the balance between exploration and exploitation. To overcome these challenges, domain-informed BO integrates physical and chemical knowledge into the probabilistic framework, steering the search toward valid candidates and enabling more efficient experimental design.[99–103,107–111]

Several strategies have been developed to achieve this integration. The first involves incorporation of prior models that embed physical or chemical knowledge into the surrogate model.[103,111] For instance, in Ni-Ti alloy optimization, a thermodynamic phase transformation model was used as the GP prior mean to enhance the physical realism of property predictions.[100] In Fig. 3a, PAL 2.0[111] employed physics-informed descriptors selected by XGBoost,[119] trained a neural-network surrogate on them, and embedded the result into the GP prior mean, thereby achieving stable optimization even with limited data. Similarly, the ChIDDO framework (Fig. 3b)[103] combined kinetic rate laws and physics-based models with data-driven surrogates to improve sampling efficiency.

The second approach focuses on encoding physical structure into the kernel function or embedding knowledge of the material/chemical search space.[99–102,107,120] Conformational BO[107] provides a representative example, where the torsional potential energy surface of molecules was explicitly integrated into a potential-informed kernel combined with a periodic kernel. This allowed exploration of molecular rotational degrees of freedom in a physically consistent manner.

A third approach is multi-fidelity BO (MFBO), which integrates data of varying fidelity.[17,101,108,121,122] Physics-aware MFBO, for instance, introduces a physics-based bias into the multi-fidelity acquisition function, enabling searches guided not only by fidelity correlations but also by explicit physical knowledge.[101] Building on this concept, Sabanza-Gil *et al.*[108] systematically tested synthetic and real chemical/material

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

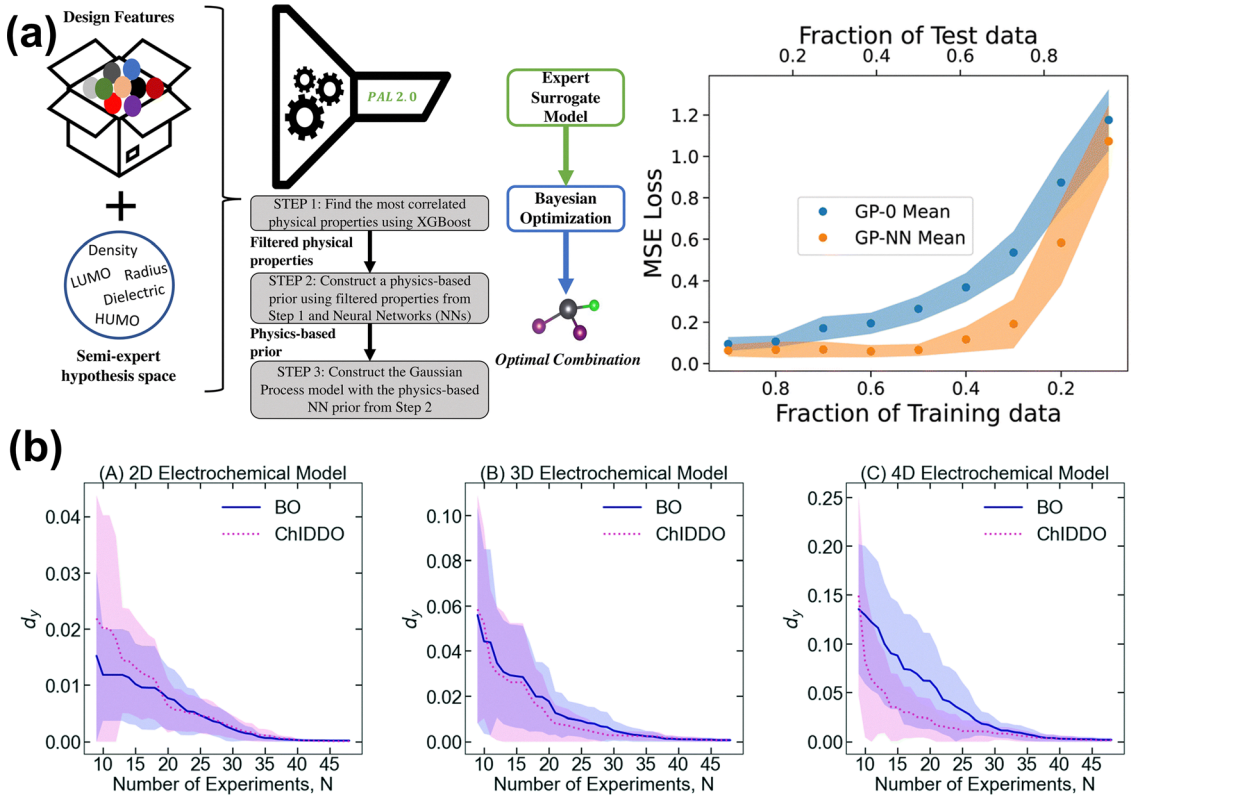

Fig. 3 Representative examples of domain-informed Bayesian optimization. (a) Workflow of the PAL 2.0 framework, where physical descriptors are selected *via* XGBoost, used to construct a physics-based prior mean through a neural network, and incorporated into a Gaussian process surrogate for Bayesian optimization. The right panel compares predictive accuracy (MSE loss) of the GP-NN prior model against a standard zero-mean GP, showing superior performance of the physics-informed surrogate with limited training data. Reproduced from ref. 111 with permission from the Royal Society of Chemistry. (b) Comparison of standard BO and chemically informed BO (ChIDDO) for electrochemical models of increasing dimensionality (2D, 3D, and 4D). In all cases, ChIDDO achieves faster convergence and lower error, highlighting the efficiency gains from embedding chemical and physical knowledge into the optimization framework. Reproduced from ref. 103 with permission from the Royal Society of Chemistry.

systems, identifying both success and failure modes of MFBO. Their work established practical guidelines, showing that MFBO is effective only when the low-fidelity source is both inexpensive ($\rho < 0.1$) and highly informative ($R^2 > 0.8$).

The fourth approach incorporates constraints and gray-box modeling to explicitly enforce feasibility and physical laws.[99,109,123] In process simulations, for example, surrogate models augmented with mass and energy balances preserved physical consistency.[109] Similarly, DKIBO[99] applied acquisition function corrections to embed domain constraints, suppress unnecessary exploration, and improve optimization efficiency.

Finally, physically meaningful descriptors represent an important means of embedding domain knowledge.[102,115] In optimizing high-entropy alloy nanozymes, descriptors derived from density functional theory and molecular dynamics—such as Gibbs free energy changes (ΔG) and d-band centers—were integrated into the surrogate model.[102] By anchoring the search on physics-based features, this strategy moved beyond statistical correlations and enabled optimization guided by chemical and physical insight. Notably, this study also demonstrated the fusion of computational data with experimental data generated by a SDL, highlighting the potential of integrated simulation-experiment workflows to accelerate computational–experimental co-discovery.

3.2. Vision AI-assisted experiment execution

In SDLs, the role of AI extends beyond experimental design to ensuring the reliable and safe execution of experiments.[48,124] Chemical processes often take place in dynamically changing environments, where unpredictable changes—such as phase separation, fluctuations in liquid levels, or accidental spillage—pose significant challenges for robotic systems relying solely on motion planning. To overcome these challenges, computer vision has emerged as a key enabling technology. By emulating human visual cognition, vision systems provide real-time situational awareness and enable adaptive responses to evolving experimental conditions.[64–73] Specifically, computer vision serves three complementary functions within SDLs: (1) monitoring and quantitative analysis of experimental progress, (2) safety assurance through anomaly detection and hazard prevention, and (3) robotic assistance for complex manipulations requiring high precision and adaptability.

3.2.1. Applications of computer vision for experimental monitoring and analysis. In automating chemical and

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

materials experiments, computer vision serves as a powerful tool for monitoring and quantitatively analyzing experimental progress.[66,72,125–130] Traditionally, researchers relied on subjective visual cues—such as color change, turbidity, bubble formation, or precipitation—to infer reaction dynamics. Computer vision replaces this manual observation with a ''digital eye'' capable of extracting quantitative information that would otherwise require specialized analytical instruments such as turbidimeters, optical microscopes, or fluorescence spectrometers.[131,132] By leveraging high-resolution imaging[133] and advanced image-processing algorithms,[134] vision-based systems can quantify changes in color, transparency, and precipitate formation, thereby reducing dependence on specialized instrumentation and establishing a scalable, cost-effective monitoring framework.

Representative applications highlight the breadth of this approach. HeinSight[66] detects liquid–liquid phase separation, interfacial boundaries, and turbidity in real time to autonomously regulate extraction and purification workflows. Unlike simple edge-detection methods, HeinSight reliably identified aqueous–organic phase boundaries under diverse conditions (Fig. 4a), while also monitoring homogeneity, solid formation, and residual impurities enhancing the efficiency of automated separation and purification workflows. Kineticolor,[72] developed by Barrington *et al.* (Fig. 4b), tracks reaction progress by quantifying color changes in the CIE–$L^*a^*b^*$ color space, enabling the simultaneous monitoring of multiple reactions from a single video feed. This represents a shift from the conventional ''one-video-one-reaction'' paradigm toward scalable, multiplexed visual analytics. Similarly, Li *et al.*[73] integrated computer vision into a high-throughput robotic colorimetric titration platform, combining an open-source liquid-handling robot (OT-2) with vision-based image analysis (Fig. 4c). By continuously recording subtle color transitions to automate tasks such as pH measurement, complexometric titrations, and redox titrations (Fig. 4c). In this system, image capture was employed to continuously record the subtle color transitions such as the faint pink endpoint in $KMnO_4$ titrations and converting them into reproducible numerical values through segmentation and CIELab color analysis, the system automated titrations spanning acid–base, redox, and complexometric chemistries.

Despite these advances, vision-based monitoring is not without limitations. Accuracy can be sensitive to lightning conditions, sample opacity, and variations across camera hardware.[73] Overcoming these challenges requires the development of standardized datasets, careful control of imaging environments, and improved algorithms capable of robust performance under diverse experimental conditions. Looking forward, the establishment of community standards for imaging device parameters and experimental protocols will be crucial to ensure reproducibility, comparability, and scalability of computer vision in SDLs.

3.2.2. Computer vision for safety in SDLs. Computer vision can serve as a critical component of laboratory safety in SDLs,[124] particularly in unmanned experimental environments where potential hazards such as gas leaks, contamination, and explosions must be carefully managed.[69,135–137] While sensor-based safety systems are already indispensable, vision-assisted monitoring provides an additional layer of protection by enabling the early detection and prevention of accidents. Examples include solution ejection during the handling of strong reducing agents (*e.g.*, $NaBH_4$), or liquid spills caused by robotic manipulation errors involving glassware such as vials, Falcon tubes, flasks. To address these challenges, vision technologies tailored to the unique characteristics of chemical experiments are increasingly being developed.

A key difficulty arises from the widespread use of transparent vessels in laboratory practice. To overcome this, Wang *et al.*[69] proposed the MVTrans architecture, an end-to-end multi-view framework incorporating depth estimation, segmentation, and pose estimation for transparent object recognition (Fig. 5a). By enabling real-time detection of overturned glassware, MVTrans allows robotic systems to take immediate corrective actions and prevent secondary accidents. Similarly, Tiong *et al.*[135] developed a DenseSSD-based transparent object detection model capable of tracking vessels in real time and issuing immediate alerts upon detecting anomalies (Fig. 5b).

Beyond glassware monitoring, computer vision has also been applied to improve the safety of automated liquid handling. Khan *et al.*[136] integrated a YOLOv8-based vision system with the Opentrons OT-2 robot to detect missing pipette tips and incorrect liquid levels, enabling real-time feedback, task interruption, and reattempts (Fig. 5c). While primarily aimed at enhancing automation reliability, such mechanisms also mitigate risks associated with mishandling and contamination. Because most chemical experiments involve liquid-phase processes, spill detection is another critical safety concern. Eppel *et al.*[65] developed segmentation approaches capable of recognizing diverse laboratory materials including liquids and solids inside transparent containers, while also identifying their boundaries and phases (Fig. 5d). To support this, they constructed the Vector-LabPics, a dataset of over 2000 annotated laboratory images encompassing glassware, liquid states, and phase separation phenomena. This resource has become foundational for advancing computer vision systems that ensure both reliability and safety in complex chemical laboratory environments.

3.2.3. Computer vision for robotic manipulation. Computer vision provides real-time visual feedback that enables robotic arms to perform sophisticated experimental tasks typically carried out by human experimenters.[64,67,70,71,139] While simple and repetitive operations such as pipetting can be automated with straightforward motion planning, more delicate manipulations like solution pouring require advanced perception and control. Yoshikawa *et al.*[138,140] demonstrated stable liquid pouring by integrating constrained motion planning with sensor-based closed-loop control, showing that robots can replicate the delicate manipulations essential to laboratory practice (Fig. 6a). Similarly, simulation-based pipelines for liquid perception have been developed to improve

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

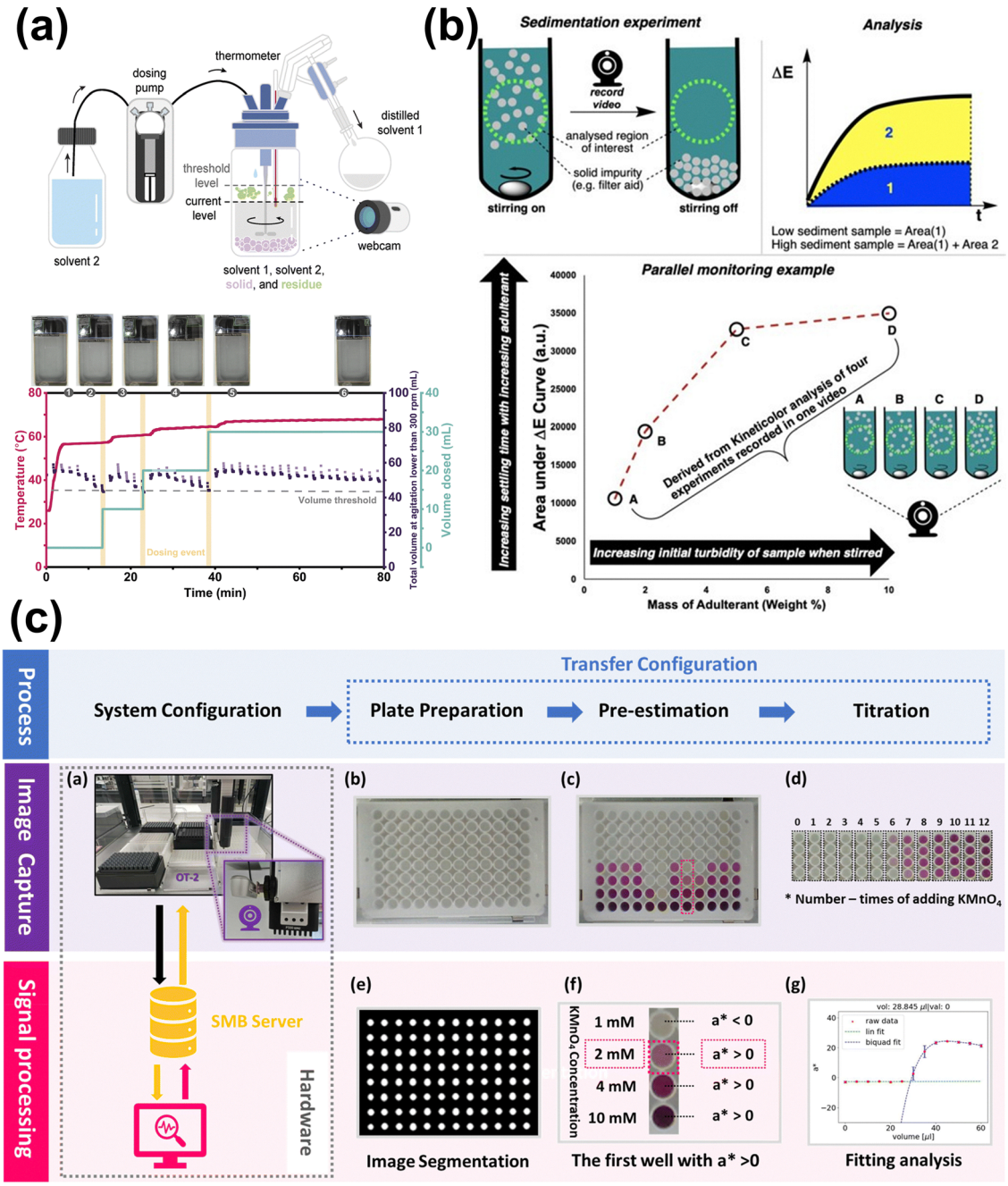

Fig. 4 (a) Real-time monitoring of solvent exchange distillation using webcam-based setup for reaction slurry, enabling automated dosing upon threshold detection. Reproduced from ref. 66 with permission from the Royal Society of Chemistry. (b) Parallel monitoring of sedimentation experiments through multi-region Kineticolor analysis, allowing simultaneous evaluation of multiple samples. Reproduced from ref. 72 with permission from Wiley-VCH. (c) High-throughput robotic colorimetric titration workflow for H_2O_2 determination, integrating automated plate handling, image segmentation, and fitting analysis. Reproduced from ref. 73 with permission from the Royal Society of Chemistry.

robotic handling of fluidic tasks under dynamic conditions (Fig. 6b).[71] Beyond liquid handling, vision-guided systems have been applied to extend automation to complex equipment. Lee *et al.*[67] developed a framework enabling robotic arms to insert and retrieve samples from within the confined geometry of a centrifuge, a task that requires precise coordination and spatial awareness (Fig. 6c). Recently, multimodal approaches have been introduced, such as the incorporation of audio–visual feedback for robust powder grinding, which enhances reliability in tasks involving solid-phase materials (Fig. 6d).[70] Collectively, these studies highlight how computer vision transforms robotic systems from simple executors of repetitive tasks into versatile agents capable of navigating complex, constrained, and multimodal laboratory environments. Such advances underscore the critical role of vision AI in pushing the boundaries of experimental automation within SDLs.

