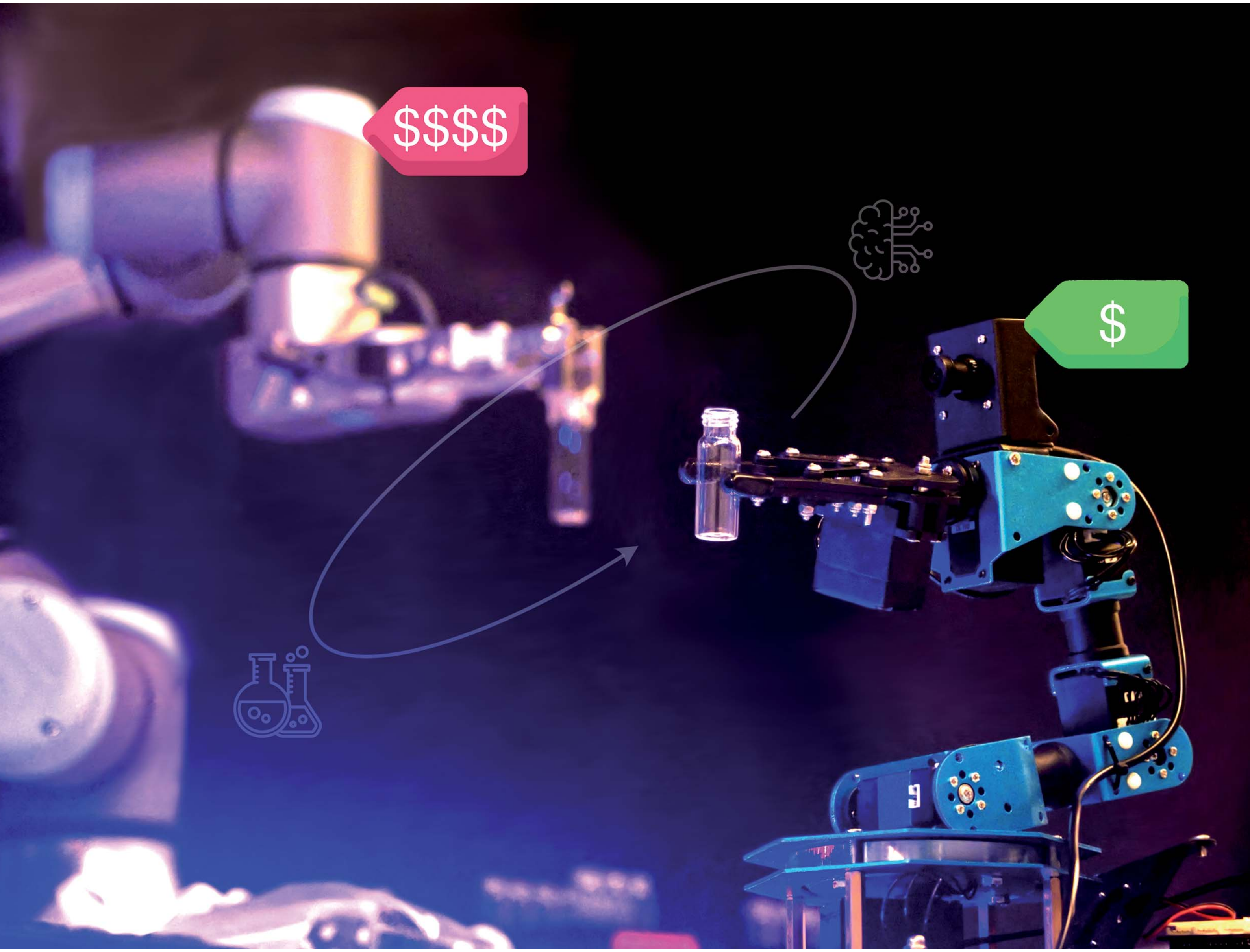
Volume 3
Number 5
May 2024
Pages 833-1072

Digital Discovery

rsc.li/digitaldiscovery



ISSN 2635-098X

ROYAL SOCIETY OF CHEMISTRY

TUTORIAL REVIEW

Stanley Lo, Sterling G. Baird, Taylor D. Sparks, Alán Aspuru-Guzik *et al.*

Review of low-cost self-driving laboratories in chemistry and materials science: the "frugal twin" concept

Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

TUTORIAL REVIEW

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Digital Discovery*, 2024, **3**, 842

Review of low-cost self-driving laboratories in chemistry and materials science: the "frugal twin" concept

Stanley Lo, †*[a] Sterling G. Baird, †*[bc] Joshua Schrier, [d] Ben Blaiszik, [eg] Nessa Carson, [f] Ian Foster, [eg] Andrés Aguilar-Granda, [h] Sergei V. Kalinin, [io] Benji Maruyama,[j] Maria Politi, [k] Helen Tran, [abl] Taylor D. Sparks †*[cm] and Alán Aspuru-Guzik †*[abln]

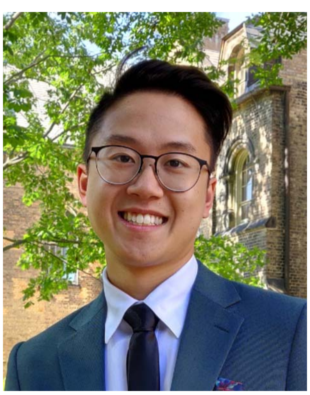
Received 15th November 2023
Accepted 13th February 2024

