The rise of self-driving labs in chemical and materials sciences

Milad Abolhasani ©[1] ✉ & Eugenia Kumacheva[2,3,4]

Accelerating the discovery of new molecules and materials, as well as developing green and sustainable ways to synthesize them, will help to address global challenges in energy, sustainability and healthcare. The recent growth of data science and automated experimentation techniques has resulted in the advent of self-driving labs (SDLs) via the integration of machine learning, lab automation and robotics. An SDL is a machine-learning-assisted modular experimental platform that iteratively operates a series of experiments selected by the machine learning algorithm to achieve a user-defined objective. These intelligent robotic assistants help researchers to accelerate the pace of fundamental and applied research through rapid exploration of the chemical space. In this Review, we introduce SDLs and provide a roadmap for their implementation by non-expert scientists. We present the status quo of successful SDL implementations in the field and discuss their current limitations and future opportunities to accelerate finding solutions for societal needs.

Finding tangible solutions for global challenges in energy, sustainability and healthcare is the cornerstone of the research, economic and societal activities; however, the current strategies to address these challenges are time, resource and labour intensive. From the first practical demonstration of a silicon solar cell in 1954, it took more than half a century to find a more cost-effective material than silicon, and yet it is not deployed at scale[1]. The timeframe of drug discovery and development is typically ten years, with a cost of more than US$1 billion (ref. [2]). Despite the worldwide acknowledgement of climate change and environmental pollution with plastics more than 20 years ago[3,4], currently there is no scalable technological solution for effective carbon capture and seawater treatment. These examples share a common challenge: the need to explore a vast number of continuous and discrete experimental variables to find the most effective composition as well as manufacturing routes of molecules and materials. Current exploration strategies in chemical and materials sciences rely on prior knowledge and, experimentally, on changing variables one at a time or in a combinatorial fashion. Despite the straightforward nature of these approaches, they do not meet the required pace of discovery in chemical and materials sciences to address the global challenges

in energy, sustainability and healthcare[5]. Although initially highly promising, combinatorial screening strategies did not make a major breakthrough in the fields of energy materials or small molecules, due to the exponential growth of the number of required experiments with every added experimental variable.

In addition, the slow progress in chemical space exploration is attributed to: 1) the physical disconnection between the stages of synthesis, characterization and performance evaluation in a conventional chemistry and materials science lab, as well as (2) the time gap between performing an experiment and making a decision about the conditions of the next experiment(s) to find a new compound or material with the targeted properties, identify an optimized synthetic route for an existing compound or unveil the underlying mechanism of a complex reaction. The physical disconnection refers to the siloed nature of the conventional research efforts on the discovery of new materials and molecules. Finding innovative solutions for large-scale global problems requires an interdisciplinary approach to experimental chemical and materials science. The siloed format of conventional chemistry and materials science labs slows down the very much required interdisciplinary research. For example, in the conventional experimental efforts

[1]Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC, USA. [2]Department of Chemistry, University of Toronto, Toronto, Ontario, Canada. [3]Institute of Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada. [4]Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario, Canada. ✉e-mail: abolhasani@ncsu.edu
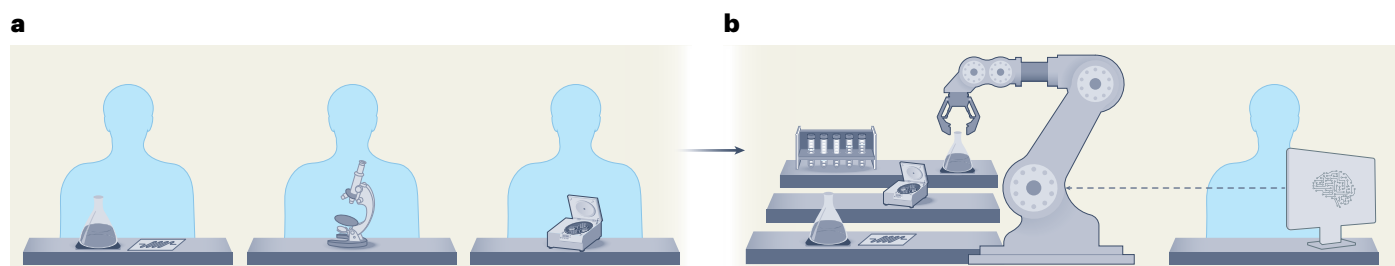
a **b**

Fig. 1 | Conventional versus self-driving labs. a,b, Illustration of the transition from a conventional chemistry and materials science lab (**a**) to an SDL for each researcher (**b**) to address the challenges of the current disconnected experimental workflows by using modular robotic experimentation and the intelligent planning of experiments.

on the discovery of clean energy technologies, different research groups study materials and develop devices. Materials scientists and device engineers typically work separately on different aspects of clean energy technologies. As a result, solution-processable clean energy materials are being sought after without considering their specific requirements at the device level, and device architectures are being optimized without the best-performing material at hand. In addition, this disconnection of the material synthesis and device-level integration results in an inefficient research operation without taking advantage of intermediate information (materials properties). These limitations stem from the current human-dependent approach to research in every step of an experimental workflow. The COVID-19 pandemic exposed the strong reliance on 'in person' presence for conventional experimental research, and the laboratory shutdowns led researchers to think about their approach to experimental research in academic and industrial settings[6].

The vast size and high dimensionality (dimension refers here to a continuous or discrete experimental variable) of the chemical design spaces that need to be experimentally explored require new integrated strategies to accelerate the discovery of new molecules and advanced functional materials, as well as to find sustainable ways for their scaled-up synthesis and manufacturing.

Recent advances in robotics[7,8] and artificial intelligence[9,10] offer an exciting opportunity to reshape research in the experimental chemical and materials sciences. Artificial intelligence, a subfield of computer science, seeks to build machines with human-programmed intelligence (for example, the ability of decision-making). Machine learning (ML), a subfield of artificial intelligence, seeks to build mathematical models for complex tasks and processes with high-dimensional spaces to perform automated operations, such as the prediction of a synthesis outcome or material properties or image classification. The convergence of ML, lab automation (for example, synthesis, separation, purification and characterization) and robotics (for example, reagent preparation and sample transfer between different experimental modules) led to the development of 'self-driving labs' (SDLs)[11]. SDLs leverage scientific and technological advancements made in academia and industry over the past decade in lab automation[12–14], reaction miniaturization (via flow chemistry and microfluidics)[15] and online analytical characterization[16]. In contrast to human-dependent experimental settings in conventional chemistry and materials science labs (Fig. 1a), the SDLs (Fig. 1b) use: (1) robots to operate multiple repetitive tasks that are time-consuming, require precision or pose safety concerns when dealing with toxic or flammable chemicals, and (2) computers that can outperform human scientists for certain tasks, such as handling high-dimensional big data. In this manner, the use of SDLs addresses three challenges of conventional chemistry and materials science labs: (1) inefficient and slow experimental space exploration, (2) physical disconnection between different experimental stages and (3) the time gap between performing an experiment and selecting the next experiment to be tested.

By the robotic integration of experimental modules, SDLs connect the otherwise physically disconnected stages of reagent preparation, synthesis, characterization and performance evaluation to establish an end-to-end experimental workflow for an accelerated synthesis and development of new molecules and materials. The end-to-end nature of SDLs[17] becomes extremely powerful to co-design materials and devices. For example, the co-design of clean energy materials and devices within an SDL equipped with the material synthesis, purification, processing and device integration modules enables another research acceleration opportunity beyond the siloed operation of SDLs only focused on the synthesis or processing aspects of materials.