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

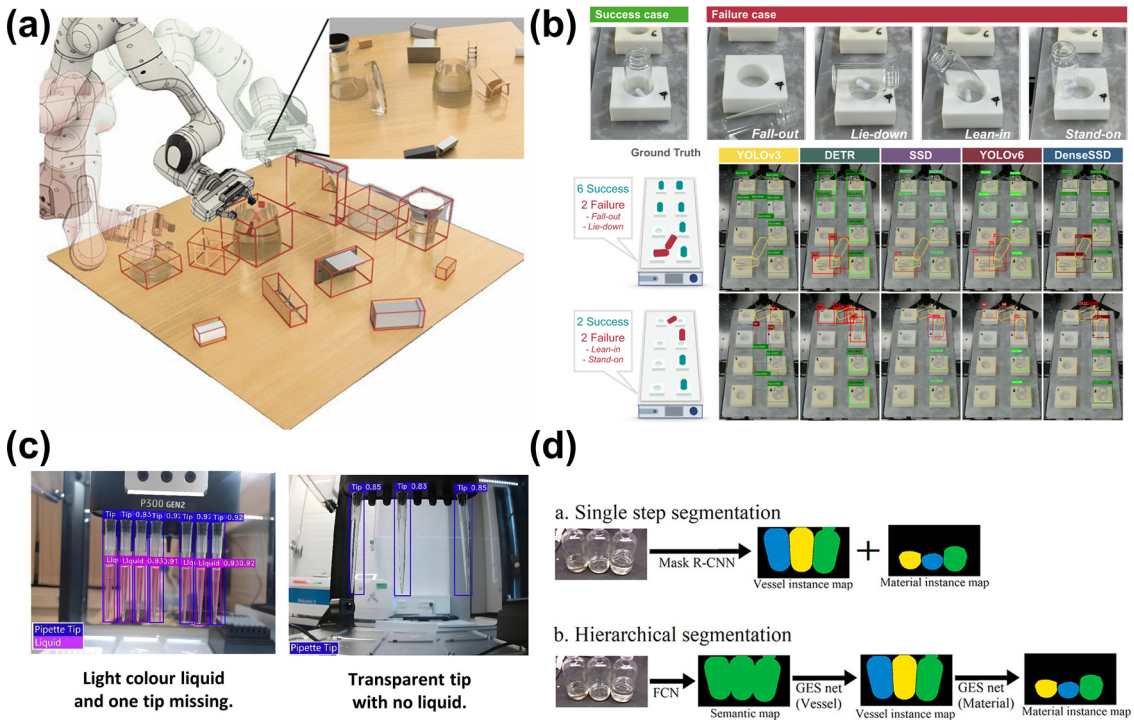

Fig. 5 Computer vision approaches for safety in SDLs. (a) Multi-view perception of transparent and opaque objects using MVTrans, an end-to-end multi-task perception network that predicts segmentation masks, depth maps, poses, and 3D bounding boxes. Reproduced from ref. [69] with permission from the authors. arXiv:2302.11683. (b) Machine vision-based detection of transparent chemical vessels to ensure safe vial positioning during automated synthesis, comparing multiple detection models across success and failure cases. Reproduced from ref. 135 with permission from Springer Nature. (c) Real-time AI-driven quality control for the OT-2 liquid handling robot, enabling detection of missing pipette tips and assessment of liquid presence under diverse laboratory conditions. Reproduced from ref. 136 with permission from Springer Nature. (d) Recognition of materials and vessels in chemistry lab settings using segmentation approaches. Reproduced from ref. 65 with permission from the American Chemical Society.

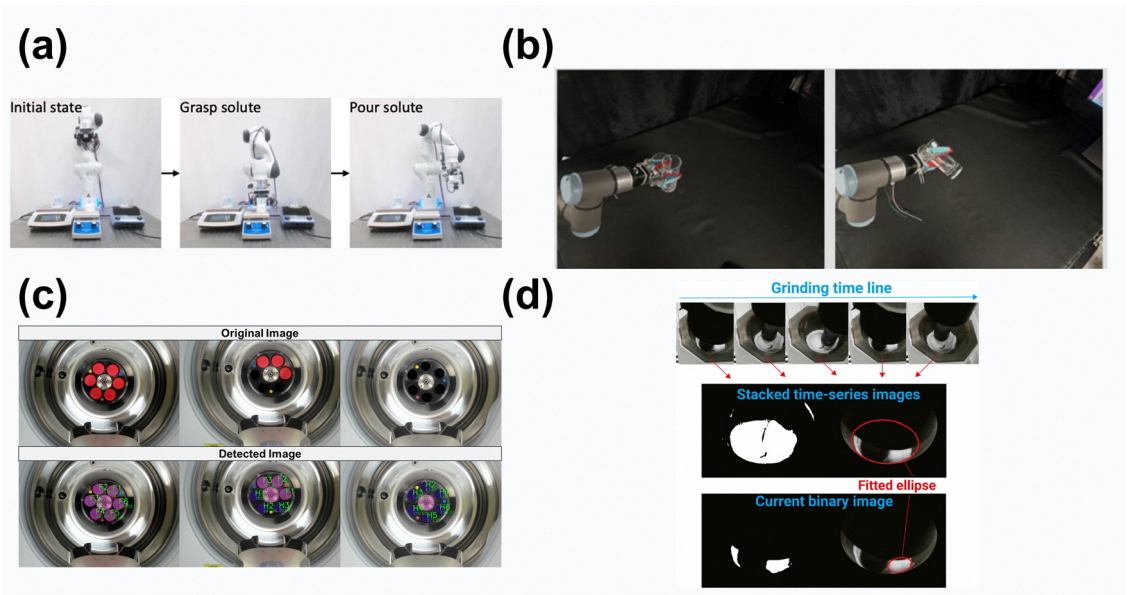

Fig. 6 Examples of assisting robotic manipulation in SDLs enabled by computer vision. (a) and (b) Task and motion planning for robotic liquid handling and pouring. (a) Reproduced from ref. 138 with permission from the authors. arXiv:2212.09672. (b) reproduced from ref. 71 with permission from Wiley-VCH. (c) Automated nanoparticle washing process guided by vision with robotic handling of centrifuge samples. Reproduced from ref. 67 with permission from the authors. Chemrxiv-2025-px33t-v2 (d) Robotic powder grinding with audio-visual feedback for real-time monitoring. Reproduced from ref. 70 with permission from IEEE.

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

3.3. LLM: a cognitive partner for scientific discovery

LLMs,[74–78,114,140–143] trained on vast corpora of scientific knowledge spanning materials science, chemistry, and biology, are emerging as a new class of cognitive partners for researchers. Within the SDL paradigm, their role is rapidly expanding beyond knowledge retrieval to that of central agents of scientific discovery.[78,144–150] Unlike traditional approaches, where years of coding expertise or specialized software training were prerequisites, LLMs provide a universal natural language interface that enables direct interaction with experimental and computational systems.[145,150,151] This capacity lowers the barrier to entry, fosters interdisciplinary collaboration, and allows even non-experts to leverage SDL infrastructures.[152,153]

The functionality of LLMs in this context is increasingly enhanced through complementary strategies. Retrieval-augmented generation (RAG)[154–156] improves factual reliability by grounding responses in curated databases; domain-specific fine-tuning injects specialized disciplinary knowledge;[155,157–160] and agentic LLM frameworks orchestrate multiple LLMs[161–165] or connect them with external computational and experimental tools. These strategies collectively extend the role of LLMs from hypothesis generation to active laboratory execution. In this sense, LLMs are no longer static language models but are evolving into adaptive, agentic systems capable of reasoning, planning, and acting within SDL environments.

This advances are already being realized in practice. LLMs can automatically extract synthesis conditions and property data from the literature (knowledge mining),[157,160,166–173] leveraging this information to propose new experiments and hypotheses (experimental planning),[145,174] and directly interface with robotic platforms to execute tasks and interpret outcomes (execution & interpretation).[68,78,159,161,175] Taken together, these applications position LLMs as powerful partners of discovery, bridging human intent and autonomous experimentation.

3.3.1. Knowledge mining & experimental planning. Knowledge mining has long been regarded as a traditional starting point of natural language processing (NLP)[167–173] and a fundamental basis for scientific inquiry. In materials science and chemistry, it has evolved into a paradigm for extracting experimental conditions, property data, and synthesis pathways from the vast scientific literature. These efforts have enabled the construction of databases, the development of knowledge graphs, and the formulation of new hypotheses, thereby accelerating discoveries across disciplines. A representative example is the A-Lab system,[27,176] which transformed text-mined synthesis conditions into machine learning–based recommendation models for proposing initial experimental conditions, demonstrating how literature-derived data can be integrated into SDL workflows (Fig. 7a). However, conventional NLP-based approaches have struggled to capture the diversity of chemical expressions and their structural context, limiting their adaptability to novel compounds or experimental scenarios.

LLMs offer a step change by capturing contextual meaning and enabling more flexible and comprehensive knowledge extraction. Recent work have explored fine-tuning LLMs[158–160] for chemistry-specific text mining tasks, significantly improving accuracy. Zhang *et al.*[160] fine-tuned LLMs for five distinct tasks: compound entity recognition, reaction role labeling, metal–organic framework (MOF) synthesis information extraction, nuclear magnetic resonance (NMR) data extraction, and the transformation of synthesis procedures into action sequences (Fig. 7b), demonstrating superior performance compared to earlier NLP models (Fig. 7b). Beyond direct extraction,

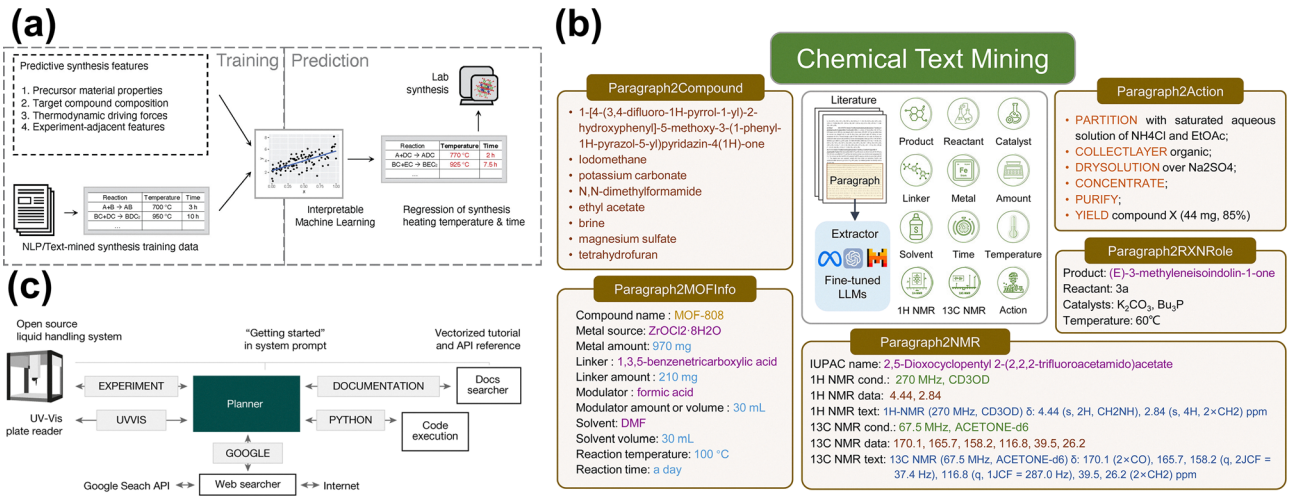

Fig. 7 Evolution of knowledge extraction approaches in autonomous chemistry. (a) Machine-learning (ML) rationalization and prediction of solid-state synthesis conditions using features mined from chemical text. Reproduced from ref. 176 with permission from the American Chemical Society. (b) Fine-tuned LLMs applied to chemical text mining, enabling extraction of compounds, actions, material properties, and spectroscopic information. Reproduced from ref. 160 with permission from the Royal Society of Chemistry. (c) Autonomous research workflows driven by LLMs, integrating search tools, experimental documentation, and code execution for end-to-end scientific reasoning. Reproduced from ref. 146 with permission from Springer Nature. Together, these examples illustrate the transition from conventional ML-based models to advanced LLM-powered frameworks for scientific knowledge extraction.