DOI: 10.1039/d3dd00223c

rsc.li/digitaldiscovery

This review proposes the concept of a "frugal twin," similar to a digital twin, but for physical experiments. Frugal twins range from simple toy examples to low-cost surrogates of high-cost research systems. For example, a color-mixing self-driving laboratory (SDL) can serve as a low-cost version of a costly multi-step chemical discovery SDL. Frugal twins already provide hands-on experience for SDLs with low costs and low risks. They can also offer as test beds for software prototyping (*e.g.*, optimization, data infrastructure), and a low barrier to entry for democratizing SDLs. However, there is room for improvement. The true value of frugal twins can be realized in three core areas. Firstly, hardware and software modularity; secondly, purpose-built design (human-inspired *vs.* hardware-centric *vs.* human-in-the-loop); and thirdly state-of-the-art (SOTA) software (*e.g.*, multi-fidelity optimization). We also

[a]Department of Chemistry, University of Toronto, 80 St George St, Toronto, ON M5S 3H6, Canada. E-mail: stanley.lo@mail.utoronto.ca; alan@aspuru.com
[b]Acceleration Consortium, University of Toronto, 80 St George St, Toronto, ON M5S 3H6, Canada. E-mail: sterling.baird@utoronto.ca
[c]Department of Materials Science and Engineering, University of Utah, Salt Lake City, UT 84108, USA. E-mail: sparks@eng.utah.edu
[d]Department of Chemistry and Biochemistry, Fordham University, The Bronx, New York 10458, USA
[e]University of Chicago, Chicago, IL 60637, USA
[f]Early Chemical Development, Pharmaceutical Sciences, R&D, AstraZeneca, Macclesfield SK10 2NA, UK
[g]Argonne National Laboratory, Lemont, IL 60439, USA
[h]Facultad de Química, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510 Ciudad de México, Mexico
[i]Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN 37916, USA
[j]Air Force Research Laboratory, Materials and Manufacturing Directorate, Wright-Patterson AFB, OH, 45433, USA
[k]Department of Chemical Engineering, University of Washington, Seattle, WA, USA
[l]Department of Chemical Engineering, University of Toronto, 80 St George St, Toronto, ON M5S 3H6, Canada
[m]Chemistry Department, University of Liverpool, Liverpool L7 3NY, UK
[n]Department of Computer Science, University of Toronto, 40 St George St, Toronto, ON M5S 2E4, Canada
[o]Physical Sciences Division, Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99354, USA
† These authors contributed equally to this work.



Stanley Lo

Stanley Lo obtained his BSc degree in Chemistry at the University of Toronto in 2022. Prior to starting graduate studies, he spent 4 months at Kebotix Inc. developing graph neural network architecture for molecular property prediction. Currently, he is a doctoral student under the joint supervision of Alán Aspuru-Guzik and Helen Tran, focusing on the automation of polymer post-functionalization for depolymerization.

Sterling G. Baird

Sterling G. Baird is the Director of Training and Programs at the Acceleration Consortium within the University of Toronto. He received his BSc in Applied Physics and his MSc in Mechanical Engineering at Brigham Young University. He completed his PhD in Materials Science & Engineering at the University of Utah. Sterling is accelerating materials discovery through advanced Bayesian optimization, self-driving laboratories, and educational platforms.

describe the ethical benefits and risks that come with the democratization of science through frugal twins. For future work, we suggest ideas for new frugal twins, SDL educational course outcomes, and a classification scheme for autonomy levels.

1 Introduction

Self-driving laboratories (SDLs) are autonomous experiment-performing systems that have the potential to accelerate the discovery of solutions for key societal needs such as carbon-neutral/net-zero processes, food and agriculture, fuels, clean energy, energy storage, drug discovery, and structural materials.[1] SDLs can improve experimental reproducibility[2] and increase researcher productivity by automating tedious, repetitive tasks. They require scientists to learn new skills relating to the supervision, modification, and maintenance of autonomous systems, at both the hardware (*e.g.*, liquid handlers, robotic arms) and software (*e.g.*, optimization algorithms, workflow orchestration, data infrastructure) levels. This concept allows scientists to focus on higher-level cognitive tasks such as hypothesis formulation, experimental design, and data interpretation, which are not easily automated.[3]

The concept of accelerated discovery *via* automation goes by several names, including SDLs,[3,4] materials acceleration platforms,[9] Lab 4.0,[10–12] Internet of Laboratory Things,[13–15] Robot Scientists,[16] the Autonomous Research System (ARES),[17] and autonomous experimentation systems.[18] While each term has its own nuances, here we use the term SDL exclusively and interpret it as referring to autonomous research systems used to accelerate materials discovery without human intervention. It is important to note that for the rest of the article, automation refers to the use of technology to perform tasks with minimal human intervention, while autonomy implies the ability of a system to operate independently, making decisions and taking actions without human control.

SDLs that are used to solve societal challenges are considered to be materials acceleration for societal solutions (MASS) platforms.[1] Such platforms need to be widely deployed and adopted if societal challenges are to be addressed. However, such "critical MASS" (in the words of Seifrid *et al.*[1]) will require lower costs, enhanced ability to reconfigure and expand, and a joint effort to make available easy-to-understand examples and systems for more advanced research tasks. Since the introduction of the concept of an artificial intelligence system to laboratory automation in 1985 by Isenhour,[19] the development of SDLs has gained traction. However, there are only a handful of fully autonomous low-cost SDLs reported in the literature. Stach *et al.*[18] provide a community perspective on SDLs in the context of academia, industry, government laboratories, and funding agencies, and supply a descriptive table of selected SDLs across a variety of applications including chemical vapor deposition,[20] nanocrystals,[21] flow-[22] and vial-based[23] chemistry, oil-in-water emulsions,[24] additive manufacturing,[25] thin films,[7] quantum materials,[26] and solid-state materials.[27] Many review and perspective articles have already been contributed to the field,[1,3,4,9,18,28–46] and a list of 25 recent low-cost SDLs is given in Table 1.