Importantly, the use of SDLs can avail a substantial amount of the researcher's time to focus on new conceptual or intellectual challenges, rather than on time-consuming repetitive tasks in the lab. Instead of changing one variable at a time, by incorporating ML, SDLs intelligently explore the chemical space and at the same time minimize or eliminate the time gap between acquiring experimental results and decision-making for the conditions of the next experiment. In contrast to the frequently misinterpreted purpose of SDLs (replacing highly trained scientists in research settings), intelligent robotic assistants are meant to accelerate discovery and avail the time of chemists and materials scientists to high-level scientific questions. For example, providing an SDL with a research acceleration of 10 times (Fig. 1b) for each of the researchers shown in Fig. 1a increases their overall research productivity by at least 30 times, which allows them to work on new scientific questions. As a result, SDLs reshape the role of the operator and/or researcher in the chemical and materials science workflow (Fig. 1b). Intelligent experimental planning and autonomous exploration of the experimental space allow scientists to see a big picture of the scientific problem, discard unfavourable synthetic routes and effectively identify impactful intrinsic and extrinsic experimental variables that control the targeted physicochemical properties of molecules or materials.

Over the past decade, promising applications of SDLs were demonstrated to accelerate the synthesis and fabrication of molecules and materials, for example, carbon nanotubes[18], complex organic compounds[13,19–21], nanomaterials[22–27], phase-change memory materials[28] and thin-film materials[29,30]. Yet, the SDL utility in chemical and materials sciences is still limited. The reasons for the slow progress of SDLs are the lack of: (1) standardized and cost-effective hardware, (2) readily accessible software, (3) user-friendly operational guidelines for chemists and materials scientists and (4) the incorporation of physics-based models with autonomous experimentation.

This Review introduces SDLs for experimental chemistry and materials science and highlights recent successful examples for the autonomous synthesis of organic molecules and functional (nano) materials. It provides a roadmap for starting an SDL in a conventional chemistry and materials science lab and discusses the steps toward its successful implementation. The discussion of current limitations and

future opportunities for SDLs serves as a catalyst for academic labs in chemical and materials sciences to accelerate the implementation of this new integrated workflow in their experimental research, and for industry to focus on the standardization of SDL hardware and software for their broad deployment to accelerate the synthesis of new compounds and the development of advanced materials that contribute to scalable future technological solutions.

SDLs in chemical and materials sciences

An SDL is an intelligent experimental platform equipped with different hardware modules that iteratively operate a series of syntheses or physical processes selected and planned by the ML algorithm in a closed-loop format to achieve a predefined objective. The SDL's closed-loop operation refers to the cycle of performing an ML-selected experiment by following an automated series of tasks, acquiring experimental data, updating an ML model and making a decision about the next set of experimental conditions to be tested by the SDL. Examples of tasks performed by the modules include reagent preparation, mixing, synthesis, purification, printing and characterization. An SDL operator defines the closed-loop 'campaign' objective, for example, to identify a new compound with the desired properties, accelerated retrosynthesis of an existing compound or low-temperature manufacturing of a thin-film material. In addition, the SDL operator can leverage the prior domain knowledge and human expertise, such as physics-based models (for example, conservation laws) and an initial hypothesis (for example, about the reaction mechanism), as well as constrains of the reaction conditions, such as the range of temperatures, pressures and reagent concentrations. In this sense, SDLs act as an assistant to scientists in the discovery, exploration, optimization and/or synthesis–structure–property mapping of new molecules and advanced materials. Furthermore, SDLs enable access to unexplored regions of the experimental design space and accelerate the pace of research towards novel compounds. With intelligent experimental planning, big data generated by SDLs can rapidly provide important information about the underlying reaction mechanisms of complex multistage reactions.

Figure 2 illustrates recent implementation of SDLs in chemical and materials sciences. For example, SDLs enabled closed-loop synthesis–property relationship mapping (Fig. 2a) and on-demand synthesis (Fig. 2b) of semiconductor[22,23,26,31–33] and metal[24,25,34] nanoparticles >1,000 times faster than conventional techniques. Chiral metal halide perovskite nanoparticles were revealed by an SDL with 250 autonomously selected and performed experiments (Fig. 2c). Furthermore, the use of SDLs accelerated the discovery of semiconductor and metal thin-film compositions[29,30,35] and their low-temperature processing conditions, 50 °C lower than that of prior art (Fig. 2d)[30]. An eight-day continuous and unattended operation of an SDL (688 experiments) unveiled an effective photocatalyst formulation for hydrogen evolution from water six times more active than that of prior art (Fig. 2e)[21]. The data-driven operation of an SDL reduced the total number of experiments required to identify a high-performing three-dimensional-printed structure with maximum toughness by 60-fold compared with a conventional grid search (Fig. 2f)[36]. In addition to the examples listed, SDLs were recently utilized for the on-demand and on-site manufacturing of active pharmaceutical ingredients[13,19,20].

The main impact of SDLs is the 'research acceleration' to generate new knowledge that leads to the discovery of novel compounds or manufacturing routes of the best-performing materials 10–1,000 times faster than by utilizing one-at-a-time variable exploration or combinatorial experiments. The acceleration factor directly translates into a substantial reduction in research time, cost, resources, waste and carbon footprint in academia and industry. We believe that the accelerated finding of innovative solutions to global problems will be the most impactful contribution of SDLs in the next decade.

A roadmap for SDLs

The most common questions asked by a scientist considering the adoption of an SDL in the chemical or materials science research are, 'Where should I start?', 'What ML algorithms can be used for experiment-selection and data mining?', 'How long does it take to build an SDL?' and 'What would be the cost of building an SDL?'. The answers to these questions are directly related to the type of molecules or materials to be prepared and the goal of research, for example, discovery, exploration, mechanistic study or optimization. Figure 3a presents a general roadmap for SDLs, aimed at answering the question 'Where should I start?'. From the hardware perspective, the targeted class of molecules or materials determines the selection of the required SDL hardware modules by specifying the reagents and the type of reaction (that is, gas, liquid or solid phase), as well as the necessary characterization techniques. From the software perspective, the goal of an SDL operation also determines the scope of the intelligent experiment planning and the required software components for the SDL's closed-loop campaigns.

As shown in Fig. 3a, a modular approach to the implementation of an SDL in a conventional chemistry or materials science lab includes the selection and integration of hardware and software modules to create the experimental design that is best suited for the targeted class of molecules or materials. The preparation of reagents involves robotic handling, stirring, heating and degassing of liquids and/or solids. Depending on the nature of the reaction, miniaturized flow reactors, parallel batch reactors or glass and/or silicon substrates (thin-film materials) are utilized for the automated synthesis under conditions selected by the ML algorithm (software). The purification module of the SDL hardware can include solvent removal (for organic synthesis), centrifugation (for nanomaterial synthesis) or spin coating (for thin-film preparation). The processing module includes the evaluation of the physical or chemical performance of the autonomously produced molecules or materials, for example, their photostability, conductivity or reactivity. Examples of the SDL processing modules include the coating and printing of thin films and nanocrystal inks, bioactivity of active pharmaceutical ingredients in medicinal chemistry and turnover frequency in (photo)catalysis. The characterization module is critically important for the evaluation of the properties of molecules and materials produced in the SDL after each module. The analytical techniques that have already been implemented in SDLs for organic synthesis include high-performance liquid chromatography and gas chromatography, mass spectrometry, nuclear magnetic resonance (NMR) spectroscopy and Fourier-transform infrared spectroscopy. Characterization techniques integrated with SDLs for the autonomous development of nanomaterials and thin films include ultraviolet–visible–near infrared absorption and photoluminescence spectroscopy. When a characterization technique is difficult to dedicate to a specific SDL due to the cost (for example, X-ray diffraction spectroscopy), complicated sample preparation (for example, transmission and scanning electron microscopy) or inaccessibility in the SDL location (for example, a synchrotron light source), ML-assisted parameter space exploration is accomplished at a slower pace than that of a fully autonomous robotic experimentation, as it requires a manual sample preparation and characterization by an operator[37]. In this format, the lack of robotic automation of one or a few experimental steps will lower the overall research throughput compared with that of a fully autonomous robotic experimentation format, but the ML-assisted experiment selection will still make it considerably faster than the conventional exploration strategies in chemical and materials sciences.