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

LLMs now complement missing knowledge through retrieval-augmented generation (RAG), which grounds responses in external sources such as Wikipedia, Google Scholar, and arXiv and domain-specific databases. Systems such as ChemCrow[145] and Coscientist[146] (Fig. 7c) exemplify this approach, combining literature extraction with web-based retrieval to extend

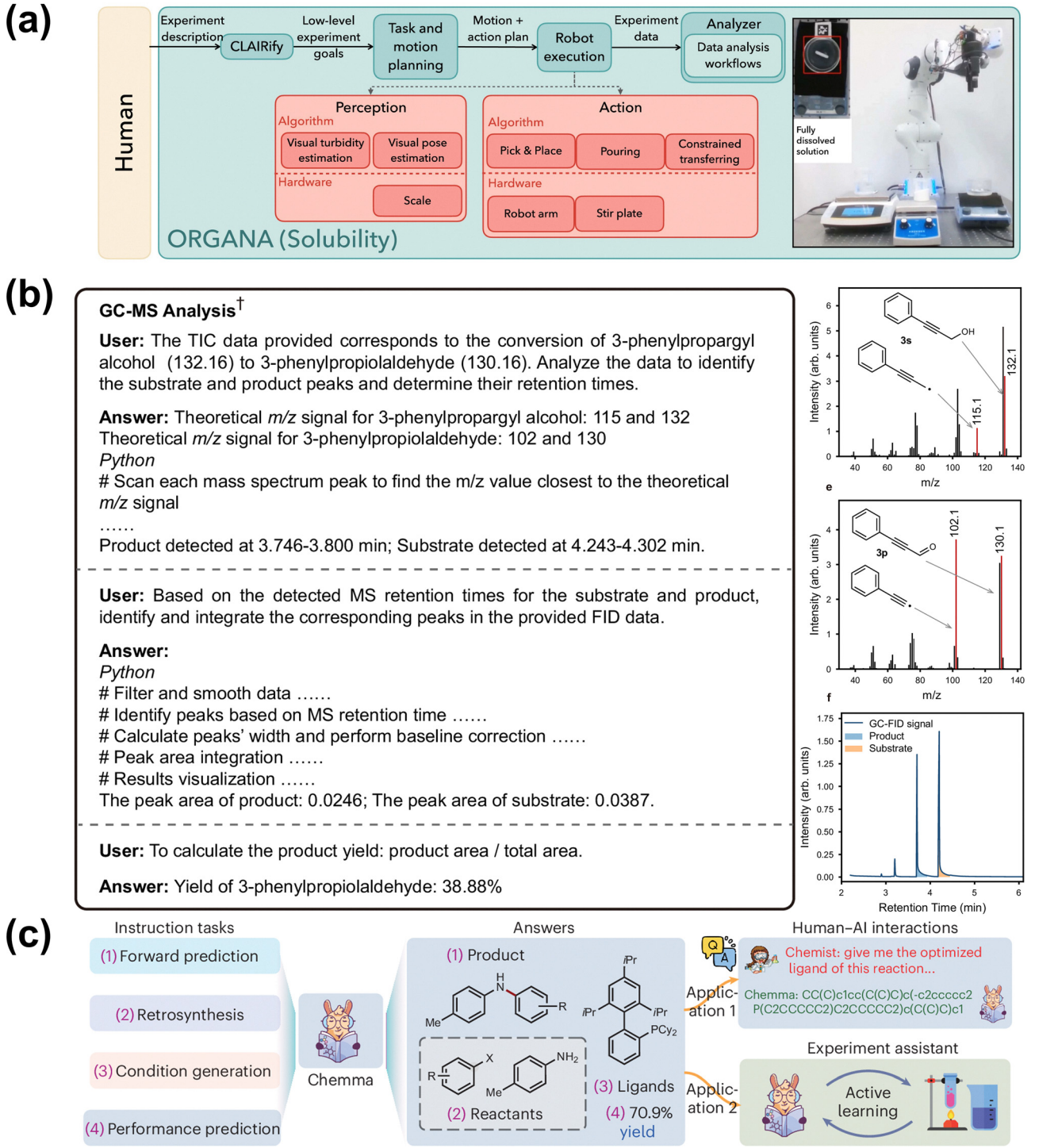

Fig. 8 (a) ORGANA, a robotic assistant that translates natural language experiment descriptions into executable laboratory workflows, enabling perception, planning, and robotic execution. Reproduced from ref. 68 with permission from Elsevier. (b) End-to-end synthesis development platform powered by LLMs, where the model interprets GC–MS data, identifies substrates and products, and quantifies yields for autonomous chemical analysis. Reproduced from ref. 161 with permission from Springer Nature. (c) LLMs support organic chemistry through forward prediction, retrosynthesis, and condition generation, while interacting with human chemists to guide experimental design and optimization. Reproduced from ref. 159 with permission from Springer Nature.

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

capabilities from knowledge mining to experimental design and execution. In this way, LLMs support strategic exploration by proposing experimental conditions hypotheses, reducing reliance on trial-and-error experimentation. Building further on this trajectory, the agentic mixture-of-workflows framework[78,79] positions retrieval not as an auxiliary step but as a central component of decision-making. By integrating multiple retrieval workflows including text retrieval, database querying, and structure-based search, it enables LLMs to operate as multi-agent RAG systems, dynamically selecting the most appropriate workflow for a given context.

3.3.2. Experimental execution & interpretation. As discussed in the previous section, LLMs effectively reduce knowledge barriers in research by enabling seamless extraction of domain-specific insights from external sources. Their utility extends beyond knowledge mining to include the control of experimental equipment and the interpretation of results within SDL environments.[78,146,161] Traditionally, researchers have interacted with instruments through direct programming or specialized interfaces, but LLMs can now translate natural language instructions into executable control commands, greatly improving accessibility to complex automation infrastructures.[68,152,161,175,177]

Notable examples include ORGANA[68] and the work of Inagaki *et al.*,[175] both of which demonstrate that LLMs can directly link natural language goals to experimental execution. ORGANA enables researchers to specify objectives interactively, interpret them into instrument control tasks, and subsequently report and explain results, thereby realizing a conversational laboratory environment reminiscent of cinematic depictions (Fig. 8a). Similarly, Inagaki *et al.*[175] showed that high-level instruction such as "replace the cell culture medium" could be translated by GPT-4 into executable robot control code, achieving automated implementation. Such systems remove the need for complex coding, enabling intuitive interactions with SDLs for both experts and non-specialists. The LLM-RDF (Reaction Development Framework) introduced by Ruan *et al.*[161] extends beyond mere experimental execution by incorporating agents such as spectrum analyzers and result interpreters to automatically analyze and interpret experimental data in Fig. 9b. This shift toward data-centered automation represents a new dimension of human–AI interaction anchored in experimental results. Likewise, Zhang *et al.*[159] recently developed *Chemma*, a chemistry-focused LLM designed to support condition recommendation, yield prediction, and ligand exploration in organic synthesis (Fig. 8c). Through iterative collaboration with human researchers, *Chemma* successfully optimized a new α-aryl N-heterocycle synthesis condition within only 15 experimental runs.

Collectively, these studies not only lower the barriers to adopting SDL approaches but also expand opportunities for participating in materials development to non-specialists as well as experts. Nevertheless, safeguards remain essential to prevent errors arising from insufficient domain knowledge during interaction, and systematic validation procedures are required to ensure that LLMs accurately analyze and interpret experimental results.

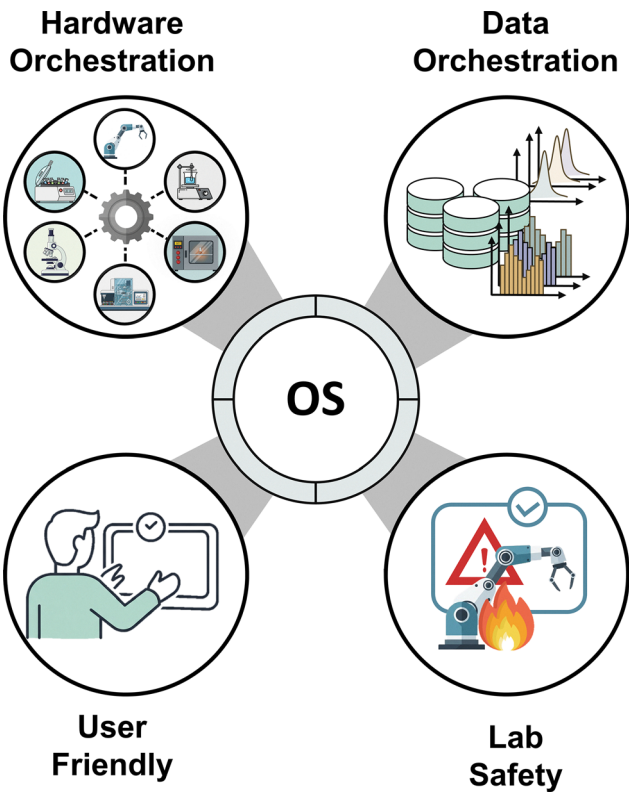

Fig. 9 The OS coordinates (i) hardware orchestration for integrating diverse robotic modules, (ii) data orchestration for managing and analyzing experimental information, (iii) lab safety mechanisms to ensure reliable and secure autonomous operation, and (iv) user-friendly interfaces to facilitate accessibility and control.

4. Orchestration software in SDLs

The OS of an SDL functions as a core infrastructure that goes beyond a simple software layer, integrating instruments, data, users, and safety into a unified framework.[34–39,178–184] Just as a conventional computer OS efficiently manages CPUs, memory, and I/O resources, the SDL OS interconnects and optimizes both physical experimental resources and digital assets. Its design can be understood around four fundamental principles: hardware orchestration, data management, safety, and user-friendliness as shown Fig. 9.

4.1. Hardware orchestration

Hardware orchestration in SDLs refers to the integrated management of heterogeneous equipment such as robotic arms, pumps, sensors, and analytical instruments through standardized control interfaces (APIs). This interoperability not only reduces conflicts and inefficiencies *via* resource scheduling but also ensures that SDLs, with their inherently modular architecture, can flexibly accommodate new instruments or software modules in a plug-and-play manner.[39,55,183–186] Different SDL orchestration software embody distinct orchestration

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

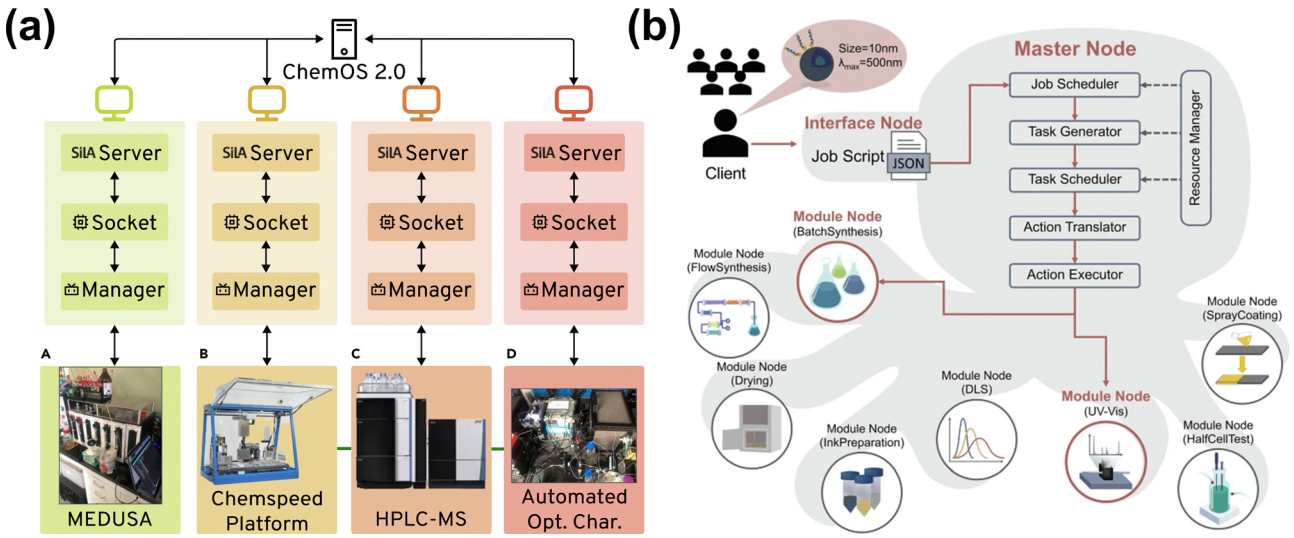

Fig. 10 (a) ChemOS 2.0 integrates diverse experimental modules using the SiLA2 protocol and a central fog computing device. Reproduced from ref. 39 with permission from Elsevier. (b) OCTOPUS coordinates distributed modules through TCP/IP communication with a master–module node framework. Reproduced from ref. 36 with permission from Springer Nature.

philosophies. For example, ChemOS 2.0[39] utilizes a fog-computing-based[187] orchestration layer—analogous to an operating system kernel—to centrally coordinate module interaction and data management in Fig. 10a. By incorporating the SiLA2[55] protocol within this architecture, it achieves robust system integration and high reproducibility.

Divergence becomes particularly evident in resource conflict management. AlabOS[35] and OCTOPUS (Fig. 10b)[36] both employ TCP/IP-based communication to manage geographically distributed instruments, yet their strategies differ fundamentally. AlabOS employs a manager–worker architecture with a centralized resource reservation system, requiring tasks to explicitly secure ownership of all necessary resources prior to execution, thereby preventing conflicts at their source. By contrast, OCTOPUS[36] focuses on throughput maximization in multi-user environments, relying on an intelligent scheduler that dynamically allocates idle time slots to parallelize jobs and enhance efficiency. Beyond these approaches, some systems emphasize specific technology stacks tailored to real laboratory conditions. For example, ARChemist[183] employs the robot operating system (ROS) as its communication backbone, enabling seamless collaboration between fixed robotic arms and mobile robots within a unified experimental pipeline.

4.2. Data management

Data management is one of the core functions of SDL orchestration software, encompassing the full pipeline of collecting, storing, standardizing, analyzing, and sharing experimental data.[36,39,178,188] The value of such data is maximized not only when the final results are recorded but also when rich metadata, such as executed commands, parameters, and environmental conditions, is included. High data reproducibility depends critically on this metadata. Accordingly, modern OS designs have moved beyond simple logging to fully embrace the FAIR principles (Findable, Accessible, Interoperable, Reusable),[41,43] providing systematic frameworks that ensure reproducibility.

Several representative implementations highlight different strategies. HELAO[178] applies an extended FAIR+ principle by recording all commands, parameters, and metadata in HDF5[189] format and automatically uploading them to institutional repositories after experiments. It further supports real-time data streaming *via* FastAPI WebSockets, enabling external observers to visually monitor ongoing experiments as they progress. ChemOS 2.0[39] introduces another layer of novelty by separating experimental and simulation data, managing them through PostgreSQL[190] and AiiDA,[191] respectively, to ensure both integrity and domain-specific organization. A central fog computing device functions as the data hub, orchestrating collection, storage, and exchange across the laboratory.

XDL,[6] though not an OS, standardizes experimental protocols as fully digital data objects (XDL files), allowing complete reproducibility of procedures in code. Combined with the ChemTorrent[192] peer-to-peer sharing model, laboratories can execute and verify protocols before redistributing them, establishing decentralized trust and reproducibility. Similarly, AlabOS[35] leverages MongoDB[193] (NoSQL) as a backend to support flexible data schemas and real-time tracking of metadata, including sample positions. Other systems highlight complementary approaches: ARChemist[183] implements a persistence manager to record the complete history of each sample, while OCTOPUS,[36] designed for multi-user environments, synchronizes instrument states in real time and standardizes all tasks in JSON scripts for consistent management.

In summary, data management within SDL OSs is evolving from simple record-keeping toward comprehensive provenance tracking, reproducibility assurance, and intelligent data utilization as a central resource for AI-driven decision-making. While

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

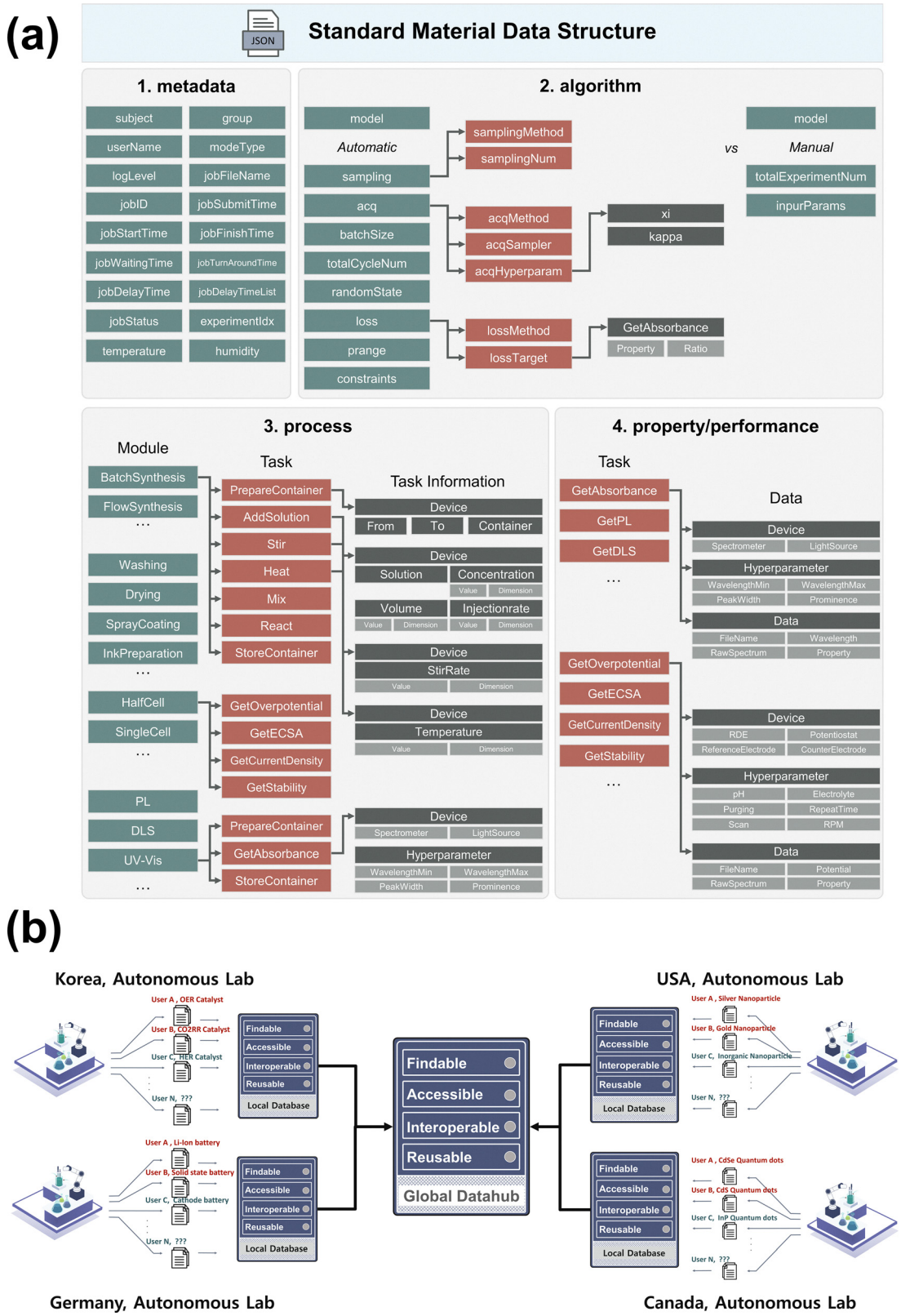

Fig. 11 (a) Standard material data structure implemented in OCTOPUS, organizing metadata, algorithms, processes, and property/performance data for reproducible autonomous experimentation. Reproduced from ref. 36 with permission from Springer Nature. (b) Conceptual illustration of an international data hub, where standardized formats and FAIR-compliant databases from distributed SDLs enable global interoperability and collaborative acceleration of materials discovery.