What sets our review apart from others is that we explicitly focus on low-cost SDLs, *i.e.*, frugal twins of high-cost SDLs. We hope that this attention to the importance of low-cost SDLs will shift perspectives on the educational and research capabilities of low-cost systems and provide a common reference point for building new solutions.

The question of what is low- *vs.* high-cost is both a subjective and contextual problem. Monetary cost and space constraints are particularly apparent in educational settings, as indicated by the large fraction of educational SDLs specifically described as low-cost, under 1000 USD,[47–49] and which occupy relatively small footprints. This is in part because the final objectives are

Taylor D. Sparks

Dr Sparks, Associate Professor of Materials Science and Engineering at the University of Utah, completed a sabbatical at the University of Liverpool, supported by the Royal Society Wolfson Visiting Fellow program. He holds a BS in MSE from UofU, MS from UCSB, and PhD in Applied Physics from Harvard. He received the NSF CAREER Award, spoke at TEDxSaltLakeCity, and actively participates in MRS, TMS, and ACERS. He's an Associate Editor for Computational Materials Science and Data in Brief journals. Beyond academia, he runs the "Materialism" podcast, creates educational YouTube content, and enjoys canyoneering in southern Utah with his four children.

Alán Aspuru-Guzik

Alán Aspuru-Guzik is a professor of Chemistry and Computer Science at the University of Toronto and is also the Canada 150 Research Chair in Theoretical Chemistry and a Canada CIFAR AI Chair at the Vector Institute. He is a CIFAR Lebovic Fellow in the Biologically Inspired Solar Energy program. Alán also holds a Google Industrial Research Chair in Quantum Computing. Alán is the director of the Acceleration Consortium, a University of Toronto-based strategic initiative that aims to gather researchers from industry, government and academia around pre-competitive research topics related to the lab of the future.

Table 1 Low-cost SDL summary

Name	Field	Purpose	Cost[a]	Ref.
Educational ARES	Education	3D printing	300	63
Additive manufacturing ARES	Mat. Sci.	3D printing	1000	17
Pioreactor for real-time … culture measurements	Biology	Cell growth	250	64
Autonomous Research System (ARES)	Mat. Sci.	CNT growth	5000	20
Closed-loop Spectroscopy Lab: Light-mixing	Education	Color opt.	50	65
Bayesian Optimization Bartender (BOB)	Education	Color opt.	200	66
Accelerate Synthesis of MOFs	Mat. Sci.	Crystallinity	830	67
Evolution of oil droplets …	Chemistry	Evolution	1000	68
A … robot for discovering … protocell behavior	Chemistry	Evolution	1000	24
… a configurable 3D printed fluidic platform	Chemistry	Evolution	2000	69
A microfluidic platform [for] chemical evolution	Chemistry	Evolution	5000	70
Chemical synthesis robot for nanomaterials	Mat. Sci.	Morphology	15 000	71
Cheap automated synthesis platform	Chemistry	Organic synth.	450	72
Networking chemical robots	Chemistry	Organic synth.	500	73
Autonomous … platform for … synthesis	Chemistry	Organic synth.	10 000	74
"The Chemputer"	Chemistry	Organic synth.	30 000	75
3D printed [microfluidic] autonomous analyzer …	Chemistry	Photometry	2050	76
High-Throughput [CdSe Nanocrystal Synthesis]	Chemistry	Quantum dots	2000	59
Crystallization Robot	Mat. Sci.	Randomness	3000	77
Scientific Inquiry in Middle Schools	Education	Titration	250	47
LEGO Low-cost Autonomous Science (LEGOLAS)	Education	Titration	300	48
Autonomous titration for chemistry classrooms	Education	Titration	600	78
Automated pH Adjustment …	Education	Titration	650	79
Automatic titrator for intro chemistry labs	Education	Titration	934	80
Automatic titration for teaching chemistry	Education	Titration	4160	81

[a] Estimated costs in USD. Abbreviations: carbon nanotube (CNT); additive manufacturing (AM); Autonomous Research System (ARES); LEGO Low-cost Autonomous Science (LEGOLAS); Bayesian Optimization Bartender (BOB); metal–organic framework (MOF).

often based on learning outcomes rather than specific research objectives.

In both contexts, there is a range between monetary costs that can be covered by business-as-usual "spare" monetary resources *vs.* costs that require dedicated support from grants and other funding sources. For example, the National Science Foundation currently places a threshold of 5000 USD to differentiate between consumables and equipment, above which a purchase must be "adequately justified" on a grant proposal. An example such as the Opentrons OT-2 platform (∼7500 USD starting cost) likely fits more clearly into the "dedicated support" category for many education-oriented systems and somewhere in-between "spare resources" and "dedicated support" for research tasks. Nevertheless, the context depends on a multitude of other factors including the specific research group, institution, country, and socioeconomic status. For example, the monetary amount a research group in a developed country considers low-cost will be significantly higher than what a local school in a developing country would consider low-cost due to practical reasons such as but not limited to lower amounts of funding, greater costs for delivery, unfair pricing, difficulty of foreign exchange, and priority to secure a livelihood.[50–52]

With an emphasis on chemistry and materials science applications and as part of a broader focus on MAPs and MASS, we walk through topics relevant to low-cost SDLs. First, we describe the development of "frugal twins" that capture the core principles of real-world systems at an education-friendly cost, and present areas where the community benefits from low-cost twins (Section 2). Next, we delineate how educational outcomes and autonomy can equip the next generation of scientists with industry-relevant skills (Section 3). Afterwards, we detail how modularity for hardware and software plays an important role in reducing redesign costs for future systems (Section 4.1). We also illustrate how using a hardware-centric approach when developing SDLs can reduce system complexity by leveraging existing hardware in unconventional ways in comparison to other design approaches (Section 4.2). Next, we highlight how discovery can be accelerated further through high-throughput and parallelized systems (Section 4.3.1). With the growth of cloud infrastructure, we show that cloud experimentation (similar to cloud computing, but for experiments) decentralizes hardware, computing, and domain expertise, reducing the barrier-of-entry for SDLs and enabling robust and efficient batch optimization (Section 4.3.4). Finally, we describe ideas for new frugal twins, suggest potential SDL course outcomes, and discuss how to classify autonomy levels in SDLs (Section 6). To encourage a continuing discussion, we also provide a list of public, community-driven discussions (Section 7).