As illustrated in Fig. 3a, sample transfer between different modules of an SDL can be handled by stationary[26,29,30,36] or mobile[21] robots or using pumps, valves and tubing[13,20,22,23,25]. When dealing with air- and/or moisture-sensitive chemical compounds, placing the SDL under an inert atmosphere can improve sample handling and data reproducibility. When the characterization module is integrated online with the synthesis module, the reaction sampling can be conducted by using
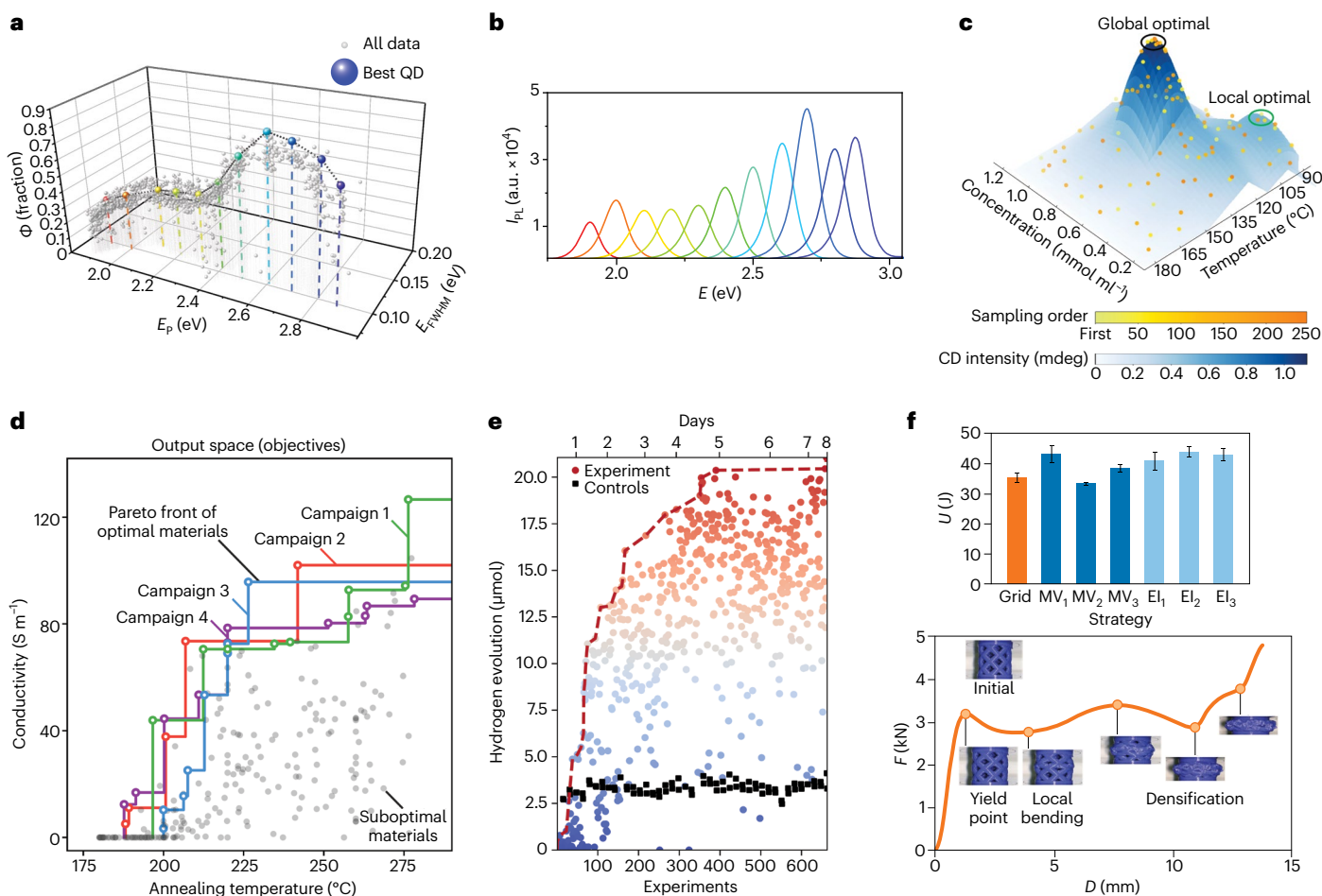
Fig. 2 | Successful examples of SDLs in chemical and materials sciences.
a,b, Autonomous material properties mapping (**a**) and bandgap engineering
of metal halide perovskite quantum dots (QDs) using a flow reactor-based
SDL[23] (**b**). **c**, Autonomous discovery of chiral perovskite nanocrystals using a
flow-reactor-based SDL[26]. **d**, Accelerated discovery of optimal low-temperature
synthesis conditions for spray coating of palladium films using an SDL equipped
with a stationary robotic arm over four closed-loop replicates (campaigns)[30].
e, Autonomous identification of an optimal photocatalyst mixture for the
maximum hydrogen production from water using a batch-reactor-based SDL
equipped with a mobile robotic arm[21]. The dot colour transition from blue to red
indicates an increase in hydrogen evolution. **f**, Top: an example of an SDL with

intelligent experiment selection outperforming a conventional grid search to
rapidly discover the optimal geometry of an additively manufactured part with
maximum toughness (U). Bottom: force (F) versus displacement (D) curve of the
three-dimensional-printed geometry with the maximum U, discovered by an SDL
equipped with a stationary robotic arm[36]. Φ, quantum yield; a.u., arbitrary units;
CD, circular dichroism; E_{FWHM}, emission linewidth; EI, expected improvement; E_p,
peak emission energy; I_{PL}, photoluminescence intensity; MV, maximum variance;
a.u., arbitrary units. Adapted with permission from: **a,b**, ref. [23], Wiley; **c**, ref. [26],
Springer Nature Ltd; **f**, ref. [36], AAAS. Reproduced with permission from: **d**, ref. [30],
Springer Nature Ltd; **e**, ref. [21], Springer Nature Ltd.

valves and pumps[13,20] without the need for the robotic arm. If the char-
acterization technique cannot be directly integrated with the synthesis
module, but can be placed within a close proximity of the synthesis
module, a stationary robotic arm can transfer the sample between the
synthesis and characterization modules of the SDL[26,29,30,36]. When the
characterization module cannot be placed within reach of a stationary
robotic arm, a mobile robotic arm can perform the sample transfer
across the SDL[21]. The configurations of the SDLs with fluidic sample
transfer and a stationary robotic arm require custom-development
and specific integration of the characterization modules with the
synthesis module of the SDLs, whereas mobile robots are a retrofit to
the conventional chemistry and materials science labs. Robotic arms,
in addition to sample transfer, can also be utilized for autonomous
reconfiguration of the end-to-end modular workflow from the starting
reagents to the final purified product, which substantially expands
the SDL's capabilities to explore continuous and discrete variables
and enable access to a larger portion of the design space than that of
conventional experimental platforms.

From the software perspective, data flow between different SDL
modules serves as a key point for closed-loop operations[38,39]. Reliable
data flow using robust data representation and metadata tracking
strategy, that is, recording and reporting the latent features of each
experiment, is required to truly digitize the synthesis and manufac-
turing of molecules and materials with scalable and transferrable
knowledge. An accelerated discovery will only become possible when
standardized and reliable digital data of all the reactions tested by
SDLs become readily available. Equipping SDLs with standardized
data representation and access to the metadata of prior experiments
performed on the same or different SDLs will address the common
challenge of lab-to-lab variations (or irreproducibility) that is faced in
the synthesis of functional materials and complex organic compounds.