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

current systems illustrate how experimental data and metadata can be organized within individual laboratories, the most critical step forward is the establishment of a standard material data structure as exemplified in Fig. 11a. Such standardization will enable interoperability across diverse platforms and foster the creation of global data hubs (Fig. 11b), where SDLs worldwide can collaborate seamlessly. Ultimately, this shared infrastructure will accelerate the growth of the AI ecosystem, allowing collective intelligence to drive faster, more reliable scientific discovery.

4.3. Safety

Safety management must be a top priority for SDL orchestration software, encompassing not only the protection of instruments but also the assurance of experimental integrity and reliability.[124] Ideally, these systems should continuously monitor environmental parameters such as pressure, temperature, and electrical load, and respond proactively by halting experiments, triggering alarms, or switching to safe modes when anomalies occur. Moreover, safety management should extend beyond local monitoring to incorporate external frameworks such as chemical safety regulations and biosafety standards, thereby ensuring both operational stability and regulatory compliance. However, most existing OSs still place greater emphasis on preventing hardware or software conflicts than on holistic environmental or regulatory safety.

Several representative systems illustrate different safety strategies. HELAO[178] ensures operational safety through sequential execution, where each task must finish before the next begins. Its distributed architecture enhances stability by isolating failures—if one server crashes, the entire system remains operational. ChemOS 2.0,[39] leveraging the SiLA2[55] protocol, implements detailed exception handling and automatic recovery routines that restore only the affected step without restarting the entire workflow. AlabOS[35] emphasizes conflict-free resource allocation: tasks must explicitly reserve all required resources prior to execution, and automatic deallocation after completion ensures fundamental avoidance of hardware conflicts. OCTOPUS[36] employs a different approach for multi-user environments, introducing a masking table and intelligent scheduler to dynamically resolve complex conflicts in multi-user environments, even invoking task rescheduling or emergency stops when risks are detected. In a different vein, AR-Chemist[183] incorporates a ROS-based rule manager that preemptively detects and mitigates spatial conflicts in workflows involving heterogeneous robots (*e.g.*, robotic arms and mobile robots).

In summary, safety management in SDL OS is evolving from static error prevention and collision avoidance toward dynamic risk prediction and real-time validation of experiment integrity. This shift reflects a broader view of safety, extending beyond hardware protection to encompass data reliability, regulatory compliance, and scientific reproducibility, ultimately framing safety as a comprehensive approach to risk management in SDLs.

4.4. User friendliness

User-friendliness of an SDL OS determines how intuitively researchers can interact with the system. Beyond hardware automation, an effective OS should support workflow design, experiment monitoring, and error handling through comprehensive UI/UX frameworks.[34,37,39,178,183] To fully unlock the potential of autonomous experimentation, interfaces must be accessible even to scientists without programming expertise. Since many materials scientists and chemists neither develop SDLs themselves nor have extensive coding backgrounds, direct interaction with SDLs can otherwise be a barrier. Thus, graphical, web-based, and drag-and-drop interfaces are critical for lowering the barrier and enabling widespread adoption across the scientific community.

HELAO[178] uses FastAPI for automatic API documentation, enabling developers to quickly add devices or actions, while its web-based operator/observer interface provides real-time control and monitoring. ChemOS2.0[39] adopts a streamlit-based GUI, allowing researchers to easily submit experiments and track progress, thereby supporting a gradual transition toward automation. IvoryOS[37] goes a step further by automatically generating GUIs from existing Python code without modification; workflows can be built *via* drag-and-drop and further enhanced with natural language input powered by LLMs, dramatically reducing the coding barrier (Fig. 12). OCTOPUS,[36] designed for multi-user environments, relies on a CLI-centered interface but integrates a GPT-powered Copilot to assist in code generation and workflow customization. Other systems highlight domain-specific accessibility. For example, NIMS-OS[34] provides a GUI tailored for candidate-space exploration, allowing researchers to visually navigate parameter settings.

In summary, user-friendliness in SDL OSs is a relative concept shaped by both target users (developers *vs.* domain scientists) and research contexts (single-user *vs.* multi-user environments). Systems such as HELAO, ChemOS and IvoryOS emphasize web-based interfaces for broad accessibility, while NIMS-OS and ARChemist tailor GUIs to specific research needs. Together, these approaches highlight a spectrum of strategies aimed at ensuring SDLs are not confined to expert programmers but can be adopted by the wider scientific community.

5. Roadmap for SDL 2.0

Although SDLs have made remarkable progress in recent years, three is still room to expand their capabilities. These platforms can already automate complex workflows, yet achieving broader interoperability, scalability, and collaboration across distributed systems remains a work in progress.[50,194–197] In many cases, SDLs operate *in silos*, with experiment data, metadata, and procedures stored in formats customized for hardware setup of each laboratory.[27,34–38,178,198] This diversity makes it harder to share, interpret, or reproduce results across different environments.[192] Even when experimental procedures are available, small differences in instrumentation (*e.g.*, heating

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

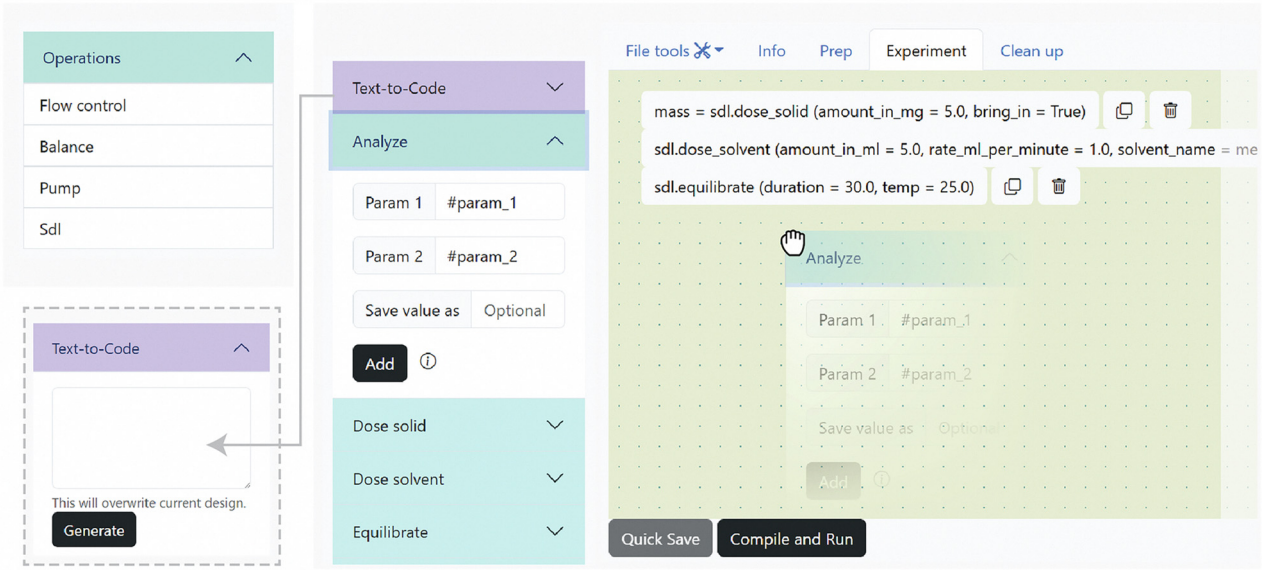

Fig. 12 Ivory OS provides an interoperable drag-and-drop interface that allows researchers to design, translate, and execute experimental workflows in Python, making self-driving laboratories more accessible to non-expert users. Reproduced from ref. 37 with permission from Springer Nature.

profiles, flow rates, sensor precision) and incomplete contextual metadata can reduce reproducibility.[36] Likewise, digital experiment recipes may not transfer seamlessly, as execution often depends on hardware-specific details or communication protocols that are not yet standardized.[36] In addition, there is currently no widely adopted framework for synchronizing results and samples across sites, which can limit multi-step, multi-laboratory campaigns. The lack of common calibration benchmarks and machine-readable formats adds further complexity. Without a shared digital infrastructure for exchanging data and protocols, SDLs may remain somewhat fragmented, making it more difficult to realize their full potential for scalability, reproducibility, and global collaboration.

In parallel with these interoperability challenges, the role of the human scientist within the SDL ecosystem remains underdeveloped. While SDLs have demonstrated the ability to execute complex experiments with limited human intervention,[3–28] the precise position of the researcher in these workflows is still not clearly defined. Current interfaces[36,199] between researchers and autonomous platforms are often developer-oriented, requiring programming skills or prior experience with automation software. This can pose a barrier for many domain scientists (*e.g.*, synthetic chemists, materials scientists) who may have strong disciplinary expertise but limited computational training.[200] Furthermore, decision-making modules such as Bayesian optimization, particularly when relying on complex surrogate models such as Gaussian processes or neural networks, may operate as black boxes, offering limited transparency about their rationale and reducing opportunities for meaningful researcher input. This makes it difficult for scientists to guide or adjust an experimental campaign mid-course based on intuition or evolving research questions. Most existing platforms also lack structured mechanisms for incorporating human-feedback. For example, determining when automation should defer to expert judgement in response to unexpected observations or uncharted experimental conditions. Without greater transparency and opportunities for collaboration, there is a risk that automation may reduce rather than enhance the role of the scientist.

A related challenge lies in the absence of standardized protocols for lab automation architectures. Many current SDLs are tailored to specific experimental setups, making them relatively inflexible and less adaptable to new reactions or materials systems. The limited modularity of such platforms often requires researchers to reconfigure or reprogram components extensively, which increases technical barriers and operational costs. Integrating diverse instruments (*e.g.* reactors, analytical devices) can also be cumbersome in the absence of common interfaces, and failures in one subsystem may disrupt the entire workflow. Overall, the lack of a generalizable architecture constrains the broader applicability of SDLs across chemistry and materials science domains.

Modular SDLs are composed of multiple interconnected robotic units and instruments, making the scheduling of experimental tasks both complex and essential. Currently, many automation modules still rely on relatively simple or serial scheduling strategies, which can lead to severe bottlenecks and reduce overall equipment utilization.[36,38,197] When multiple experiments or users compete for access to the same module, resource conflicts and idle times can occur. For example, overlapping use of a shared robotic arm or analytical instrument may lead to operational delays or even safety concerns such as collisions and deadlocks.[36] Inadequate scheduling can also result in imbalanced workloads—some modules remain idle while others are overburdened—thereby prolonging experimentation time and diminishing the benefits of parallel operations. These challenges become even more pronounced in multi-user environments, where simultaneous

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

experiment requests can easily interfere with one another without proper coordination. In this context, the absence of advanced, centralized scheduling algorithms limits the efficiency of current SDLs, preventing them from fully realizing their potential. As such, intelligent scheduling emerges as a key requirement for next-generation SDLs.

Safety remains one of major barriers to achieving fully autonomous, continuous operation in SDLs.[135] Most existing platforms are unable to manage safety incidents without human oversight, making reliable 24/7 operation challenging. Hazard detection is typically limited to threshold-based alarms (*e.g.*, temperature or pressure sensors), which lack contextual awareness. As a result, subtle or early signs of failure, such as solution leakages, color changes, corrosion, explosion, or abnormal noise/vibration, often go undetected. When issues are detected, current platforms[36,135] generally respond by shutting down the entire system, even in minor cases, leading to unnecessary downtime and experiment interruption.

Equal significant is the lack of automatic recovery mechanisms. As far as we know, no current systems can, for example, clean a spill, re-park a malfunctioning robotic arm, or reroute a sample autonomously. Safety considerations are also not fully integrated into scheduling or orchestration software, allowing tasks to be planned without accounting for safety constraints (*e.g.*, scheduling exothermic reaction in enclosed spaces without ventilation). Predictive diagnostics are largely absent as well, with little proactive monitoring for pumps, valves, heaters, or other hardware for early signs of degradation. These shortcomings are particularly problematic in shared labs or remote operations, where human response time may be delayed or unavailable. Without proactive detection, predictive maintenance, and real-time mitigating strategies, SDLs remain vulnerable to cascading failures, compromised data quality, and equipment damage. Overall, current systems remain heavily dependent on human supervision, limiting both their scalability and their ability to function safely in unattended or high-risk environments.

On the other hand, although Transformer[172]-based generative AI[60,74–76,141] holds great promise for accelerating discovery within SDLs,[48,194] its practical deployment remains constrained by several challenges. Many generative models in chemistry or materials science are typically trained only on structural or compositional data, without incorporating domain knowledge such as reactivity rules, thermodynamics, or synthetic feasibility. As a result, they often generate suggestions that are unrealistic, unsafe, or experimentally infeasible. Interpretability is another limitation: most generative models operate as black boxes, offering little chemical rationale for their predictions. This lack of transparency hampers their adoption as reliable scientific tools. Moreover, integration with laboratory

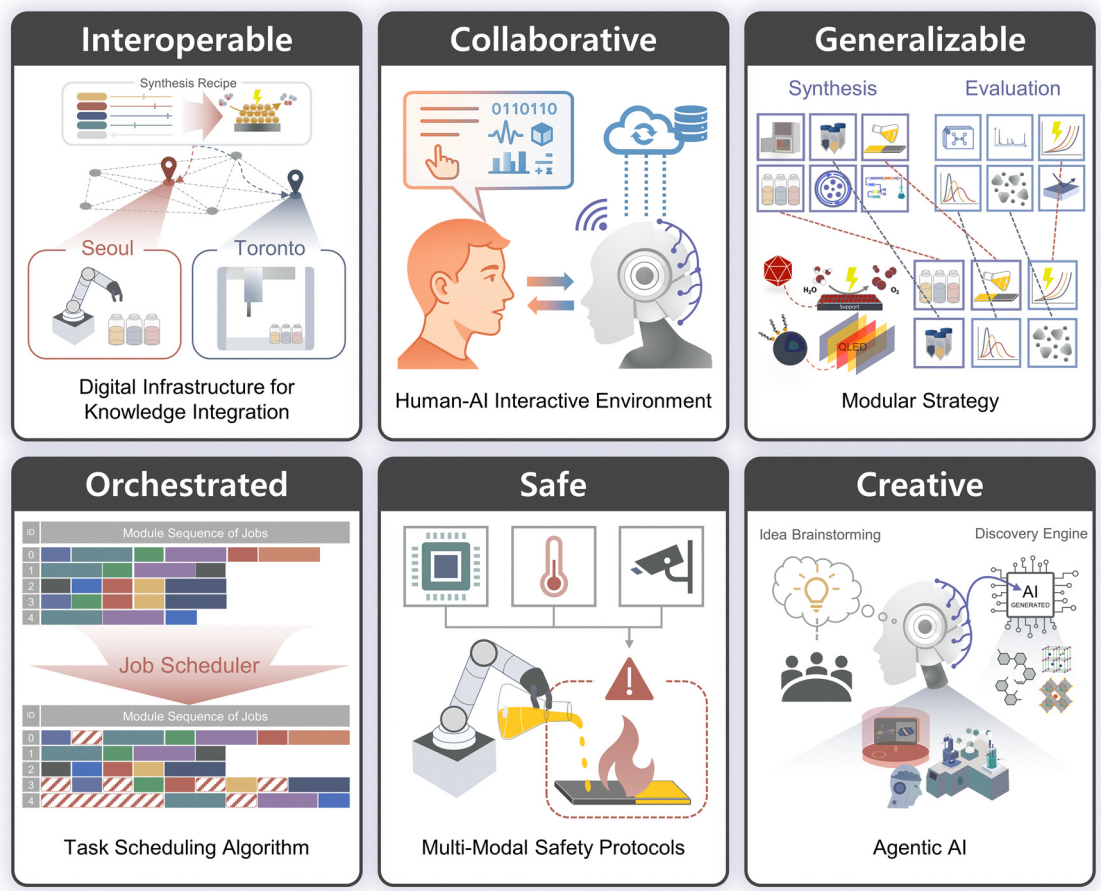

Fig. 13 Six characteristics of SDL 2.0: Interoperable, Collaborative, Generalizable, Orchestrated, Safe, and Creative.