2 What are frugal twins, and why do we need them?

Inspired by the digital twin, a virtual counterpart of a physical entity, we introduce the concept of the frugal twin, a low-cost counterpart of a physical entity.[53] A digital twin is designed for simulation, modelling, and evaluation, and can offer

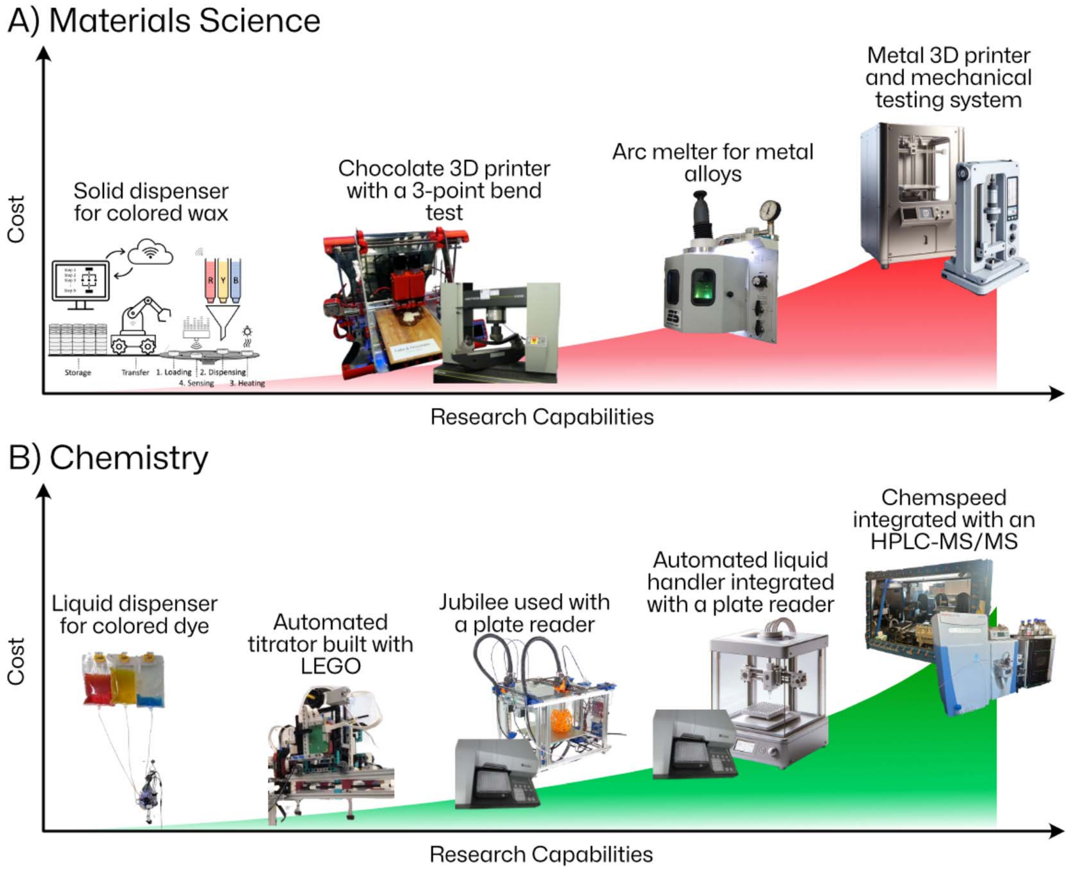


Fig. 1 Spectrum of frugal twin capability *vs.* cost trade-off. (A) From left to right: solid dispenser for colored wax,[55] chocolate 3D printer (*)[56] with a 3-point bend test,[57] arc melter (*),[58] metal 3D printer (**) and mechanical testing system for metals (**). (B) From left to right: liquid handling for dye mixing,[55] automated titrator built from LEGO,[48] Jubilee sonochemical synthesis platform used with a plate reader (*) for absorbance and fluorescence measurements,[59–61] automated liquid handler (**) integrated directly with a plate reader (*),[61,62] and Chemspeed integrated with an HPLC-MS/MS. Images marked with (*) were reproduced with permission under the Creative Commons Attribution license (CC-BY). (**) Marked images were rendered using ChatGPT 4.0.

insights into the physical entity, either before its inception[53] or during its lifetime.[54] Likewise, a low-cost SDL can serve as a frugal twin of a high-cost SDL. Frugal twins present a low-risk environment for rapid prototyping and a new educational platform which can offer insights into the high-cost entity.

Any frugal twin of an SDL is located within a trade-off spectrum between cost and research capabilities, with the balance between two factors determining its usefulness for particular education and research activities (Section 2.1). We show in Fig. 1 some illustrative examples of these trade-offs for materials science and chemistry, and a list in Table 1 of various low-cost SDLs.

2.1 Trade-offs between cost and capabilities

There are two primary ways to reduce costs when creating a frugal twin: scale back research capabilities, or reduce accuracy and precision. The appropriate balance between cost and capability will typically be governed by available resources, and necessary functions to perform the desired task. We illustrate in Fig. 1 possible trade-offs in the context of two experiments: one in materials science and one in chemistry. Although some of the examples shown in the figure are not standalone SDLs, each could be integrated into an SDL for various research purposes.