SDLs incorporate ML for modelling and the uncertainty quan-
tification of experimental data or genetic algorithms to efficiently
explore the synthesis design space of molecules or (nano)materials in
a sequential, closed-loop and adaptive manner[40–42]. This critical adap-
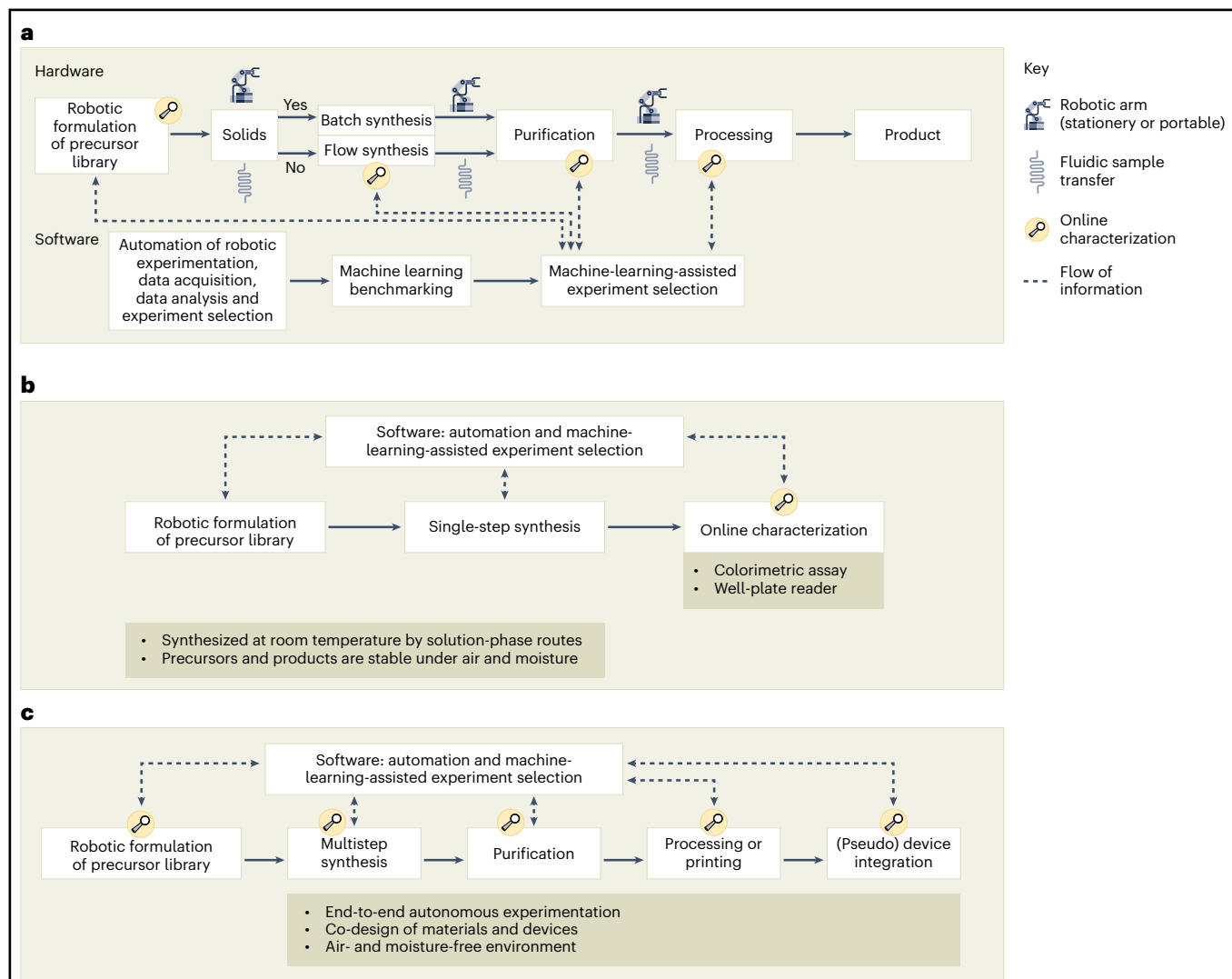tive aspect of autonomous experimentation leverages the uncertainty

Fig. 3 | A roadmap of SDLs. a, An overview of the approach to building an SDL in a chemical and materials science lab. **b**, A flowchart of an SDL built using robotic liquid handling and well-plate-based synthesis integrated with online colorimetric or spectroscopic characterization modules for room-temperature solution-phase chemistries stable under air and moisture. **c**, A flowchart of an end-to-end SDL for the accelerated co-design of clean energy materials and devices under an air- and moisture-free environment.

quantification of data-driven ML models to overcome the limitations of non-adaptive combinatorial screening techniques. Closed-loop formulation–synthesis–structure–property mapping of the targeted class of molecules or materials can be performed by using genetic algorithms or by integrating an ML model (thus improving the model prediction accuracy with every new data point) of single or multiple experimental objectives, for example, reaction yield and regioselectivity or film thickness and manufacturing temperature with uncertainty quantification. The uncertainty quantification of ML models can be utilized to select the next experimental condition by using exploration (design space navigation), exploitation (optimization) or balanced exploration–exploitation decision policies. Existing open-access SDL software packages, which include ChemOS[43] and ARES OS[44], provide a user-friendly starting point for researchers in chemical and materials sciences to initiate an autonomous experimentation. The closed-loop operation of SDLs can be utilized for fundamental studies, for example, to uncover reaction mechanisms, as well as in applied research, for example, the identification of the most sustainable manufacturing route of the target molecule or material. Using ML algorithms that are not properly selected, designed or tuned to achieve a specific objective

of the SDL operation substantially increases the number of closed-loop experimental iterations and, hence, the total cost of experiments[42]. This is why comparing the suitability of ML algorithms for different classes of molecules and (nano)materials[45] by using freely accessible and reproducible data libraries is a vitally important feature of the future developments of SDLs. Providing open-access ML benchmarking resources will be crucial to answer the question 'What ML algorithms can be used for experiment selection and data mining?'.