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

automation remains underdeveloped. In most workflows, human experts must translate AI-generated candidates into executable experimental protocols,[201] creating bottlenecks and increasing the risk of human error. Accessibility also poses a significant barrier. Training or fine-tuning large generative models requires substantial computational resources, making them impractical for smaller laboratories.[113] Without lightweight or cloud-based alternatives, access to these tools is largely restricted to well-funded institutions or corporation. Finally, despite the emergence of several SDL communities worldwide,[202–206] progress is fragmented: the lack of shared and community-driven datasets or models limits generalizability and stifles collaborative progress.

Looking ahead, SDL 2.0 must embody six defining characteristics (Fig. 13): *interoperable* (accessible and usable by a broad community of researchers), *collaborative* (human-AI interactive), *generalizable* (adaptable across diverse experimental domains and tasks), *orchestrated* (equipped with intelligent scheduling to minimize bottlenecks and resource conflicts), *safe* (ensuring reliable operations through proactive safety protocols), and *creative* (capable of curating and testing novel hypotheses). Together, these attributes will enable SDLs to evolve from isolated automation tools into transformative engines of scientific discovery.

5.1. Interoperable: digital infrastructures for knowledge transfer and integration across distributed platforms

Building interoperability in SDLs requires a shared digital infrastructure that treats experimental procedures, chemical and material data, and physical workflows as portable, standardized knowledge.[29,35,36,207] This vision hinges on several key elements. First, the community must establish ontologies and metadata schemas for chemistry and materials science. These standards should describe not only what was done, but also how and why, in machine-readable form. Such schemas must capture procedural context, hardware-specific tolerances, calibration status, and external environmental variables. Standardized vocabularies for experimental actions (*e.g.*, inject, stir, heat) and units of measure are also essential to ensure semantic clarity.[32,35,36] Second, digital recipes should be version-controlled and provenance- tracked, enabling modifications to be transparent and reproducible. To maximize portability, these recipes must be hardware-agnostic yet interpretable across different setups using a normalization layer analogous to hardware drivers in a computer OS.[35,36,58] Third, reference experiments and calibration materials should be defined to benchmark reproducibility across laboratories. Such standards would allow different SDLs to validate external workflows and align their platforms in terms of experimental fidelity.[34] Finally, inter-platform exchange protocols are needed to standardize the transfer of physical samples.[36,39] For instance, QR-coded or digitally tagged vials containing machine-readable metadata (*e.g.*, precursor identity, synthesis conditions, storage environment) would allow a lab to seamlessly continue an experimental workflow initiated elsewhere. The ultimate vision is a globally interconnected network of SDLs—platforms that do not merely replicate workflows but co-evolve them, continuously learning from each other's data and experimental results.

A standardized, interoperable SDL ecosystem would unlock a new paradigm of distributed, collective experimentation. Once experimental knowledge complete with metadata and experimental procedure can be shared like open-source software, research becomes dramatically more scalable. Scientists anywhere could retrieve validated digital recipes, adapt them to local objectives, and contribute results back to shared repositories. This creates a real-time feedback loop for rapid iteration, broader validation, and a cumulative knowledge building. Such infrastructure also democratizes access to advanced experimentation. Smaller institutions or labs in under-resourced regions could leverage standardized protocols without duplicating costly hardware setups, provided their platforms comply with common standards. They could also participate in global initiatives by running modular workflow or contributing validation experiments. Interoperability further enables meta-analysis and machine learning at scale. Harmonized datasets[208] spanning different labs, countries, and experiment types form a powerful foundation for training generalizable AI models. These models can uncover transferable synthesis-structure–property relationships, guide future experiments, and even autonomously propose new research directions. Ultimately, by facilitating reproducibility, inclusivity, and accelerated knowledge generation, a standardized knowledge-transfer ecosystem positions SDLs as collaborative engines of global scientific discovery.

5.2. Collaborative: human-AI collaborative environments

To fully realize the potential of SDLs, future systems must be designed to amplify human intelligence rather than simply automate processes.[57] A key direction is the creation of intuitive, user-friendly interfaces that allow researchers to design, query, and adapt experiments through natural and convenient interactions.[57] Instead of hard-coding robotic sequences, scientists should be able to construct workflows using drag-and-drop GUIs[34,209–211] or natural language commands *via* LLM-based assistants.[67] Such application-level interfaces abstract away low-level hardware control, enabling researchers to focus on scientific objectives rather than infrastructure details.[212] Equally important is the integration of explainable AI (xAI) and uncertainty-aware models. Rather than acting as opaque decision-makers, future SDLs should provide justifications for their choices explaining why a particular recipe was recommended or which data supported a conclusion. This transparency allows scientists to evaluate, critique, or override AI reasoning when necessary.

Another promising direction is the implementation of graduated levels of autonomy. Routine tasks can be delegated to automation, while decision points remain under human supervision. For example, experimental checkpoints[213] can pause workflows for expert review in cases involving anomalous data, safety concerns, or high-stakes resource use. Beyond this, translational AI frameworks could combine algorithmic optimization with human insight, incorporating real-time feedback

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

into the experimental loop. Researchers might adjust reward functions, inject new hypotheses, or halt unsafe directions mid-loop, ensuring that experimentation remains both flexible and safe. Achieving this vision will require not only new technical architectures but also sociotechnical studies to better understand how scientists interact with autonomous systems and how workflows can be designed to feel natural, productive, and empowering.

In this paradigm, the scientist becomes a strategic guide and creative partner, not a passive observer. SDLs that learn from, communicate with, and adapt to their users transform automation into an inclusive and empowering technology. Lowering technical barriers through intuitive interfaces ensures that even researchers without programming or robotic expertise can immediately benefit. For instance, a graduate student in materials science could design and optimize a complex experimental campaign without writing a single line of code. Such collaborative SDLs also act as intellectual force multipliers:[11,192] small research groups with modest automation setups can achieve scales of experimentation once reserved for large facilities, enabled by AI-enhanced throughput and real-time feedback. Moreover, this approach fosters global collaboration. Teams across continents could jointly supervise experiments, with AI mediating their input in a shared could-based platform. Time zone differences even become strengths—one group adjusts parameters during the day, another analyzes results overnight—creating a continuous, 24-hour scientific discovery cycle. Ultimately, designing SDLs as human-AI collaborative platforms ensures that automation is not a replacement for human reasoning but a partner in discovery. By paring human intuition and creativity with AI-driven rigor and scalability, this approach promises a more inclusive, transparent, and transformative research culture—one capable of unlocking scientific advances that neither humans nor machines could achieve alone.

5.3. *Generalizable*: modular strategies for flexible and adaptable platforms

To overcome the rigidity of conventional SDLs, future directions point toward modular and reconfigurable "plug-and-play" architectures.[35–37,63] In this framework, hardware and software are developed as interoperable modules with standardized interfaces, allowing seamless integration and rapid reconfiguration. For example, a modular chemistry lab could enable researchers to interchange reactor types (*e.g.*, flow, batch, or photoreactors) and analytical instruments (*e.g.*, GC–MS, X-ray, HPLC) depending on the experimental need.[59] This flexibility dramatically reduces the barrier to adopting new protocols and broadens the range of accessible applications. Realizing this vision requires progress on several fronts: the development of universal communication standards such as SiLA[55] and OPCUA,[215] modular robotics (*e.g.*, versatile robotic arms and liquid handlers[53,65]), and open-source orchestration software capable of coordinating diverse devices through high-level commands rather than low-level coding. Such orchestration tools[35,36,39] should also simplify hardware control into intuitive workflows that can be configured by non-experts. Current initiatives including research consortia and open-hardware projects are beginning to establish these foundations, moving SDLs closer to "scientific LEGO systems": extensible, interoperable, and adaptable to evolving research needs.

Adopting modular design and standardization will help democratize automated experimentation. By reducing reliance on custom engineering, modular SDLs lower the entry barrier for smaller research groups: instead of building bespoke systems, scientists could assemble SDLs from ready-made modules much like customizing a computer. Standardization further promotes reproducibility and knowledge sharing, as protocols and hardware setups developed in one lab could be reproduced elsewhere simply by reusing the same module configurations. For chemists or material scientists, a single modular SDL can flexibly switch between synthesizing pharmaceuticals, catalysts, or polymers, thereby accelerating progress across subfields. Ultimately, modular strategies make SDLs more generalizable and accessible. As costs decrease and open-source designs spread, even resource-constrained labs will gain the ability to participate in high-throughput, AI-driven research, expanding the global impact of SDLs.

5.4. *Orchestrated*: scheduling algorithms for modular and scalable platforms

As SDLs become more modular and multi-user, the need for intelligent scheduling grows increasingly important. Future systems should incorporate centralized scheduling architectures that operate much like an operating system (OS), dynamically allocating tasks across robotic modules with awareness of device status and availability. Instead of relying on static timelines or hard-coded queues, advanced schedulers can integrate task-priority rules, Boolean-based masking tables to monitor resource usage, and predictive analytics to anticipate conflicts before they occur.[36] For example, by cross-checking a task's device requirements against real-time usage, the scheduler can prevent collisions and idle periods. Workflows represented as directed graphs[35] can be analyzed by constraint solvers and optimization algorithms (*e.g.*, makespan minimization) to enable efficient parallelization, while reinforcement learning or graph neural networks could further adapt scheduling to live experimental feedback.[184] Crucially, schedulers must remain scalable and interoperable as labs evolve—recognizing new hardware, balancing workloads, and orchestrating workflows across diverse domains such as synthesis, analysis, measurement. In this way, this scheduler functions as the brain of the SDL, translating high-level experimental goals into coordinated, low-level robotic actions in the most efficient order possible.

The benefits of advanced scheduling extend beyond efficiency. A robust, OS-like scheduler enables SDLs to operate as flexible, continuously running platforms, improving throughput, reducing bottlenecks, and enhancing reproducibility by ensuring that experiments follow standardized, precisely timed sequences. For shared labs and core facilities, intelligent scheduling makes multi-user environments more practical, balancing tasks fairly across different projects and embedding safety

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

features that prevent hazardous overlaps and enforce cool-down intervals. On a larger scale, such infrastructure supports cloud-based SDLs, where researchers can submit experiments remotely and rely on the scheduler to allocate time and resources. This democratizes access to automation: experimentation becomes as seamless as submitting a computing job to a server, lowering barriers, increasing equity, and transforming SDLs into globally accessible engines for discovery.

5.5. *Safe*: multi-modal safety protocols based on detection-recovery processes

Ensuring safe, fully autonomous operation requires SDLs to adopt a two-step safety framework: intelligent hazard detection followed by autonomous recovery. Safety detection[135] must extend beyond traditional vision system and static sensors toward AI-driven situational awareness, integrating multimodal sensing such as computer vision, gas analysis, acoustic and thermal profiling, and real-time signal interpretation. These systems would learn baseline normal behavior and continuously compare it to live data, enabling early detection of abnormal trends such as subtle discoloration before a spill, or vibrations indicating imminent pump failure. Once a hazard is identified, the system should initiate an active recovery protocol rather than simply halting all activity. For example, a robotic arm could clean a spill, a failed reaction might be quenched with a built-in reservoir, or a secondary robotic unit could isolate and reconfigure specific modules. Predictive diagnostics will also play a critical role, flagging equipment issues in advance (*e.g.*, declining pump flow rates, rising motor resistance) to support preemptive maintenance. Importantly, the task scheduler itself should embed safety logic, refusing to execute or even queue unsafe experimental sequences. Existing frameworks such as OCTOPUS[36] and CLAIRify,[140] which employ formal program verification, highlight how only validated and safe experiment paths can be authorized. In the long term, SDLs will evolve toward a degree of ''safety self-awareness'', monitoring themselves like vigilant lab technicians, responding to hazards within seconds, and maintaining unaffected tasks in parallel. Such systems will not only remove human exposure to risk but also increase uptime and operational robustness,[64,65,68,135,214] laying the foundation for lights-out, globally distributed laboratories.

Robust autonomous safety mechanisms promise to reshape experimental practice. First, they enable continuous operation across nights, weekends, and time zones, substantially increasing throughput. A single SDL could sustain multi-week optimization campaigns with minimal human intervention. Second, intelligent safety improves scientific rigor and reproducibility: when experiments are executed under consistent, well-monitored conditions, free from human timing errors, data reliability improves. Automatic logging of sensor states, interventions, and recovery actions provides a transparent digital record, supporting *post hoc* analysis and reproducibility checks. These safety pipelines also protect equipment and the surrounding environment by immediately addressing hazards such as leaks, overheating, or emissions. This result is reduced waste, improved sustainability, and longer instrument lifetimes. Finally, the widespread adoption of such safety protocols will help shape new laboratory safety standards. Institutions and regulators will be more willing to adopt fully automated labs once autonomous safety is demonstrably more robust than traditional methods. Embedding safety into the core design of SDLs ensures that autonomy scales responsibly, making advanced science faster, safer, and more equitable.

5.6. *Creative*: generative AI assistants for advanced platforms

To realize the creative potential of generative AI within SDLs, future efforts must prioritize integration, accessibility, and scientific rigor. Domain-aware models are essential—systems that not only learn from structural or compositional datasets but also encode principles of chemistry and materials science, including safety, feasibility, and physical constraints. Such models could incorporate prior knowledge *via* rule-based constraints or physics-informed learning. Within a closed-loop setting, generative AI can leverage real-time feedback from experimental outcomes to iteratively refine hypotheses and guide subsequent experimental decisions in a data-driven manner.

Seamless interfacing between generative models and lab automation software is also critical. Instead of requiring human translation, models[68,214] should generate machine-readable protocols that autonomous platforms can execute directly. Shared standards—whether protocol languages or APIs—will be central to bridging ideation and execution. At the same time, accessible interfaces are necessary to broaden adoption: intuitive web platforms, drag-and-drop GUIs, or natural language interaction *via* LLMs[68] would enable any researcher to harness generative AI without specialized expertise. Lightweight solutions based on transfer learning or model distillation could further democratize, lowering computational barriers for smaller institutions. Community-driven development will also be key: shared, open generative models trained on disturbed experimental datasets can evolve into collective scientific resources, improving as new results are contributed globally. In this vision, generative AI becomes a partner in discovery rather than a solitary tool. A materials scientist could request a compound with a targeted conductivity, and the SDL would autonomously propose candidates, test them, and refine its suggestions, while the scientist provides strategic guidance, interprets results, setting priorities. This synergy between human creativity and AI-driven hypothesis generation represents the next stage of SDL development.