In the materials science experiment, the high-cost capability is to 3D print various metal alloys at extremely high temperatures, as can be accomplished, for example, by a metal 3D printer. As cost decreases, the capabilities of frugal twins stray further away from the high-cost capabilities (Fig. 1). The arc melter can form alloys at high temperatures, but cannot 3D print them. The next drop in cost renders the instrument only capable of toy problems: the 3D chocolate printer can form and 3D print various chocolate compositions. Lastly, the "Hello World" of a materials science SDL, at the lowest cost shown, is the solid dispenser for colored wax, capable of producing candle wax in customized colours.[55]

Likewise in the chemistry context, the high-cost capability of multi-step, multi-batch synthesis and characterization can be accomplished by a Chemspeed integrated with high-performance liquid chromatography coupled with mass spectrometry (HPLC-MS). At a significantly lower cost, the Opentrons OT-2 platform can perform single-step, multi-batch synthesis and limited characterization techniques using an integrated plate reader, focused primarily on biological applications.[62,82]

The next lowest in cost is the Jubilee system which can be adapted to perform sonochemical synthesis and used with an offline plate reader.[59,60] The automated titrator built from LEGO, one step lower in cost than the Jubilee, can no longer perform synthesis but only multi-batch liquid dispensing, and uses a pH probe for characterization.[48] Lastly, the cheapest SDL is a liquid handler for dye mixing, tasked with obtaining a customized color as characterized by a light sensor.[65,66,73,83,84]

We note that it may not always be possible to create a useful frugal twin for an SDL. For example, a large part of cutting-edge research relies on expensive analytical instrumentation to be able to obtain sufficient information about experiments. In the context of compound characterization, instruments such as nuclear magnetic resonance spectroscopy (NMR) and HPLC-MS apparatus can cost hundreds of thousands of dollars to acquire and operate. However, infrared radiation can be a cheaper alternative to expensive analytical techniques like the ones mentioned before for tasks such as in-line reaction monitoring.[85,86] This can be sufficient for low-fidelity reaction monitoring but is incapable of unknown compound characterization. Sacrificing research capabilities for lower costs is sometimes infeasible depending on the task at hand. To perform robust unknown compound characterization, low-cost ($\leq$10 000 USD) alternatives to NMR or HPLC-MS do not currently exist on the market.

Analogous to the trade-off between cost and research capabilities, there can be a trade-off between throughput and fidelity.[38] For example, a benchtop NMR is lower cost ($\geq$40 000 USD)[87] and easily adapted to flow chemistry SDLs, but sacrifices measurement precision and accuracy. The cost/benefit analysis must consider the expected speedup in the rate of progress for a lower fidelity analysis tool and the cost from potential inaccuracies compared to the gold standard analysis tools.

2.2 Rapid, low-risk prototyping and proofs of concept for research

SDLs are feats of both science and engineering which are typically both complex and expensive such that rapid prototyping is challenging. As a result, there is often a gap between state-of-the-art (SOTA) technologies and technologies found in current SDLs. Typically, researchers building SDLs risk the "jack of all trades, master of none" effect relative to more traditional researchers in terms of scientific research, hardware, and software advancements. Oftentimes, one or more of these components are sacrificed and/or large and diverse teams are required to build the SDL in an appropriate time-frame. This is where frugal twins can close the gap between SOTA technologies and high-cost SDLs. Frugal twins can enable researchers to easily prototype and engage in an iterative loop to explore new design concepts, gain new knowledge, refine and validate existing designs, and easily share information within a group of researchers.[89] This relaxed requirement prototyping approach[89] leverages trade-offs between accuracy and cost. As an example, advanced optimization algorithms for SDLs can be integrated and tested on the frugal twin of an SDL. In principle, any of the three components (scientific objective, hardware, or software) can have a relaxed requirement to accelerate the prototyping of the other components.

Preliminary evidence acquired from a low-cost SDL can serve as a proof of concept for solving an analogous research problem that can then justify the funding for a more capable high-cost SDL. The low-cost SDL may have lower accuracy and reliability, but still provide evidence of feasibility for the proposed research, as well as answering some relevant research questions. In addition, the low-cost SDL can act as a proxy for estimating the acceleration factor that an SDL can offer in comparison to manual experimentation.

An example that compellingly captures how a frugal twin can promote rapid prototyping and teach transferable skills to students in a low-risk setting is the MIK-I, a frugal twin of "The Machine".[72,88] The initial goal for researchers is to build "The Machine".[88] However, prior to assembling this SOTA research tool, they built MIK-I with approximately 450 USD (Fig. 2), the main purpose of MIK-I's creation being to familiarize the researchers with automated synthesis platforms. The frugal twin is designed to handle liquids of different physicochemical properties such as density, viscosity, and surface tension. However, when building MIK-I, liquid handling became an issue because the pumps needed to be calibrated differently for each liquid in the system. This problem gave students hands-on experience with an issue that would also occur with the SOTA research tool, which would allow them to solve the eventual problem more readily. To evaluate the scope of MIK-I, the researchers successfully performed C–C bond formation reactions widely used in organic chemistry such as the Claisen–Schmidt condensations, Suzuki–Miyaura coupling,

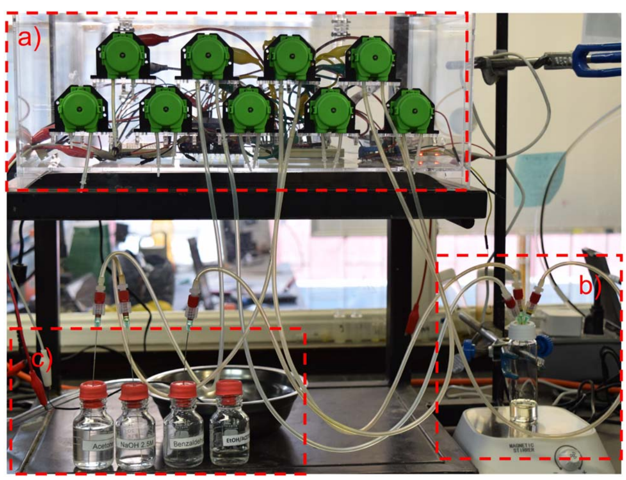


Fig. 2 MIK-I, a low-cost automated synthesis workflow platform. (a) Peristaltic pumps controlled by a Raspberry Pi, (b) synthesis reactor, (c) reagent bottles.[72]