The answers to questions 'How long does it take to build an SDL?' and 'What would be the cost of building an SDL?' are directly related to the complexity of the required experimental modules (for example, single versus multistage experimental stages), the range and number of operating process conditions (for example, pressure and temperature), type of solvent (aqueous versus organic), required characterization technique(s) and acceptable precision. Building a reliable SDL with a high level of reproducibility for chemistry and materials science labs can take from several weeks to 1–2 years and cost from less than US$1,000 to more than US$1,000,000. For example, the hardware and software requirements of an SDL that performs at room-temperature with a colorimetric or spectroscopic readout (Fig. 3b)[35,46] are different
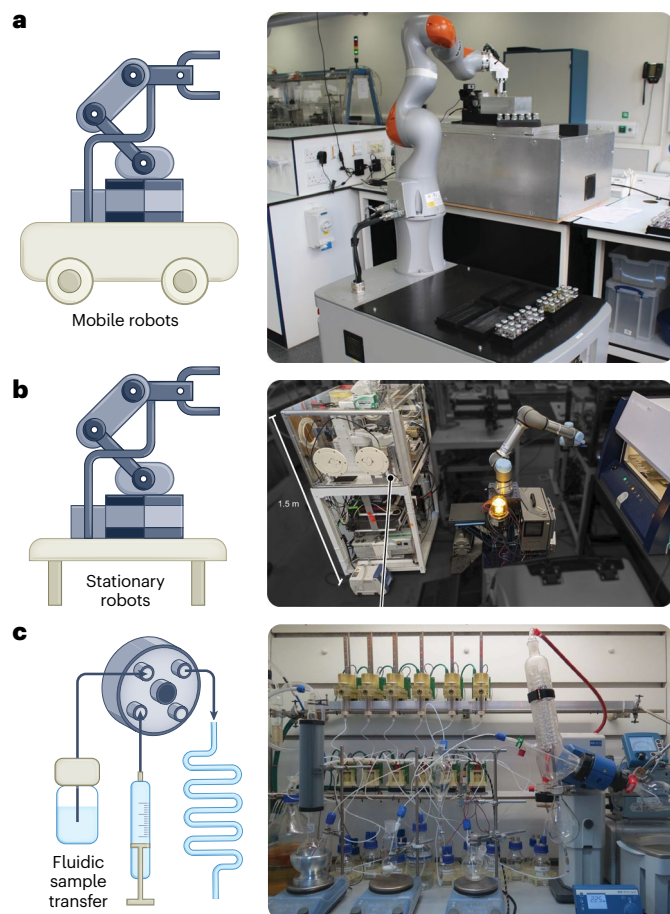
Fig. 4 | The use of robotics in SDLs. a–c, Photographs of mobile[21] (**a**), stationary[30] (**b**) or fluidic (**c**) robots[13] utilized for the automated sample transfer between experimental modules of SDLs. Reprinted with permission from: **a** (right), ref. [21], Springer Nature Ltd; **b** (right), ref. [30], Springer Nature Ltd; **c** (right), ref. [13], AAAS.

to those of an end-to-end autonomous robotic experimentation platform working under an inert atmosphere for the co-design of clean energy materials and devices (Fig. 3c). The hardware and software modularization and standardization of SDLs, along with providing open-access communication protocols with different characterization instrumentations for in situ or online product analysis, can reduce the development timeframe of SDLs from 1–2 years to 1–2 months.

Successful examples of SDLs

Over the past five years, proof-of-concept SDLs—for example, Chemputer[13,20], BEAR[36,47], CAMEO[28] and Artificial Chemist[22,23]— were successfully utilized for the autonomous synthesis of nanoparticles[22–27,32,34], polymers[48] and copolymers[49], thin-film materials[29,50,51], carbon nanotubes[52], supramolecular clusters[53], complex organic molecules[13,19,20,54], photocatalysts[21] and shape-memory materials[28] for applications in additive manufacturing[36], liquid product formulations[55,56], pharmaceuticals[57] and clean energy technologies[58–60]. Figure 4 shows three approaches to the hardware and robotic integration of SDLs: portable robotic arms that access an entire SDL (Fig. 4a)[21] or connect different modules of SDLs[61], stationary robots that supply manufactured parts[36], collected nanomaterial inks[26] or thin film substrates (Fig. 4b)[29,30] to different SDL modules, and compact workstations for tube and/or pump-based reagent transfer between the synthesis and characterization modules of SDLs (Fig. 4c)[13,62]. The unique aspect of mobile robots (Fig. 4a) is the facile access to conventional characterization techniques available in a chemical lab without the need

for a direct integration with the synthesis module of SDLs. Despite this advantage, the high cost of mobile robots that offer a precise and reproducible sample transfer with multiple grippers poses a major bottleneck for such SDLs.

Figure 5 shows examples of parallel batch (Fig. 5a)[25] and flow reactors (Fig. 5b)[22–24,26,34] utilized to automatically perform reactions in SDLs. In the case of organic or nanomaterial synthesis with no solid reagent or precipitation during the synthesis, flow reactors provide an excellent opportunity for reaction miniaturization, reduced chemical consumption and waste generation, facile integration with online characterization techniques and access to synthesis conditions, for example, mixing and heating or cooling rates that are not accessible to batch reactors[19,22–24,32,63]. These advantages of flow reactors make them a promising candidate to access unexplored regions of the design spaces for emerging molecules and (nano)materials. For solid-phase synthesis and processing (for example, preparation of thin films, battery materials or solid-state polymerization), or reactions with the precipitation of solid products or by-products, parallel batch reactors are more suitable reactor candidates for SDLs.

From the characterization perspective, both online and offline modules, such as custom-developed spectroscopy techniques[22–24,32,34] and imaging tools[29,30], and off-the-shelf analytical units, for example, high-performance liquid chromatography, Fourier-transform infrared spectroscopy, NMR spectroscopy and gas chromatography[13,19–21], have been successfully integrated with SDLs for the autonomous synthesis and development of functional materials and molecules. Furthermore, online characterization modules can provide access to measurements after each stage of multistage syntheses or material fabrication. Such intermediate-stage information can be leveraged to accelerate a search through the high-dimensional space of multistage processes by the early identification of more advantageous synthetic routes. The integration of SDLs with online characterization techniques leverages the extensive hardware development and online reaction sampling techniques developed during the past two decades via the emergence and growth of lab-on-a-chip technologies. In addition to common spectral characterization techniques, the structural characterization of fabricated (nano)materials using electron microscopy (transmission electron microscopy and scanning electron microscopy) and small- and wide-angle X-ray scattering can also be integrated with SDLs; however, the high capital cost and the need for additional complex hardware development and integration limit their integration with SDLs to specially dedicated facilities. From the ML perspective, a range of strategies suitable for handling continuous and discrete parameters, from Bayesian optimization to evolutionary algorithms (for example, covariance matrix adaptation evolution strategy and genetic algorithms) have been successfully implemented in SDLs for the accelerated development and on-demand synthesis of organic molecules, nanomaterials and thin-film materials. For details of different ML algorithms utilized in SDLs relevant to chemical and materials sciences, we refer the reader to recent comprehensive reviews of such algorithms[40,64–68].

Current limitations and future opportunities of SDLs

Despite successful proof-of-concept examples of SDLs in the accelerated synthesis of complex organic molecules and advanced (nano)materials, many opportunities exist for further research and development. First and foremost, for non-experts in autonomous robotic experimentation, the transition of SDLs from sophisticated custom-developed technologies to a mainstream approach in experimental chemical and materials sciences requires major advances in hardware development, which include module engineering and online characterization techniques to reduce the entry barriers, such as cost, module assembly, operation and troubleshooting. The high cost of robots and characterization modules, the complicated assembly of custom-developed
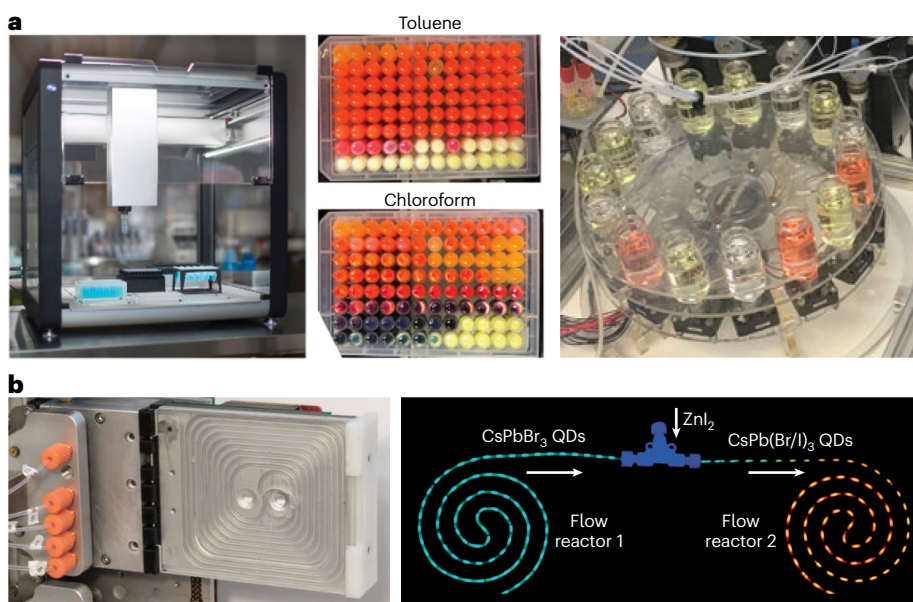
Fig. 5 | Diversity of SDL reactors. a,b, Parallel batch reactors[25,35] (**a**) and miniaturized flow reactors[19,22] (**b**) utilized for the controlled synthesis of thin-film materials (**a**, left), metal nanoparticles (**a**, right), active pharmaceutical ingredients (**b**, left), and colloidal quantum dots, QDs (**b**, right). Adapted with permission from: **a** (left), ref. [35], ACS; **a** (right), ref. [25], Springer Nature Ltd; **b** (left), ref. [19], AAAS; **b** (right), ref. [22], Wiley.