Generative AI also expand the creative scope of experimentation. Beyond accelerating established workflows, it enables exploration of new molecular and materials spaces, discovery of novel materials. By navigating vast combinatorial design spaces with targeted precision, generative AI helps avoid combinatorial confusion while unlocking innovative possibilities. When these models are shared and accessible, they become engines of collective progress: each experiment contributes knowledge that improves the next generation of ideas, creating a self-reinforcing cycle of discovery across institutions and continents. Such democratization ensures that creativity is no

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

longer limited by local resources. A student in a resource-constrained lab could discuss or collaborate with a global model trained on millions of reactions, using it to design and execute experiments through cloud-connected SDLs. This not only amplifies inclusivity but also accelerates global progress. Generative and agentic AIs also hold educational value: early-career researchers can interact with it to test hypotheses, explore mechanisms, or critique model proposals, turning it into both a teaching tool and a discovery engine. In the long term, integration of generative AI into autonomous experimentation will define a new scientific paradigm: one that is not only faster but also more creative, collaborative, and inclusive. SDLs will evolve from executing predefined instructions to co-writing them, catalyzing scientific innovation at scales and speeds that traditional approaches cannot match.

6. Conclusion

Across these six dimensions—interoperability, collaboration, modularity, orchestration, safety, and creativity through generative AI—a coherent vision for the next generation of SDLs comes into focus. Future SDLs will not be narrow, task-specific tools, but generalizable platforms capable of adapting to diverse experimental challenges. They will be democratized infrastructures, accessible to a broad range of researchers regardless of institutional resources, and globally interconnected engines of discovery, where knowledge and results flow seamlessly across borders. Realizing this vision will require parallel advances in technology, shared standards, and community practices. Technical progress must be matched by efforts to build trust, ensure inclusivity, and foster open knowledge exchange. The payoff, however, is profound: SDLs can accelerate discovery cycles, close the gap between hypothesis and experiment, and bringing more voices into the process of scientific innovation. Crucially, next-generation SDLs will not remain confined to a handful of high-tech laboratories. Instead, they will evolve into a distributed, cooperative ecosystem that enables scientists everywhere to pursue research more creatively, efficiently, and safely. By following this roadmap, the chemistry and materials science communities can ensure that SDLs mature in line with core values of openness, reproducibility, and shared benefit. In this light, the emergence of SDLs is not only a technical trajectory but also a collective mission to democratize and accelerate discovery. By linking human creativity with autonomous experimentation on a global scale, we have the opportunity to transform not just the pace of science, but also its accessibility and inclusiveness. The journey is already underway, and with coordinated global collaboration SDLs are poised to redefine how scientific knowledge is created and shared in the decades to come.

Author contributions

S. S. H. and Y. J. P.: supervision. S. S. H., H. L., and H. J. Y.: conceptualization. H. J. Y., H. L., and S. S. H.: writing – original draft, writing – review & editing. H. L., H. S. J., B. P., and H. J. Y.: data curation, visualization. All authors contributed to manuscript writing and approved the final version of the manuscript.

Conflicts of interest

The authors declare no competing interests.

Data availability

No primary research results, software or code have been included and no new data were generated or analyzed as part of this review. Data for this review article are presented in the cited papers mentioned in the figures and texts.

Acknowledgements

This work was supported by the National Research Foundation of Korea funded by the Ministry of Science and ICT [NRF-2022M3H4A7046278 & RS-2024-00450102].

References

1 F. Häse, L. M. Roch and A. Aspuru-Guzik, *Trends Chem.*, 2019, **1**, 282–291.
2 E. Stach, B. DeCost, A. G. Kusne, J. Hattrick-Simpers, K. A. Brown, K. G. Reyes, J. Schrier, S. Billinge, T. Buonassisi, I. Foster, C. P. Gomes, J. M. Gregoire, A. Mehta, J. Montoya, E. Olivetti, C. Park, E. Rotenberg, S. K. Saikin, S. Smullin, V. Stanev and B. Maruyama, *Matter*, 2021, **4**, 2702–2726.
3 J. X.-Y. Lim, D. Leow, Q.-C. Pham and C.-H. Tan, *IEEE Trans. Autom. Sci. Eng.*, 2021, **18**, 2185–2190.
4 C. W. Coley, D. A. Thomas, J. A. M. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers, H. Gao, R. W. Hicklin, P. P. Plehiers, J. Byington, J. S. Piotti, W. H. Green, A. John Hart, T. F. Jamison and K. F. Jensen, *Science*, 2019, **365**, eaax1566.
5 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, V. Lai, G. J. Ng, H. Situ, R. H. Zhang, M. S. Elliott, T. H. Haley, D. J. Dvorak, A. Aspuru-Guzik, J. E. Hein and C. P. Berlinguette, *Sci. Adv.*, 2020, **6**, 1–8.
6 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, *Science*, 2019, **363**, eaav2211.
7 H. Tao, T. Wu, S. Kheiri, M. Aldeghi, A. A. Aspuru-Guzik and E. Kumacheva, *Adv. Funct. Mater.*, 2021, **31**, 2106725.
8 D. Angelone, A. J. S. Hammer, S. Rohrbach, S. Krambeck, J. M. Granda, J. Wolf, S. Zalesskiy, G. Chisholm and L. Cronin, *Nat. Chem.*, 2021, **13**, 63–69.
9 F. Mekki-Berrada, Z. Ren, T. Huang, W. K. Wong, F. Zheng, J. Xie, I. P. S. Tian, S. Jayavelu, Z. Mahfoud, D. Bash, K. Hippalgaonkar, S. Khan, T. Buonassisi, Q. Li and X. Wang, *npj Comput. Mater.*, 2021, **7**, 55.

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

10 B. P. MacLeod, F. G. L. Parlane, C. C. Rupnow, K. E. Dettelbach, M. S. Elliott, T. D. Morrissey, T. H. Haley, O. Proskurin, M. B. Rooney, N. Taherimakhsousi, D. J. Dvorak, H. N. Chiu, C. E. B. Waizenegger, K. Ocean, M. Mokhtari and C. P. Berlinguette, *Nat. Commun.*, 2022, **13**, 995.
11 N. H. Angello, D. M. Friday, C. Hwang, S. Yi, A. H. Cheng, T. C. Torres-Flores, E. R. Jira, W. Wang, A. A. Aspuru-Guzik, M. D. Burke, C. M. Schroeder, Y. Diao and N. E. Jackson, *Nature*, 2024, **633**, 351–358.
12 F. Bateni, R. W. Epps, K. Antami, R. Dargis, J. A. Bennett, K. G. Reyes and M. Abolhasani, *Adv. Intell. Syst.*, 2022, **4**, 2200017.
13 Y. Jiang, D. Salley, A. Sharma, G. Keenan, M. Mullin and L. Cronin, *Sci. Adv.*, 2022, **8**, 1–12.
14 K. Molga, S. Szymkuć, P. Gołębiowska, O. Popik, P. Dittwald, M. Moskal, R. Roszak, J. Mlynarski and B. A. Grzybowski, *Nat. Synth.*, 2022, **1**, 49–58.
15 S. H. M. Mehr, D. Caramelli and L. Cronin, *Proc. Natl. Acad. Sci. U. S. A.*, 2023, **120**, e2220045120.
16 D. Salley, G. Keenan, J. Grizou, A. Sharma, S. Martín and L. Cronin, *Nat. Commun.*, 2020, **11**, 2771.
17 B. A. Koscher, R. B. Canty, M. A. McDonald, K. P. Greenman, C. J. McGill, C. L. Bilodeau, W. Jin, H. Wu, F. H. Vermeire, B. Jin, T. Hart, T. Kulesza, S.-C. Li, T. S. Jaakkola, R. Barzilay, R. Gómez-Bombarelli, W. H. Green and K. F. Jensen, *Science*, 2023, **382**, eadi1407.
18 A. A. Volk, R. W. Epps, D. T. Yonemoto, B. S. Masters, F. N. Castellano, K. G. Reyes and M. Abolhasani, *Nat. Commun.*, 2023, **14**, 1403.
19 Q. Zhu, Y. Huang, D. Zhou, L. Zhao, L. Guo, R. Yang, Z. Sun, M. Luo, F. Zhang, H. Xiao, X. Tang, X. Zhang, T. Song, X. Li, B. Chong, J. Zhou, Y. Zhang, B. Zhang, J. Cao, G. Zhang, S. Wang, G. Ye, W. Zhang, H. Zhao, S. Cong, H. Li, L.-L. Ling, Z. Zhang, W. Shang, J. Jiang and Y. Luo, *Nat. Synth.*, 2023, **3**, 319–328.
20 H. J. Yoo, N. Kim, H. Lee, D. Kim, L. T. C. Ow, H. Nam, C. Kim, S. Y. Lee, K.-Y. Lee, D. Kim and S. S. Han, *Adv. Funct. Mater.*, 2024, **34**, 2312561.
21 K. Higgins, S. M. Valleti, M. Ziatdinov, S. Kalinin and M. Ahmadi, *ACS Energy Lett.*, 2020, **5**, 3426–3436.
22 D. S. Salley, G. A. Keenan, D.-L. Long, N. L. Bell and L. Cronin, *ACS Cent. Sci.*, 2020, **6**, 1587–1593.
23 L. Porwol, D. J. Kowalski, A. Henson, D. L. Long, N. L. Bell and L. Cronin, *Angew. Chem., Int. Ed.*, 2020, **59**, 11256–11261.
24 B. Burger, P. M. Maffettone, V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R. S. Sprick and A. I. Cooper, *Nature*, 2020, **583**, 237–241.
25 X. Du, L. Lüer, T. Heumueller, J. Wagner, C. Berger, T. Osterrieder, J. Wortmann, S. Langner, U. Vongsaysy, M. Bertrand, N. Li, T. Stubhan, J. Hauch and C. J. Brabec, *Joule*, 2021, **5**, 495–506.
26 R. W. Epps, M. S. Bowen, A. A. Volk, K. Abdel-Latif, S. Han, K. G. Reyes, A. Amassian and M. Abolhasani, *Adv. Mater.*, 2020, **32**, 1–9.
27 N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng and G. Ceder, *Nature*, 2023, **624**, 86–91.
28 E. Gu, X. Tang, S. Langner, P. Duchstein, Y. Zhao, I. Levchuk, V. Kalancha, T. Stubhan, J. Hauch, H. J. Egelhaaf, D. Zahn, A. Osvet and C. J. Brabec, *Joule*, 2020, **4**, 1806–1822.
29 R. J. Hickman, M. Aldeghi, F. Häse and A. A. Aspuru-Guzik, *Digital Discovery*, 2022, **1**, 732–744.
30 F. Häse, L. i M. Roch, C. Kreisbeck and A. a Aspuru-Guzik, *ACS Cent. Sci.*, 2018, **4**, 1134–1145.
31 F. Hase, L. M. Roch and A. Aspuru-Guzik, *Chem. Sci.*, 2018, **9**, 7642–7655.
32 J. M. Granda, L. Donina, V. Dragone, D. L. Long and L. Cronin, *Nature*, 2018, **559**, 377–381.
33 J. Mockus, *Bayesian Approach to Global Optimization*, Springer, Dordrecht, 1989.
34 R. Tamura, K. Tsuda and S. Matsuda, *Sci. Technol. Adv. Mater.:Methods*, 2023, **3**, 2232297.
35 Y. Fei, B. Rendy, R. Kumar, O. Dartsi, H. P. Sahasrabuddhe, M. J. McDermott, Z. Wang, N. J. Szymanski, L. N. Walters and D. Milsted, *Dig. Discovery*, 2024, **3**, 2275–2288.
36 H. J. Yoo, K.-Y. Lee, D. Kim and S. S. Han, *Nat. Commun.*, 2024, **15**, 9669.
37 W. Zhang, L. Hao, V. Lai, R. Corkery, J. Jessiman, J. Zhang, J. Liu, Y. Sato, M. Politi, M. E. Reish, R. Greenwood, N. Depner, J. Min, R. El-khawaldeh, P. Prieto, E. Trushina and J. E. Hein, *Nat. Commun.*, 2025, **16**, 5182.
38 J. Zhou, M. Luo, L. Chen, Q. Zhu, S. Jiang, F. Zhang, W. Shang and J. Jiang, *Digital Discovery*, 2025, **4**, 636–652.
39 M. Sim, M. G. Vakili, F. Strieth-Kalthoff, H. Hao, R. J. Hickman, S. Miret, S. Pablo-García and A. Aspuru-Guzik, *Matter*, 2024, **7**, 2959–2977.
40 J. J. Günzl, 2025, Accelerate conference, Workshop: Self-driving lab for all: build, automate, discovery.
41 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, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. T. Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S. A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons, *Sci Data*, 2016, **3**, 160018.
42 K. Stracke and J. D. Evans, *Commun. Chem.*, 2024, **7**, 63.
43 L. C. Brinson, L. M. Bartolo, B. Blaiszik, D. Elbert, I. Foster, A. Strachan and P. W. Voorhees, *MRS Bull.*, 2024, **49**, 12–16.
44 H. Kim, H. Choi, D. Kang, W. B. Lee and J. Na, *Chem. Sci.*, 2024, **15**, 7908–7925.
45 C. Karpovich, E. Pan and E. A. Olivetti, *npj Comput. Mater.*, 2024, **10**, 287.

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

46 S. P. Stier, C. Kreisbeck, H. Ihssen, M. A. Popp, J. Hauch, K. Malek, M. Reynaud, T. P. M. Goumans, J. Carlsson, I. Todorov, L. Gold, A. Rader, W. Wenzel, S. T. Bandesha, P. Jacques, F. Garcia-Moreno, O. Arcelus, P. Friederich, S. Clark, M. Maglione, A. Laukkanen, I. E. Castelli, J. Carrasco, M. C. Cabanas, H. S. Stein, O. Ozcan, D. Elbert, K. Reuter, C. Scheurer, M. Demura, S. S. Han, T. Vegge, S. Nakamae, M. Fabrizio and M. Kozdras, *Adv. Mater.*, 2024, **36**, e2407791.

47 K. Fushimi, Y. Nakai, A. Nishi, R. Suzuki, M. Ikegami, R. Nimura, T. Tomono, R. Hidese, H. Yasueda, Y. Tagawa and T. Hasunuma, *Sci. Rep.*, 2025, **15**, 6648.

48 G. Tom, S. P. Schmid, S. G. Baird, Y. Cao, K. Darvish, H. Hao, S. Lo, S. Pablo- García, E. M. Rajaonson, M. Skreta, N. Yoshikawa, S. Corapi, G. D. Akkoc, F. Strieth-Kalthoff, M. Seifrid and A. A. Aspuru-Guzik, *Chem. Rev.*, 2024, **124**, 9633–9732.

49 S. Lo, S. G. Baird, J. Schrier, B. Blaiszik, N. Carson, I. Foster, A. Aguilar-Granda, S. V. Kalinin, B. Maruyama, M. Politi, H. Tran, T. D. Sparks and A. Aspuru-Guzik, *Digital Discovery*, 2024, **3**, 842–868.

50 M. Seifrid, R. Pollice, A. E. Aguilar-Granda, Z. Morgan Chan, K. Hotta, C. T. Ser, J. Vestfrid, T. C. Wu and A. A. Aspuru-Guzik, *Acc. Chem. Res.*, 2022, **55**, 2454–2466.

51 O. Bayley, E. Savino, A. Slattery and T. Noël, *Matter*, 2024, **7**, 2382–2398.

52 C. Wang, Y. J. Kim, A. Vriza, R. Batra, A. Baskaran, N. Shan, N. Li, P. Darancet, L. Ward, Y. Liu, M. K. Y. Chan, S. Sankaranarayanan, H. C. Fry, C. S. Miller, H. Chan and J. Xu, *Nat. Commun.*, 2025, **16**, 1498.

53 H. Zhao, W. Chen, H. Huang, Z. Sun, Z. Chen, L. Wu, B. Zhang, F. Lai, Z. Wang, M. L. Adam, C. H. Pang, P. K. Chu, Y. Lu, T. Wu, J. Jiang, Z. Yin and X.-F. Yu, *Nat. Synth.*, 2023, **2**, 505–514.