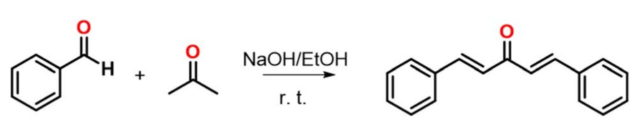


Fig. 3 The scheme for a general crossed aldol condensation reaction as a proof of concept.[72]

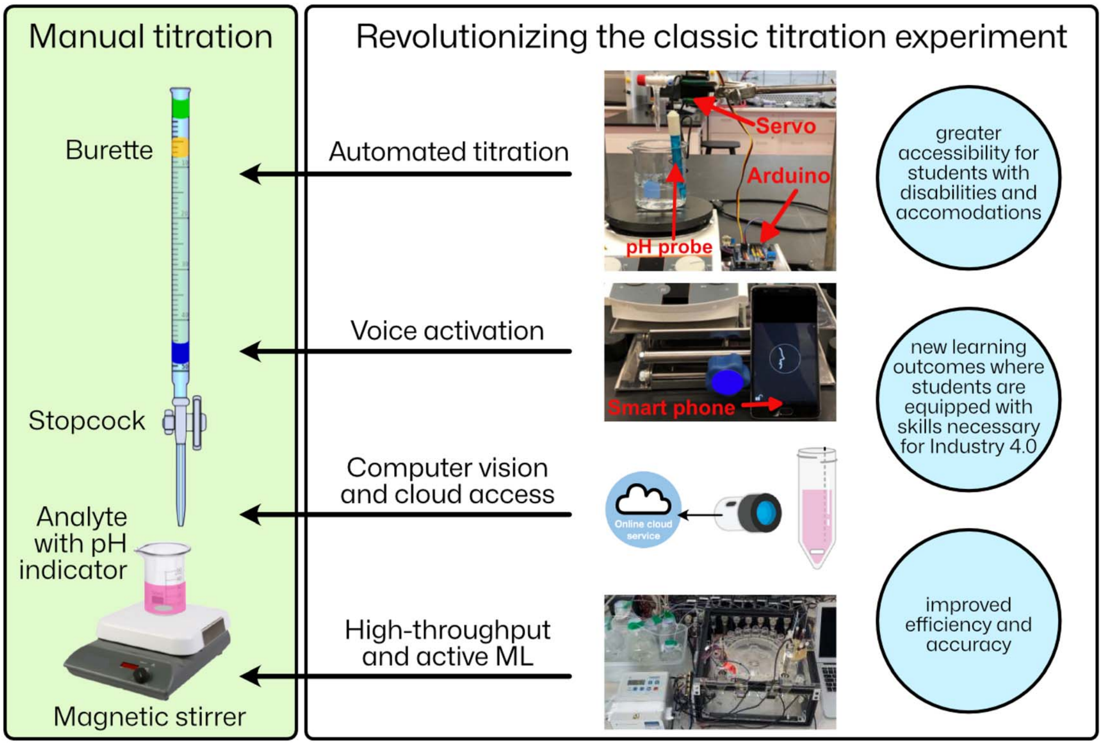


Fig. 4 Example of a titration setup that can be equipped with automation, voice activation, computer vision, high-throughput capabilities, and machine learning. Adapted with permission from ref. 80, 81, 90 and 91. Copyright 2016, 2019, 2021 American Chemical Society. Adapted with permission from ref. 79 under the Creative Commons Attribution license (CC-BY). Copyright Elsevier 2022.

Knoevenagel condensations, and Morita–Baylis–Hillman reactions in an automated fashion (Fig. 3).[72]

2.3 Education of future workforce will be critical for self-driving labs

New skills, including AI, autonomy, and complex data analysis will be required to design, build and operate SDLs. SOTA SDLs can have a high training burden and potentially high cost from mistakes. The frugal twin can provide a potential solution to these problems by enabling new users to gain transferable skills for the SOTA SDLs in a low-cost, low-risk setting. Low-cost SDLs create an environment conducive to experiential learning *via* trial and error, which acts as a stepping stone for new users with limited robotics and programming experience. Furthermore, by making SDLs affordable and easier to access, barriers to entry to citizen scientists will be reduced, which enables a wider group of citizens, both in terms of quantity and diversity, to partake in the pursuit of scientific research. This feat requires overcoming both financial and technical barriers, by providing detailed schematics, parts lists, assembly instructions, code documentation, and troubleshooting guides.

3 How are frugal twins being used in education and research?

In this section, we offer an in-depth overview of low-cost SDLs in materials science and chemistry designed for education (Section 3.1) and research (Section 3.2). From these examples, there are many lessons to be learned and areas to be improved, which are later discussed in Section 4.

3.1 Designed for education

Two pertinent educational topics are examples of automated/autonomous titration setups (Section 3.1.1) and minimal working examples of SDLs (Section 3.1.2).