modules and extensive troubleshooting, all combined with the lack of standardization of hardware modules, data flow, data representation and intelligent experiment-selection algorithms, are the current major limitations of SDLs. We see the initial cost barrier of SDLs as an enabling opportunity for the research acceleration community in chemical and materials sciences. The large capital expenditure of current SDLs provides a unique opportunity for researchers interested in hardware development to focus on low-cost and open-source SDL modules, such as liquid-handling robots[69], syringe pumps[70], three-dimensional-printed reactionware[71] and field-deployable diagnostics[72]. Moreover, the recent growth of cloud labs around the world[73] provides another potential avenue for early career researchers to access state-of-the-art robotic experimentation facilities without major capital investments.

The adoption of SDLs by scientists across chemical and materials sciences would entail a highly intelligent and flexible automation of research labs with autonomously reconfigurable experimental modules. The challenge of the autonomous development of advanced functional materials, in contrast to that of small molecules, is the lack of reproducible data in the literature. Although automated data extraction from the literature, despite a proved bias[74], has been achieved for organic synthesis[19,75] and successfully enabled data-driven retrosynthesis or highly accurate reaction prediction, it has largely failed for advanced (nano)materials. This failure, however, creates a unique opportunity for SDLs. The sparse data availability for advanced (nano) materials (for example, clean energy materials), in combination with their lab-to-lab variations, makes SDLs an ideal research platform to provide reproducible data for ML modelling and design space navigation and for knowledge transfer within each class of targeted material. In general, SDLs improve the experimental data reproducibility through digitization, enhanced accuracy, transferrable knowledge and minimization of the impact of human errors.

Although mobile or stationary robotic arms can be utilized for the transfer of liquid-phase reagents or products between different modules or the automatic reconfiguration of SDLs, they are mostly required for SDLs that handle solid-phase reagents, or in cases for which more powerful characterization techniques, for example, NMR spectroscopy, are required. A critical requirement of SDLs working with solid-phase reactions, reagents or samples is the need to use robotics for precise solid-powder dosing and a fast and reliable sample transfer between different SDL modules. Despite the rapid progress of robots and solid-dispensing technologies over the past two decades, the high cost of precise solid-dispensing and robotic arms, with the required precision, reproducibility, mobility and speed, poses a limitation for the widespread implementation of SDLs. Reductions in the costs of solid- and/or liquid-dispensing and stationary and/or mobile robots are enabling factors for the broad deployment and adoption of SDLs across chemical and materials sciences. We believe that a critical next step for SDL adoption is the development of cost-effective mobile robotic manipulators designed to enable flexibility in the automatic reconfiguration of the SDL design and adaptation to dynamic changes in the workspace. Furthermore, robotic manipulators should provide precise and reproducible high-speed operations to maximize the reproducibility and agility of SDLs. A reduced cost of mobile robotic manipulators would enable the incorporation of multiple robots in the SDLs, which would prevent disruption in the closed-loop SDL operation in the case of a potential failure of a specific robot. Such open-access and mobile robotic manipulators will be able to make agile actions in an environment, similar to conventional human-centred research labs, without the need for a special lab space design or modification of the SDL operation.

An important software aspect of SDLs is their robust and flexible integration with ML to provide autonomy for navigation through the design space of molecules and materials. The rapidly growing list of ML modelling and experiment selection strategies makes the algorithm selection a challenging task for non-experts. This challenge is an exciting opportunity for the future development of SDLs towards the standardization of ML algorithms suitable for different end-to-end experimental workflows, operation modes (exploration, exploitation or mechanistic studies) and targeted classes of molecules or (nano) materials (for example, prior knowledge versus physics-based models versus black-box search).

Industry plays an important role in addressing the hardware and software challenges for SDLs by leveraging the prior advancements in the development of experimental tools for combinatorial screening applications in medicinal chemistry and molecular biology. By focusing on cost reduction and the standardization of robots, experimental modules and characterization techniques for SDLs, industry can reduce

the entry barrier to SDLs for scientists. Standard experimental modules and equipment communication protocols are a critically important advancement for future SDLs[39,76]. The main pieces of equipment for the online or in situ characterization of materials or molecules using conventional spectroscopy and chromatography techniques already exist. However, SDLs generally need to use custom-built hardware (for example, a flow cell for the in situ monitoring of reactions performed in a flow reactor) or a triggering method (for example, online gas chromatography sampling) to integrate the existing characterization units with other SDL modules. As the number of SDL users increases, it is expected the companies that manufacture characterization instrumentation, such as spectrometers and chromatographs, as well as NMR spectroscopy, mass spectrometry and X-ray diffraction equipment, will focus on the design and development of sampling and integration modules with open-access software for the in situ and online characterization of materials and molecules. In addition, the leading SDL research groups around the world are strongly encouraged to work with instrumentation companies to expand the available in situ and online characterization modules. A successful example of such an academia–industry collaboration in the advancements of online reaction monitoring modules is the powerful ReactIR probe for integration with flow reactors developed by Mettler Toledo in collaboration with the Ley group at the University of Cambridge[77].

We encourage the ML community in chemical and materials sciences to focus their future efforts on the facile benchmarking of application-specific algorithms[45], expanding open-access databases and making the design space exploration and/or exploitation software user-friendly. Another important aspect of SDLs that is still not well studied is how to carefully choose the best ML algorithm to generate new fundamental knowledge about an underlying phenomenon or an unexpected relationship between input parameters and output properties for the class of reactions or materials explored by the SDL. As the number of experimental modules and independent input parameters of SDLs increases over the next few years, more data- and/or physics-informed ML strategies will be needed to reduce the total cost of computation and experimentation to discover new materials and molecules or the sustainable way to manufacture them at scale[78–80]. Such information can be provided to the SDL either from open-source reaction databases[81], or by ML models that are created using prior data generated by the same or another SDL (for example, the model built on a different subset of materials or reactions from the same general class of materials or reactions)[82]. Data- and/or physics-informed autonomous experimentation is a necessary next step of the SDL's software development to realize their largest impact in the autonomous discovery of materials and molecules. This aspect of future SDLs requires cross-disciplinary training[83] and collaboration between the ML and chemical and materials science communities to enable implementation of the most suitable ML algorithms that are accessible and understandable to non-experts. Such collaborations are necessary to accelerate the intelligent search through the chemical space with constrains, metrics and objectives defined by domain experts.

One of the most intriguing aspects of SDLs, which is largely unexplored and directly tied to the future hardware and software advancements, is their remote operation capabilities through the cloud or remote connection to define the next goal of the SDL operation[27]. Automatic access to a library of starting reagents, in combination with reliable and reproducible automated sample preparation, synthesis and online and offline characterization techniques substantially reduces the required amount of 'in-person' presence of the researcher in the lab during the SDL operations. Furthermore, the remote operation of SDLs in different physical locations provides the unique advantage of reproducible knowledge-sharing (data fusion) opportunities via open databases for different classes of emerging materials and molecules.