54 J. Bai, S. Mosbach, C. J. Taylor, D. Karan, K. F. Lee, S. D. Rihm, J. Akroyd, A. A. Lapkin and M. Kraft, *Nat. Commun.*, 2024, **15**, 462.

55 D. Juchli, *Adv. Biochem. Eng. Biotechnol.*, 2022, **182**, 147–174.

56 M. B. Rooney, B. P. MacLeod, R. Oldford, Z. J. Thompson, K. L. White, J. Tungjunyatham, B. J. Stankiewicz and C. P. Berlinguette, *Digital Discovery*, 2022, **1**, 382–389.

57 A. V. Tobias and A. Wahab, *R. Soc. Open Sci.*, 2025, **12**, 250646.

58 H. Hysmith, E. Foadian, S. P. Padhy, S. V. Kalinin, R. G. Moore, O. S. Ovchinnikova and M. Ahmadi, *Digital Discovery*, 2024, **3**, 621–636.

59 R. J. Hickman, M. Sim, S. Pablo-García, G. Tom, I. Woolhouse, H. Hao, Z. Bao, P. Bannigan, C. Allen, M. Aldeghi and A. Aspuru-Guzik, *Digital Discovery*, 2025, **4**, 1006–1029.

60 A. Priyanshu, Y. Maurya and Z. Hong, *arXiv preprint*, 2024, arXiv:2407.01557, DOI: 10.48550/arXiv.2407.01557.

61 A. Mottafegh, G. N. Ahn and D. P. Kim, *Lab Chip*, 2023, **23**, 1613–1621.

62 D. Packwood, *Bayesian Optimization for Materials Science*, Springer, 2017.

63 R. R. Griffiths and J. M. Hernandez-Lobato, *Chem. Sci.*, 2020, **11**, 577–586.

64 S. S. Sajjan, M. Moore, M. Pan, G. Nagaraja, J. Lee, A. Zeng and S. Song, 2020 IEEE International Conference on Robotics and Automation (ICRA), 2020, pp. 3634–3642.

65 S. Eppel, H. Xu, M. Bismuth and A. Aspuru-Guzik, *ACS Cent. Sci.*, 2020, **6**, 1743–1752.

66 R. El-Khawaldeh, M. Guy, F. Bork, N. Taherimakhsousi, K. N. Jones, J. M. Hawkins, L. Han, R. P. Pritchard, B. A. Cole, S. Monfette and J. E. Hein, *Chem. Sci.*, 2024, **15**, 1271–1282.

67 H. Lee, D. Kim, H. Lee, N. Gwak, N. Kim, H. J. Yoo, T. Yu, N. Oh, S. S. Sohn and S. S. Han, 2025, DOI: 10.26434/chemrxiv-2025-px33t-v2.

68 K. Darvish, M. Skreta, Y. Zhao, N. Yoshikawa, S. Som, M. Bogdanovic, Y. Cao, H. Hao, H. Xu, A. A. Aspuru-Guzik, A. Garg and F. Shkurti, *Matter*, 2025, **8**, 101897.

69 Y. R. Wang, Y. Zhao, H. Xu, S. Eppel, A. Aspuru-Guzik, F. Shkurti and A. Garg, *arXiv*, preprint, 2023, arXiv:2302.11683, DOI: 10.48550/arXiv.2302.11683.

70 Y. Nakajima, M. Hamaya, K. Tanaka, T. Hawai, F. von Drigalski, Y. Takeichi, Y. Ushiku and K. Ono, Presented in part at the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2023.

71 Y. Huang, J. Zhang, R. Yu, S. Li and W. Ding, *J. Field Robot.*, 2025, **42**, 2908–2919.

72 H. Barrington, A. Dickinson, J. McGuire, C. Yan and M. Reid, *Org. Process Res. Dev.*, 2022, **26**, 3073–3088.

73 Y. Li, B. Dutta, Q. J. Yeow, R. Clowes, C. E. Boott and A. I. Cooper, *Digital Discovery*, 2025, **4**, 1276–1283.

74 A. I. DeepSeek, *arXiv*, *preprint*, 2024, arXiv:2401.02954, DOI: 10.48550/arXiv.2401.02954.

75 H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. E. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave and G. Lample, *arXiv*, preprint, 2023, arXiv:2302.13971, DOI: 10.48550/arXiv.2302.13971.

76 G. Yenduri, M. Ramalingam, G. C. Selvi, Y. Supriya, G. Srivastava, P. K. R. Maddikunta, G. D. Raj, R. H. Jhaveri, B. Prabadevi, W. Wang, A. Vasilakos and T. R. Gadekallu, *IEEE Access*, 2024, **12**, 54608–54649.

77 Y. Kang and J. Kim, *Nat. Commun.*, 2024, **15**, 4705.

78 T. Song, M. Luo, X. Zhang, L. Chen, Y. Huang, J. Cao, Q. Zhu, D. Liu, B. Zhang, G. Zou, G. Zhang, F. Zhang, W. Shang, Y. Fu, J. Jiang and Y. Luo, *J. Am. Chem. Soc.*, 2025, **147**, 12534–12545.

79 T. J. Callahan, N. H. Park and S. Capponi, *arXiv*, preprint, 2025, arXiv:2502.19629, DOI: 10.48550/arXiv.2502.19629.

80 C. Z. Shuyi Jia and V. Fung, *arXiv*, *preprint*, 2025, arXiv.2406.13163, DOI: 10.48550/arXiv.2406.13163.

81 J.-P. Correa-Baena, K. Hippalgaonkar, J. van Duren, S. Jaffer, V. R. Chandrasekhar, V. Stevanovic, C. Wadia, S. Guha and T. Buonassisi, *Joule*, 2018, **2**, 1410–1420.

82 T. Bartz-Beielstein, J. Branke, J. Mehnen and O. Mersmann, *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.*, 2014, **4**, 178–195.

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

83 J. Kennedy and R. Eberhart, Proceedings of ICNN'95 – International Conference on Neural Networks, 1995.
84 J. H. Holland, *Sci. Am.*, 1992, **267**, 66–73.
85 L. P. Kaelbling, M. L. Littman and A. W. Moore, *J. Artif. Intell. Res.*, 1996, **4**, 237–285.
86 B. J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J. I. M. Alvarado, J. M. Janey, R. P. Adams and A. G. Doyle, *Nature*, 2021, **590**, 89–96.
87 D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue and T. Lookman, *Nat. Commun.*, 2016, **7**, 11241.
88 J. Sun, P. Lin, L. Zeng, Z. Guo, Y. Jiang, C. Xiao, Q. Jian, J. Ren, L. Pan, X. Xu, Z. Li, L. Wei and T. Zhao, *Nat. Commun.*, 2025, **16**, 6528.
89 D. F. Colin Doumont, N. Maus, J. R. Gardner, H. Moss and G. Pleiss, *arXiv*, *preprint*, 2025, arXiv.2512.00170, DOI: 10.48550/arXiv.2512.00170.
90 S. Ju, T. Shiga, L. Feng, Z. Hou, K. Tsuda and J. Shiomi, *Phys. Rev. X*, 2017, **7**, 021024.
91 K. Kandasamy, J. Schneider and B. Poczos, Presented in part at the Proceedings of the 32nd International Conference on Machine Learning, Proceedings of Machine Learning Research, 2015, **37**, 295–304.
92 Z. Wang, C. Gehring, P. Kohli and S. Jegelka, Presented in part at the Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, 2018, **84**, 745–754.
93 F. Thelen, R. Zehl, R. Zerdoumi, J. L. Bürgel, L. Banko, W. Schuhmann and A. Ludwig, *Adv. Sci.*, 2025, **12**, e07302.
94 D. Eriksson, M. Pearce, J. Gardner, R. D. Turner and M. Poloczek, 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), 2019, **32**, 5496–5507.
95 K. Wang and A. W. Dowling, *Curr. Opin. Chem. Eng.*, 2022, **36**, 100728.
96 J. Snoek, H. Larochelle and R. P. Adams, Advances in Neural Information Processing Systems 25 (NIPS 2012), 2012, **25**, 2951–2959.
97 R. J. Hickman, G. Tom, Y. Zou, M. Aldeghi and A. Aspuru-Guzik, *Digital Discovery*, 2025, **4**, 2104–2122.
98 R. Liang, S. Zheng, K. Wang and Z. Yuan, *ACS Electrochem.*, 2025, **1**, 360–368.
99 Z. Xie, X. Evangelopoulos, J. C. R. Thacker and A. I. Cooper, in ECAI 2023, 2023, DOI: 10.3233/faia230587.
100 D. Khatamsaz, R. Neuberger, A. M. Roy, S. H. Zadeh, R. Otis and R. Arróyave, *npj Comput. Mater.*, 2023, **9**, 221.
101 F. Di Fiore and L. Mainini, *Comput. Struct.*, 2024, **296**, 107302.
102 M. Luo, Z. Xie, H. Li, B. Zhang, J. Cao, Y. Huang, H. Qu, Q. Zhu, L. Chen, J. Jiang and Y. Luo, *Matter*, 2025, **8**, 102009.
103 D. Frey, J. H. Shin, C. Musco and M. A. Modestino, *React. Chem. Eng.*, 2022, **7**, 855–865.
104 K. Taha, *IEEE Access*, 2020, **8**, 80855–80878.
105 A. Biswas, C. Fuentes and C. Hoyle, *J. Mech. Design*, 2022, **144**, 011703.
106 S. Daulton, D. Eriksson, M. Balandat and E. Bakshy, Presented in part at the Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence, Proceedings of Machine Learning Research, 2022, **180**, 507–517.
107 I. A. Bespalov, N. V. Krivoshchapov, A. A. Lisov, V. A. Chaliy and M. G. Medvedev, *J. Chem. Inf. Model.*, 2025, **65**, 6048–6056.
108 V. Sabanza-Gil, R. Barbano, D. Pacheco Gutierrez, J. S. Luterbacher, J. M. Hernandez-Lobato, P. Schwaller and L. Roch, *Nat. Comput. Sci.*, 2025, **5**, 572–581.
109 N. Asprion, R. Böttcher, R. Pack, M. E. Stavrou, J. Höller, J. Schwientek and M. Bortz, *Chem. Ing. Tech.*, 2018, **91**, 305–313.
110 Q. Ke and C. M. Simon, *Nat. Comput. Sci.*, 2025, **5**, 518–519.
111 M. S. Priyadarshini, O. Romiluyi, Y. Wang, K. Miskin, C. Ganley and P. Clancy, *Mater. Horiz.*, 2024, **11**, 781–791.
112 A. K. Y. Low, F. Mekki-Berrada, A. Gupta, A. Ostudin, J. Xie, E. Vissol-Gaudin, Y.-F. Lim, Q. Li, Y. S. Ong, S. A. Khan and K. Hippalgaonkar, *npj Comput. Mater.*, 2024, **10**, 104.
113 H. J. Yoo, D. Kim, S. Yim and S. S. Han, 2025, DOI: 10.26434/chemrxiv-2025-96bmc-v2.
114 X. E. Abdoulatif Cisse, V. V. Gusev and A. I. Cooper, *arXiv*, preprint, 2025, arXiv.2501.16224, DOI: 10.48550/arXiv.2501.16224.
115 F. Häse, M. Aldeghi, R. J. Hickman, L. M. Roch and A. Aspuru-Guzik, *Appl. Phys. Rev.*, 2021, **8**, 031406.
116 B. Ranković and P. Schwaller, *arXiv*, *preprint*, 2025, arXiv:2504.06265, DOI: 10.48550/arXiv.2504.06265.
117 M. Rajabi-Kochi, N. Mahboubi, A. P. S. Gill and S. M. Moosavi, *Chem. Sci.*, 2025, **16**, 5464–5474.
118 P. Shetty, A. C. Rajan, C. Kuenneth, S. Gupta, L. P. Panchumarti, L. Holm, C. Zhang and R. Ramprasad, *NPJ Comput. Mater.*, 2023, **9**, 52.
119 T. Chen and C. Guestrin, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, 785–794.
120 I. Sekulic, J. Schaible, G. Müller, M. Plock, S. Burger, V. J. Martínez-Lahuerta, N. Gaaloul and P.-I. Schneider, *Mach. Learn.: Sci. Technol.*, 2025, **6**, 040503.
121 N. Gantzler, A. Deshwal, J. R. Doppa and C. M. Simon, *Digital Discovery*, 2023, **2**, 1937–1956.
122 J. Kim, M. Li, Y. Li, A. Gómez, O. Hinder and P. W. Leu, *Digital Discovery*, 2024, **3**, 381–391.
123 L. D. González and V. M. Zavala, *Ind. Eng. Chem. Res.*, 2025, **64**, 2168–2182.
124 S. X. Leong, C. E. Griesbach, R. Zhang, K. Darvish, Y. Zhao, A. Mandal, Y. Zou, H. Hao, V. Bernales and A. Aspuru-Guzik, *Nat. Rev. Chem.*, 2025, **9**, 707–722.
125 Y. Kosenkov and D. Kosenkov, *J. Chem. Educ.*, 2021, **98**, 4067–4073.
126 S. Uyanik, S. Parkinson, G. Killick, B. Dutta, R. Clowes, C. E. Boott and A. I. Cooper, *Digital Discovery*, 2025, **4**, 2816–2826.
127 S. Zhou, B. Chen, E. S. Fu and H. Yan, *Microsyst. Nanoeng.*, 2023, **9**, 116.