3.1.1 Titration. Titrations are a common experiment type in high school and undergraduate chemistry curricula where students determine the concentration of an unknown solution by adding a titrant, a solution with a known concentration. In acid–base titrations, the pH of an unknown solution is determined by quantitatively adding a titrant (acid or base) while monitoring the pH using an indicator or detector (refer to Fig. 4 for a visualization).[89] The automation of a titrator allows many students, including those with certain disabilities who may otherwise be excluded, to further their understanding of chemistry, while simultaneously providing an opportunity to learn about electronics and robotics[90] (Fig. 4). A variety of features can be incorporated around an automated titrator, such as a web interface for remote work, a liquid (acid/base) dispenser using a solenoid valve or peristaltic pump, a pH probe for characterization, a pH indicator with computer vision, voice activation *via* digital assistants such as Siri, and a LEGO framework for modularity and high-throughput.[48,80,81,90,91]

A programmable titrator can also support a variety of other educational tasks. Students can be tasked with developing their own automation methods for this previously manual procedure, a problem that is engaging, encourages critical thinking and provides additional opportunities for learning. Typically, students develop their own heuristics, such as adding large amounts of titrant at the start of the experiment and slowly reducing the addition of titrant until the endpoint is reached, with the goal of optimizing for efficiency and accuracy. An automated titrator can accelerate the pace at which students can quantify and test multiple titration strategies for optimal efficiency and/or accuracy.[78]

The applicability of skills acquired from educational settings to research and industry settings is critical,[92] and modification of a titration experiment presents a direct example of this transferability. For instance, Pomberger *et al.*[79] designed their titration apparatus with high-throughput batch samples, and active machine learning (ML) to model the pH response of multi-buffered polyprotic systems, a challenging yet important task for many chemical labs and industrial plants. For context, educational titration setups with a single-buffered system like those mentioned above can be accurately described by the Henderson–Hasselbalch equation;[79] however, this does not hold for multi-buffered polyprotic systems.[79] Although the multi-buffered polyprotic problem has greater complexity, students can learn to adapt solutions to fit their needs and work around the limitations. By exploiting the benefits of modularity (outlined in Section 4.1), students can choose from several optimization algorithms such as ML, proportional-integral-derivative control, and model predictive control.[79] Although automated solutions improve efficiency and robustness, an educational apparatus should also provide the option for a student to be put back in the loop (*i.e.*, manual mode) because it can provide the student with more direct interactions with the hardware.

3.1.2 Color-matching. Another straightforward demo for SDLs is color-matching, where the goal is to find the optimal mixture of a set of colors (*e.g.*, primary colors) that will mix to produce a target color. The concept is low-cost and straightforward and has been demonstrated for both light-mixing[65,83] and liquid-mixing examples.[66,93]

For the light-mixing example, Baird and Sparks[65] developed a system known as Closed-loop Spectroscopy Lab: Light-mixing (CLSLab:Light) as a teaching and prototyping platform that entails mixing the light from red, green, and blue light-emitting diodes (LEDs) (Fig. 5). The demo employs light rather than matter while retaining the principles of SDLs. Taking language from the software community, it is a "minimal working example" of an SDL. The primary benefits of this device relative to more costly, time-intensive, higher-footprint (and, of course, more chemistry-relevant) liquid handlers such as Opentrons OT-2,[82] Sidekick,[94] evoBOT,[95] OpenLH,[96] OTTO,[97] and OpenWorkstation[98] are that it costs under 100 USD, requires less than an hour of setup time, takes up minimal desk space, and does not require chemical consumables. While CLSLab:Light cannot provide experimental data directly relevant to materials discovery, its features make it a prime candidate for classroom settings, allowing each student or team to obtain hands-on experience. Additionally, the platform can be used to prototype concepts such as creating a network of geographically distant experiments and implementing advanced optimization topics such as batch (Section 4.3.1) and multi-fidelity optimization (Section 4.3.2). Over a dozen tutorials and examples for basic optimization, advanced optimization, device communication, and data ecosystems are given in the Closed-loop Spectroscopy Lab documentation.

CLSLab:Light has also evolved as an example and suggestion of SDL best practices. The software is modular, and open-source. Build instructions[83] and a video build tutorial are provided, with parts lists designed to be modular and robust to supply chain issues. Additional features of the CLSLab:Light platform that helps students to learn and implement best practices are summarized in Table 2.

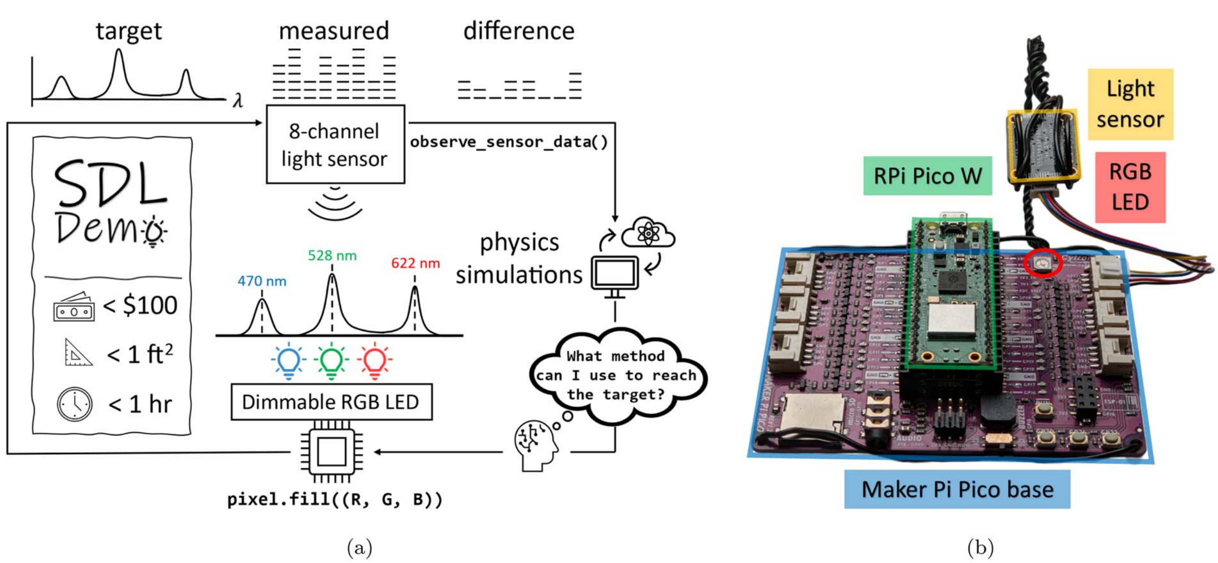


Fig. 5 The CLSLab:Light demo. (a) A summary schematic of CLSLab:Light. (b) An annotated image of the CLSLab:Light. (c) Was adapted with permission from ref. 65. Copyright Elsevier 2022.