We note that the remote operation of SDLs will require different workforce training than that of the current paradigm in chemical

and material sciences. The rapidly emerging remote connectivity tools, such as virtual reality[84] and augmented reality[85], along with digital communication platforms provided stimulating avenues to explore for future SDLs and workforce development during the pandemic and continued thereafter. As SDLs start to penetrate different applications of experimental sciences, one of the major challenges in the next decade will be the required talent pool of a new generation of interdisciplinary trained scientists to utilize SDLs to their full potentials. The need for this new generation of scientists will require us to re-evaluate our student's training and focus on multidisciplinary skills in academia.

References

1. Park, N.-G. & Zhu, K. Scalable fabrication and coating methods for perovskite solar cells and solar modules. *Nat. Rev. Mater.* **5**, 333–350 (2020).
2. Wouters, O. J., McKee, M. & Luyten, J. Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *J. Am. Med. Assoc.* **323**, 844–853 (2020).
3. Helm, D. The Kyoto approach has failed. *Nature* **491**, 663–665 (2012).
4. MacLeod, M., Arp, H. P. H., Tekman, M. B. & Jahnke, A. The global threat from plastic pollution. *Science* **373**, 61–65 (2021).
5. Hanna, R. & Victor, D. G. Marking the decarbonization revolutions. *Nat. Energy* **6**, 568–571 (2021).
6. Gao, J., Yin, Y., Myers, K. R., Lakhani, K. R. & Wang, D. Potentially long-lasting effects of the pandemic on scientists. *Nat. Commun.* **12**, 6188 (2021).
7. Yang, G.-Z. et al. Ten robotics technologies of the year. *Sci. Robot.* **4**, eaaw1826 (2019).
8. MacLeod, B. P., Parlane, F. G. L., Brown, A. K., Hein, J. E. & Berlinguette, C. P. Flexible automation accelerates materials discovery. *Nat. Mater.* **21**, 722–726 (2022).
9. Silver, D. et al. Mastering the game of Go without human knowledge. *Nature* **550**, 354–359 (2017).
10. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
11. Epps, R. W., Volk, A. A., Ibrahim, M. Y. S. & Abolhasani, M. Universal self-driving laboratory for accelerated discovery of materials and molecules. *Chem* **7**, 2541–2545 (2021).
12. Bédard, A.-C. et al. Reconfigurable system for automated optimization of diverse chemical reactions. *Science* **361**, 1220–1225 (2018).
13. Steiner, S. et al. Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, eaav2211 (2019).
14. Tabor, D. P. et al. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **3**, 5–20 (2018).
15. Volk, A. A., Campbell, Z. S., Ibrahim, M. Y. S., Bennett, J. A. & Abolhasani, M. Flow Chemistry: a sustainable voyage through the chemical universe en route to smart manufacturing. *Annu. Rev. Chem. Biomol. Eng.* **13**, 45–72 (2022).
16. Kaminski, T. S. & Garstecki, P. Controlled droplet microfluidic systems for multistep chemical and biological assays. *Chem. Soc. Rev.* **46**, 6210–6226 (2017).
17. Wagner, J. et al. The evolution of materials acceleration platforms: toward the laboratory of the future with AMANDA. *J. Mater. Sci.* **56**, 16422–16446 (2021).
18. Nikolaev, P., Hooper, D., Perea-López, N., Terrones, M. & Maruyama, B. Discovery of wall-selective carbon nanotube growth conditions via automated experimentation. *ACS Nano* **8**, 10214–10222 (2014).

19. Coley, C. W. et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**, eaax1566 (2019).

20. Granda, J. M., Donina, L., Dragone, V., Long, D.-L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **559**, 377–381 (2018).

21. Burger, B. et al. A mobile robotic chemist. *Nature* **583**, 237–241 (2020).

22. Abdel-Latif, K. et al. Self-driven multistep quantum dot synthesis enabled by autonomous robotic experimentation in flow. *Adv. Intell. Syst.* **3**, 2000245 (2021).

23. Epps, R. W. et al. Artificial chemist: an autonomous quantum dot synthesis bot. *Adv. Mater.* **32**, 2001626 (2020).

24. Tao, H. et al. Self-driving platform for metal nanoparticle synthesis: combining microfluidics and machine learning. *Adv. Funct. Mater.* **31**, 2106725 (2021).

25. Salley, D. et al. A nanomaterials discovery robot for the Darwinian evolution of shape programmable gold nanoparticles. *Nat. Commun.* **11**, 2771 (2020).

26. Li, J. et al. Autonomous discovery of optically active chiral inorganic perovskite nanocrystals through an intelligent cloud lab. *Nat. Commun.* **11**, 2046 (2020).

27. Li, J., Tu, Y., Liu, R., Lu, Y. & Zhu, X. Toward 'on-demand' materials synthesis and scientific discovery through intelligent robots. *Adv. Sci.* **7**, 1901957 (2020).

28. Kusne, A. G. et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* **11**, 5966 (2020).

29. MacLeod, B. P. et al. Self-driving laboratory for accelerated discovery of thin-film materials. *Sci. Adv.* **6**, eaaz8867 (2020).

30. MacLeod, B. P. et al. A self-driving laboratory advances the Pareto front for material properties. *Nat. Commun.* **13**, 995 (2022).

31. Bateni, F. et al. Autonomous nanocrystal doping by self-driving fluidic micro-processors. *Adv. Intell. Syst.* **4**, 2200017 (2022).

32. Vikram, A., Brudnak, K., Zahid, A., Shim, M. & Kenis, P. J. A. Accelerated screening of colloidal nanocrystals using artificial neural network-assisted autonomous flow reactor technology. *Nanoscale* **13**, 17028–17039 (2021).

33. Bezinge, L., Maceiczyk, R. M., Lignos, I., Kovalenko, M. V. & deMello, A. J. Pick a color MARIA: adaptive sampling enables the rapid identification of complex perovskite nanocrystal compositions with defined emission characteristics. *ACS Appl. Mater. Interfaces* **10**, 18869–18878 (2018).

34. Mekki-Berrada, F. et al. Two-step machine learning enables optimized nanoparticle synthesis. *npj Comput. Mater.* **7**, 55 (2021).

35. Higgins, K., Ziatdinov, M., Kalinin, S. V. & Ahmadi, M. High-throughput study of antisolvents on the stability of multicomponent metal halide perovskites through robotics-based synthesis and machine learning approaches. *J. Am. Chem. Soc.* **143**, 19945–19955 (2021).

36. Gongora, A. E. et al. A Bayesian experimental autonomous researcher for mechanical design. *Sci. Adv.* **6**, eaaz1708 (2020).

37. Liu, Z. et al. Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing. *Joule* **6**, 834–849 (2022).

38. Bai, J. et al. From platform to knowledge graph: evolution of laboratory automation. *J. Am. Chem. Soc. Au* **2**, 292–309 (2022).

39. Seifrid, M. et al. Autonomous chemical experiments: challenges and perspectives on establishing a self-driving lab. *Acc. Chem. Res.* **55**, 2454–2466 (2022).

40. Gromski, P. S., Henson, A. B., Granda, J. M. & Cronin, L. How to explore chemical space using algorithms and automation. *Nat. Rev. Chem.* **3**, 119–128 (2019).

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

42. Epps, R. W., Volk, A. A., Reyes, K. G. & Abolhasani, M. Accelerated AI development for autonomous materials synthesis in flow. *Chem. Sci.* **12**, 6025–6036 (2021).