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

128 A. C. Sun, J. A. Jurica, H. B. Rose, G. Brito, N. R. Deprez, S. T. Grosser, A. M. Hyde, E. E. Kwan and S. Moor, *Org. Process Res. Dev.*, 2023, **27**, 1954–1964.
129 R. El-khawaldeh, A. Mandal, N. Yoshikawa, W. Zhang, R. Corkery, P. Prieto, A. Aspuru-Guzik, K. Darvish and J. E. Hein, *Device*, 2024, **2**, 100404.
130 C. Yan, M. Cowie, C. Howcutt, K. M. P. Wheelhouse, N. S. Hodnett, M. Kollie, M. Gildea, M. H. Goodfellow and M. Reid, *Chem. Sci.*, 2023, **14**, 5323–5331.
131 P. Shiri, V. Lai, T. Zepel, D. Griffin, J. Reifman, S. Clark, S. Grunert, L. P. E. Yunker, S. Steiner, H. Situ, F. Yang, P. L. Prieto and J. E. Hein, *iScience*, 2021, **24**, 102176.
132 C. Fyfe, R. Duncan, T. J. D. McCabe, K. Donnachie, H. Barrington and M. Reid, *ACS Sustain. Chem. Eng.*, 2025, **13**, 17241–17256.
133 D. Capel and A. Zisserman, *IEEE Signal Process. Mag.*, 2003, **20**, 75–86.
134 V. Wiley and T. Lucas, *Int. J. Artif. Intell. Res.*, 2018, **2**, 29–36.
135 L. C. O. Tiong, H. J. Yoo, N. Kim, C. Kim, K.-Y. Lee, S. S. Han and D. Kim, *npj Comput. Mater.*, 2024, **10**, 42.
136 S. U. Khan, V. K. Møller, R. J. N. Frandsen and M. Mansourvar, *Appl. Intell.*, 2025, **55**, 524.
137 T. Zepel, V. Lai, L. P. Yunker and J. E. Hein, 2020, DOI: 10.26434/chemrxiv.12798143.v1.
138 A. Z. L. Naruki Yoshikawa, K. Darvish, Y. Zhao, H. Xu, A. Kuramshin, A. Aspuru-Guzik, A. Garg and F. Shkurti, *arXiv*, *preprint*, 2022, arXiv:2212.09672, DOI: 10.48550/arXiv.2212.09672.
139 M. Kennedy, K. Queen, D. Thakur, K. Daniilidis and V. Kumar, 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017, 1260–1266.
140 N. Yoshikawa, M. Skreta, K. Darvish, S. Arellano-Rubach, Z. Ji, L. Bjørn Kristensen, A. Z. Li, Y. Zhao, H. Xu, A. Kuramshin, A. A. Aspuru-Guzik, F. Shkurti and A. Garg, *Auton. Robots*, 2023, **47**, 1057–1086.
141 Team Gemini Google, *arXiv*, *preprint*, 2023, arXiv.2312.11805, DOI: 10.48550/arXiv.2312.11805.
142 M. C. Ramos, C. J. Collison and A. D. White, *Chem. Sci.*, 2025, **16**, 2514–2572.
143 OpenAI, ChatGPT, https://chat.openai.com/.
144 T. Hartung, *Front. Artif. Intell.*, 2025, **8**, 1649155.
145 A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White and P. Schwaller, *Nat. Mach. Intell.*, 2024, **6**, 525–535.
146 D. A. Boiko, R. MacKnight, B. Kline and G. Gomes, *Nature*, 2023, **624**, 570–578.
147 I. Mandal, J. Soni, M. Zaki, M. M. Smedskjaer, K. Wondraczek, L. Wondraczek, N. N. Gosvami and N. M. A. Krishnan, *Nat. Commun.*, 2025, **16**, 9104.
148 T. Zheng, Z. Deng, H. T. Tsang, W. Wang, J. Bai, Z. Wang and Y. Song Suzhou, *arXiv*, preprint, 2025, arXiv:2505.13259, DOI: 10.48550/arXiv.2505.13259.
149 M. Schilling-Wilhelmi, M. Ríos-García, S. Shabih, M. V. Gil, S. Miret, C. T. Koch, J. A. Márquez and K. M. Jablonka, *Chem. Soc. Rev.*, 2025, **54**, 1125–1150.
150 Z. Zhao, D. Ma, L. Chen, L. Sun, Z. Li, Y. Xia, B. Chen, H. Xu, Z. Zhu, S. Zhu, S. Fan, G. Shen, K. Yu and X. Chen, *Cell Rep. Phys. Sci.*, 2025, **6**, 102523.
151 D. Zhang, W. Liu, Q. Tan, J. Chen, H. Yan, Y. Yan, J. Li, W. Huang, X. Yue and W. Ouyang, *arXiv*, preprint, 2024, arXiv:2402.06852, DOI: 10.48550/arXiv.2402.06852.
152 J. Liang, W. Huang, F. Xia, P. Xu, K. Hausman, B. Ichter, P. Florence and A. Zeng, Presented in part at the 2023 IEEE International Conference on Robotics and Automation (ICRA), 2023, 9493–9500.
153 T. Schick, J. Dwivedi-Yu, R. Dessì, R. Raileanu, M. Lomeli, E. Hambro, L. Zettlemoyer, N. Cancedda and T. Scialom, *Adv. Neural Inf. Process. Syst.*, 2023, **36**, 68539–68551.
154 G. Izacard, P. Lewis, M. Lomeli, L. Hosseini, F. Petroni, T. Schick, J. Dwivedi-Yu, A. Joulin, S. Riedel and E. Grave, *J. Mach. Learn. Res.*, 2023, **24**, 1–43.
155 K. Guu, K. Lee, Z. Tung, P. Pasupat and M. Chang, *International conference on machine learning*, 2020.
156 P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. T. Yih and T. Rocktäschel, *Adv. Neural Inform. Process. Syst.*, 2020, **33**, 9459–9474.
157 T. Gupta, M. Zaki, N. M. A. Krishnan and Mausam, *Npj Comput. Mater.*, 2022, **8**, 102.
158 J. Van Herck, M. V. Gil, K. M. Jablonka, A. Abrudan, A. S. Anker, M. Asgari, B. Blaiszik, A. Buffo, L. Choudhury, C. Corminboeuf, H. Daglar, A. M. Elahi, I. T. Foster, S. Garcia, M. Garvin, G. Godin, L. L. Good, J. Gu, N. Xiao Hu, X. Jin, T. Junkers, S. Keskin, T. P. J. Knowles, R. Laplaza, M. Lessona, S. Majumdar, H. Mashhadimoslem, R. D. McIntosh, S. M. Moosavi, B. Mourino, F. Nerli, C. Pevida, N. Poudineh, M. Rajabi-Kochi, K. L. Saar, F. Hooriabad Saboor, M. Sagharichiha, K. J. Schmidt, J. Shi, E. Simone, D. Svatunek, M. Taddei, I. Tetko, D. Tolnai, S. Vahdatifar, J. Whitmer, D. C. F. Wieland, R. Willumeit-Romer, A. Zuttel and B. Smit, *Chem. Sci.*, 2025, **16**, 670–684.
159 Y. Zhang, Y. Han, S. Chen, R. Yu, X. Zhao, X. Liu, K. Zeng, M. Yu, J. Tian, F. Zhu, X. Yang, Y. Jin and Y. Xu, *Nat. Mach. Intell.*, 2025, **7**, 1010–1022.
160 W. Zhang, Q. Wang, X. Kong, J. Xiong, S. Ni, D. Cao, B. Niu, M. Chen, Y. Li, R. Zhang, Y. Wang, L. Zhang, X. Li, Z. Xiong, Q. Shi, Z. Huang, Z. Fu and M. Zheng, *Chem. Sci.*, 2024, **15**, 10600–10611.
161 Y. Ruan, C. Lu, N. Xu, Y. He, Y. Chen, J. Zhang, J. Xuan, J. Pan, Q. Fang, H. Gao, X. Shen, N. Ye, Q. Zhang and Y. Mo, *Nat. Commun.*, 2024, **15**, 10160.
162 H. Zhang, Y. Song, Z. Hou, S. Miret and B. Liu, *arXiv*, preprint, 2024, arXiv:2409.00135, DOI: 10.48550/arXiv.2409.00135.
163 Y. L. Shuaihang Chen, W. Han, W. Zhang and T. Liu, *arXiv*, preprint, 2024, arXiv:2412.17481, DOI: 10.48550/arXiv.2412.17481.
164 J. Becker, *arXiv*, preprint, 2024, arXiv:2410.22932, DOI: 10.48550/arXiv.2410.22932.
165 Q. Wu, G. Bansal, J. Zhang, Y. Wu, B. Li, E. Zhu, L. Jiang, X. Zhang, S. Zhang, J. Liu, A. H. Awadallah, R. W. White,

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

D. Burger and C. Wang, First Conference on Language Modeling (COLM), 2024.
166 X. Huang, M. Surve, Y. Liu, T. Luo, O. Wiest, X. Zhang and N. V. Chawla, Presented in part at the Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, Boise, ID, USA, 2024, 3797–3801.
167 T. Hishiki, N. Collier, C. Nobata, T. Okazaki-Ohta, N. Ogata, T. Sekimizu, R. Steiner, H. S. Park and J. Tsujii, *Genome. Inform. Ser. Workshop Genome. Inform.*, 1998, **9**, 81–90.
168 S. Liu, A. B. McCoy, Q. Chen and A. Wright, *Int. J. Med. Inform.*, 2025, **205**, 106104.
169 M. Krauthammer and G. Hripcsak, *Proc. AMIA Symp.*, 2001, 339–343.
170 B. Libbus and T. C. Rindflesch, *Proc. AMIA Symp.*, 2002, 445–449.
171 I. S. Tomas Mikolov, K. Chen, G. Corrado and J. Dean, *arXiv*, preprint, 2013.
172 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin, Advances in Neural Information Processing Systems, 2017, **30**, 5998–6008.
173 J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, 4171–4186.
174 Q. Liu, M. P. Polak, S. Y. Kim, M. D. A. A. Shuvo, H. S. Deodhar, J. Han, D. Morgan and H. Oh, *Acta Mater.*, 2025, **297**, 121307.
175 T. Inagaki, A. Kato, K. Takahashi, H. Ozaki and G. N. Kanda, *arXiv*, preprint, 2023, arXiv.2304.10267, DOI: 10.48550/arXiv.2304.10267.
176 H. Huo, C. J. Bartel, T. He, A. Trewartha, A. Dunn, B. Ouyang, A. Jain and G. Ceder, *Chem. Mater.*, 2022, **34**, 7323–7336.
177 B. Zitkovich, T. Yu, S. Xu, P. Xu, T. Xiao, F. Xia, J. Wu, P. Wohlhart, S. Welker and A. Wahid, *Conference on Robot Learning*, 2023, **229**, 2165–2183.
178 F. Rahmanian, J. Flowers, D. Guevarra, M. Richter, M. Fichtner, P. Donnely, J. M. Gregoire and H. S. Stein, *Adv. Mater. Interfaces*, 2022, **9**, 2101987.
179 A. Angelopoulos, C. Baykal, J. Kandel, M. Verber, J. F. Cahoon and R. Alterovitz, *IEEE International Conference on Robotics and Automation*, 2025, 15900–15906.
180 J. Gao, J. Chang, H. Que, Y. Xiong, S. Zhang, X. Qi, Z. Liu, J.-J. Wang, Q. Ding and X. Li, *arXiv*, preprint, 2025, arXiv:2512.21766, DOI: 10.48550/arXiv.2512.21766.
181 J. Li, Y. Tu, R. Liu, Y. Lu and X. Zhu, *Adv. Sci.*, 2020, **7**, 1901957.
182 L. M. Roch, F. Häse, C. Kreisbeck, T. Tamayo-Mendoza, L. P. Yunker, J. E. Hein and A. Aspuru-Guzik, *Sci. Robotics*, 2018, **3**, eaat5559.
183 H. Fakhruldeen, G. Pizzuto, J. Glowacki and A. I. Cooper, 2022 International Conference on Robotics and Automation (ICRA), 2022, 6013–6019.
184 R. B. Canty, B. A. Koscher, M. A. McDonald and K. F. Jensen, *Digital Discovery*, 2023, **2**, 1259–1268.
185 W. Á. P. Galambos and K. Széll, IEEE 24th International Conference on Intelligent Engineering Systems (INES),2020, 171–177.
186 G. Gauglitz, *Anal. Bioanal. Chem.*, 2018, **410**, 5093–5094.
187 B. Costa, J. Bachiega Jr, L. R. De Carvalho and A. P. Araujo, *ACM Computing Surveys (CSUR)*, 2022, **55**, 1–34.
188 T. J. Jacobsson, A. Hultqvist, A. García-Fernández, A. Anand, A. Al-Ashouri, A. Hagfeldt, A. Crovetto, A. Abate, A. G. Ricciardulli and A. Vijayan, *Nat. Energy*, 2022, **7**, 107–115.
189 M. Folk, G. Heber, Q. Koziol, E. Pourmal and D. Robinson, Proceedings of the EDBT/ICDT 2011 workshop on array databases, 2011, 36–47.
190 A. Viloria, G. C. Acuña, D. J. A. Franco, H. Hernández-Palma, J. P. Fuentes and E. P. Rambal, *Proc. Comput. Sci.*, 2019, **155**, 575–580.
191 S. P. Huber, S. Zoupanos, M. Uhrin, L. Talirz, L. Kahle, R. Häuselmann, D. Gresch, T. Müller, A. V. Yakutovich, C. W. Andersen, F. F. Ramirez, C. S. Adorf, F. Gargiulo, S. Kumbhar, E. Passaro, C. Johnston, A. Merkys, A. Cepellotti, N. Mounet, N. Marzari, B. Kozinsky and G. Pizzi, *Sci. Data*, 2020, **7**, 300.
192 R. Rauschen, M. Guy, J. E. Hein and L. Cronin, *Nat. Synth.*, 2024, **3**, 488–496.
193 MongoDB, MongoDB Atlas: Fully Managed Cloud Database, https://www.mongodb.com/products/platform/atlas-database.
194 B. Madika, A. Saha, C. Kang, B. Buyantogtokh, J. Agar, C. M. Wolverton, P. Voorhees, P. Littlewood, S. Kalinin and S. Hong, *ACS Nano*, 2025, **19**, 27116–27158.
195 M. Abolhasani and E. Kumacheva, *Nat. Synth.*, 2023, **2**, 483–492.
196 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, M. Hereld, H. Joress, N. Jung, H.-K. Kwon, G. Pizzuto, J. Rintamaki, C. Steinmann, L. Torresi and S. Sun, *Digital Discovery*, 2023, **2**, 1644–1659.
197 Y. Kim, H. Doo, D. Shin, S. Y. Lee, Y. Roh, S. Park, H. Song, Y. Jung, H. J. Yoo, S. S. Han, J. W. Kim, M. O. Besenhard, Y. S. Lee and J. Na, *Comput. Chem. Eng.*, 2025, **203**, 109266.
198 D. Guevarra, K. Kan, Y. Lai, R. J. R. Jones, L. Zhou, P. Donnelly, M. Richter, H. S. Stein and J. M. Gregoire, *Digital Discovery*, 2023, **2**, 1806–1812.
199 C. Elliott, V. Vijayakumar, W. Zink and R. Hansen, *J. Lab. Autom.*, 2007, **12**, 17–24.
200 K. L. Snapp and K. A. Brown, *Digital Discovery*, 2023, **2**, 1620–1629.
201 N. Kim, H. J. Yoo, D. Kim, H. Lee and S. S. Han, 2025, DOI: 10.26434/chemrxiv-2025-3lj65-v2.
202 B. Pelkie, S. Baird, E. Aissi, K. Aspuru-Takata, Y. Cao, J. H. Chang, K. Gambhir, W. S. Hale, L. Hao, C. Hattrick, J. Hein, D. Luo, O. Melville, M. Ngan, L. L. B. Nyeland, N. Peek, M. Politi, E. E. Rajkumar, A. Siemenn, B. Subbaraman, S. Vasquez, J. Watchorn, W. Zhang, R. O. Ziskason, L. Pozzo, T. Buonassisi and T. Vegge, 2025, DOI: 10.26434/chemrxiv-2025-zhkrf.

 This journal is © The Royal Society of Chemistry 2026

Open Access Article. Published on 04 März 2026. Downloaded on 07.06.2026 10:33:45.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

203 Acceleration Consortium, https://acceleration.utoronto.ca.
204 CAPeX, https://capex.dtu.dk/.
205 N. Savage, *Biopharma Dealmakers*, 2021, DOI: 10.1038/d43747-021-00045-7.
206 Umicore enters AI platform agreement with Microsoft, https://www.umicore.com/en/newsroom/umicore-enters-ai-platform-agreement-with-microsoft-to-accelerate-and-scale-its-battery-materials-technologies-development.
207 S. Hessam, M. Craven, A. I. Leonov, G. Keenan and L. Cronin, *Science*, 2020, **370**, 101–108.
208 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.
209 A. I. Leonov, A. J. S. Hammer, S. Lach, S. H. M. Mehr, D. Caramelli, D. Angelone, A. Khan, S. O'Sullivan, M. Craven, L. Wilbraham and L. Cronin, *Nat. Commun.*, 2024, **15**, 1240.
210 T. Ha, D. Lee, Y. Kwon, M. S. Park, S. Lee, J. Jang, B. Choi, H. Jeon, J. Kim, H. Choi, H. T. Seo, W. Choi, W. Hong, Y. J. Park, J. Jang, J. Cho, B. Kim, H. Kwon, G. Kim, W. S. Oh, J. W. Kim, J. Choi, M. Min, A. Jeon, Y. Jung, E. Kim, H. Lee and Y. S. Choi, *Sci. Adv.*, 2023, **9**, eadj0461.
211 A. Slattery, Z. Wen, P. Tenblad, J. U. Sanjosé-Orduna, D. Pintossi, T. den Hartog and T. Noël, *Science*, 2024, **383**, eadj1817.
212 A. J. S. Hammer, A. I. Leonov, N. L. Bell and L. Cronin, *JACS Au*, 2021, **1**, 1572–1587.
213 R. B. Canty and K. F. Jensen, *Nat. Synth.*, 2024, **3**, 428–429.
214 M. H. Schwarz and J. Börcsök, 2013 XXIV International Conference on Information, Communication and Automation Technologies (ICAT), 2013, 1–6.
215 Y. Zhao, M. Bogdanovic, C. Luo, S. Tohme, K. Darvish, A. A. Aspuru-Guzik, F. Shkurti and A. Garg, *arXiv*, preprint, 2025. arXiv:2502.04531, DOI: 10.48550/arXiv.2502.04531.