43. Roch, L. M. et al. ChemOS: an orchestration software to democratize autonomous discovery. *PLoS ONE* **15**, e0229862 (2020).

44. Deneault, J. R. et al. Toward autonomous additive manufacturing: Bayesian optimization on a 3D printer. *MRS Bull.* **46**, 566–575 (2021).

45. Liang, Q. et al. Benchmarking the performance of Bayesian optimization across multiple experimental materials science domains. *npj Comput. Mater.* **7**, 188 (2021).

46. Vaddi, K., Chiang, H. T. & Pozzo, L. D. Autonomous retrosynthesis of gold nanoparticles via spectral shape matching. *Digital Discov.* **1**, 502–510 (2022).

47. Gongora, A. E. et al. Using simulation to accelerate autonomous experimentation: a case study using mechanics. *iScience* **24**, 102262 (2021).

48. Salley, D. S., Keenan, G. A., Long, D.-L., Bell, N. L. & Cronin, L. A modular programmable inorganic cluster discovery robot for the discovery and synthesis of polyoxometalates. *ACS Cent. Sci.* **6**, 1587–1593 (2020).

49. Reis, M. et al. Machine-learning-guided discovery of ^{19}F MRI agents enabled by automated copolymer synthesis. *J. Am. Chem. Soc.* **143**, 17677–17689 (2021).

50. Langner, S. et al. Beyond ternary OPV: high-throughput experimentation and self-driving laboratories optimize multicomponent systems. *Adv. Mater.* **32**, 1907801 (2020).

51. Li, Z. et al. Robot-accelerated perovskite investigation and discovery. *Chem. Mater.* **32**, 5650–5663 (2020).

52. Nikolaev, P. et al. Autonomy in materials research: a case study in carbon nanotube growth. *npj Comput. Mater.* **2**, 16031 (2016).

53. Porwol, L. et al. An autonomous chemical robot discovers the rules of inorganic coordination chemistry without prior knowledge. *Angew. Chem. Int. Ed.* **59**, 11256–11261 (2020).

54. Schweidtmann, A. M. et al. Machine learning meets continuous flow chemistry: automated optimization towards the Pareto front of multiple objectives. *Chem. Eng. J.* **352**, 277–282 (2018).

55. Grizou, J., Points, L. J., Sharma, A. & Cronin, L. A curious formulation robot enables the discovery of a novel protocell behavior. *Sci. Adv.* **6**, eaay4237 (2020).

56. Cao, L. et al. Optimization of formulations using robotic experiments driven by machine learning DoE. *Cell Rep. Phys. Sci.* **2**, 100295 (2021).

57. Sagmeister, P. et al. Autonomous multi-step and multi-objective optimization facilitated by real-time process analytics. *Adv. Sci.* **9**, 2105547 (2022).

58. Zhao, Y. et al. Discovery of temperature-induced stability reversal in perovskites using high-throughput robotic learning. *Nat. Commun.* **12**, 2191 (2021).

59. Du, X. et al. Elucidating the full potential of OPV materials utilizing a high-throughput robot-based platform and machine learning. *Joule* **5**, 495–506 (2021).

60. Sun, S. et al. Accelerated development of perovskite-inspired materials via high-throughput synthesis and machine-learning diagnosis. *Joule* **3**, 1437–1451 (2019).

61. Nambiar, A. M. K. et al. Bayesian optimization of computer-proposed multistep synthetic routes on an automated robotic flow platform. *ACS Cent. Sci.* https://doi.org/10.1021/acscentsci.2c00207 (2022).

62. Li, S. et al. Using automated synthesis to understand the role of side chains on molecular charge transport. *Nat. Commun.* **13**, 2102 (2022).

63. Volk, A. A. & Abolhasani, M. Autonomous flow reactors for discovery and invention. *Trends Chem.* **3**, 519–522 (2021).

64. Pollice, R. et al. Data-driven strategies for accelerated materials design. *Acc. Chem. Res.* **54**, 849–860 (2021).
65. Epps, R. W. & Abolhasani, M. Modern nanoscience: convergence of AI, robotics, and colloidal synthesis. *Appl. Phys. Rev.* **8**, 041316 (2021).
66. Li, J. et al. AI applications through the whole life cycle of material discovery. *Matter* **3**, 393–432 (2020).
67. Tao, H. et al. Nanoparticle synthesis assisted by machine learning. *Nat. Rev. Mater.* **6**, 701–716 (2021).
68. Yano, J. et al. The case for data science in experimental chemistry: examples and recommendations. *Nat. Rev. Chem.* **6**, 357–370 (2022).
69. Saar, L. et al. The LEGOLAS Kit: A low-cost robot science kit for education with symbolic regression for hypothesis discovery and validation. *MRS Bull.* **47**, 881–885 (2022).
70. Baas, S. & Saggiomo, V. Ender3 3D printer kit transformed into open, programmable syringe pump set. *HardwareX* **10**, e00219 (2021).
71. Hou, W. et al. Automatic generation of 3D-printed reactionware for chemical synthesis digitization using ChemSCAD. *ACS Cent. Sci.* **7**, 212–218 (2021).
72. Koydemir, H. C. & Ozcan, A. Smartphone-based sensors and imaging devices for global health. *Adv. Opt. Technol.* **10**, 87–88 (2021).
73. Arnold, C. Cloud labs: where robots do the research. *Nature* **606**, 612–613 (2022).
74. Beker, W. et al. Machine learning may sometimes simply capture literature popularity trends: a case study of heterocyclic Suzuki–Miyaura coupling. *J. Am. Chem. Soc.* **144**, 4819–4827 (2022).
75. Coley, C. W., Green, W. H. & Jensen, K. F. Machine learning in computer-aided synthesis planning. *Acc. Chem. Res.* **51**, 1281–1289 (2018).
76. Gao, W., Raghavan, P. & Coley, C. W. Autonomous platforms for data-driven organic synthesis. *Nat. Commun.* **13**, 1075 (2022).
77. Carter, C. F. et al. ReactIR flow cell: a new analytical tool for continuous flow chemical processing. *Org. Process Res. Dev.* **14**, 393–404 (2010).
78. Correa-Baena, J.-P. et al. Accelerating materials development via automation, machine learning, and high-performance computing. *Joule* **2**, 1410–1420 (2018).
79. Ahmadi, M., Ziatdinov, M., Zhou, Y., Lass, E. A. & Kalinin, S. V. Machine learning for high-throughput experimental exploration of metal halide perovskites. *Joule* **5**, 2797–2822 (2021).
80. Sun, S. et al. A data fusion approach to optimize compositional stability of halide perovskites. *Matter* **4**, 1305–1322 (2021).
81. Kearnes, S. M. et al. The open reaction database. *J. Am. Chem. Soc.* **143**, 18820–18826 (2021).
82. Gongora, A. E. et al. Designing lattices for impact protection using transfer learning. *Matter* **5**, 2829–2846 (2022).
83. Sun, S., Brown, K. & Kusne, A. G. Teaching machine learning to materials scientists: lessons from hosting tutorials and competitions. *Matter* **5**, 1620–1622 (2022).
84. Skibba, R. Virtual reality comes of age. *Nature* **553**, 402–404 (2018).
85. Matthews, D. Virtual-reality applications give science a new dimension. *Nature* **557**, 127–128 (2018).

Acknowledgements

Author contributions
M.A. wrote the original draft and M.A. and E.K.G. reviewed and edited the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence should be addressed to Milad Abolhasani.

Peer review information *Nature Synthesis* thanks Mahshid Ahmadi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Peter Seavill, in collaboration with the *Nature Synthesis* team.

Reprints and permissions information is available at www.nature.com/reprints.