Table 2 Summary of best practice topics (Topic) that address development pain points (Pain point). Related resources/tools (Resources) and corresponding implementations in the CLSLab:Light framework (CLSLab:Light) are also given. In other words, the Resources column links directly to the tools while the CLSLab:Light column typically links to various places in https://github.com/sparks-baird/self-driving-lab-demo

Topic	Pain point	Resources	CLSLab:Light
Version control	Keep detailed, accessible, and efficient snapshots of your code at any point in time	Git, GitHub	GitHub repo/history
Project generator	Streamline setting up modular code for a new project while conforming to best practices	PyScaffold, cookiecutter-pypackage	PyScaffold and initial commit
Python packages	Make installation and setup easier for users	PyPI (pip), Anaconda	PyPI *via* setup.cfg
Unit tests	Catch bugs and ensure functionality	pytest	Tests folder
Continuous integration	Regularly and automatically validate code, run tests, and publish new versions	GitHub actions	Actions *via* ci.yml
Secure wireless communication	Safely communicate within and between software and hardware	MQTT	MQTT[a] host/client
Data management	Store data that is "Findable, Accessible, Interoperable, Reusable" (FAIR)	MongoDB, SQL	MongoDB[a] main.py
Installation-free notebook tutorials	Make it easy for users to learn, test, and adapt the functionality	Google Colab, Binder	Tutorials page
Documentation web host	Host a website with your documentation for free	Readthedocs, GitHub pages	Readthedocs site
Documentation builder	Package your documentation, tutorials, and API as web-friendly HTML files	Sphinx, Jekyll	Source files, conf.py

[a] Detailed setup instructions for MQTT and MongoDB are provided in Baird and Sparks.[83]

Baird and Sparks[83] have explored the commercialization of CLSLab:Light as an at-cost kit, with two successful rounds of crowdfunding *via* the GroupGets platform (see Campaign #1112 and Campaign #1129), totalling 39 kits; many kits have already been used in classroom settings at the University of Toronto, Massachusetts Institute of Technology, and University of Chicago. For continuing discussion related to packaging open-source hardware as commercial kits, see Discussion #124.

CLSLab:Light has already seen success, but domain-specific communities (biology, chemistry, solid-state materials science) will benefit from their own minimal working examples. Baird and Sparks[83] have explored extensions that adapt the instructive

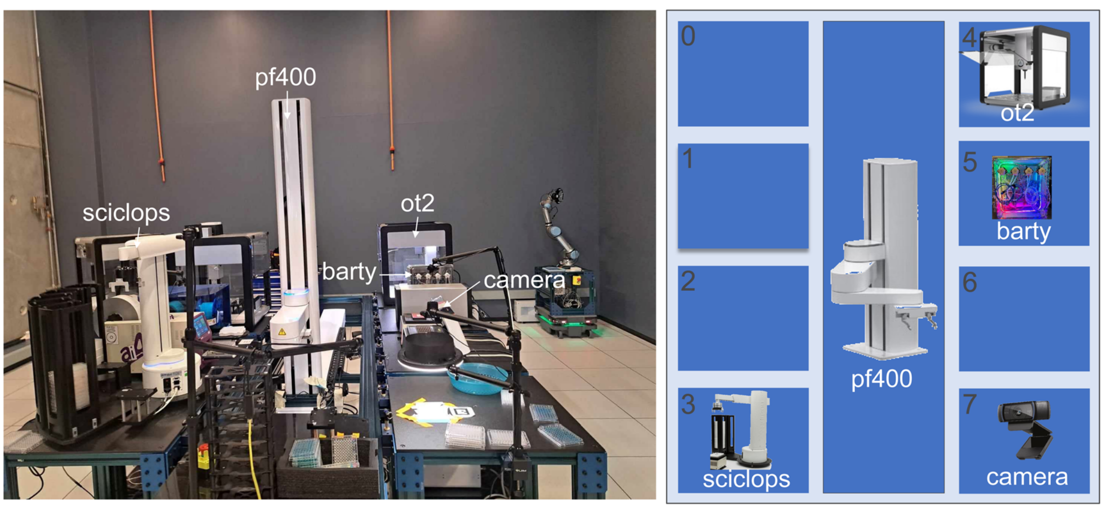


Fig. 6 A photograph and diagram of the robotic work cell (indicated by each blue box) used for a WEI-based color mixing experiment. The Sciclops picks up a 96-well plate from its plate storage towers and transfers it to its exchange location. The PF400 then transfers the plate to the Opentrons OT-2, which mixes the three target colors. When the liquid reservoirs in the system are empty, the custom robot, Barty, refills them by using peristaltic pumps. Once mixing is completed, the plate is transferred to the camera location to be imaged. The plate is then looped between the camera and the Opentrons OT-2 until the experiment is over. The empty work cells (*i.e.* blue boxes) provide additional space for the robotic platform to expand its capabilities, showcasing modular design. Reprinted from Ginsburg *et al.*[84]

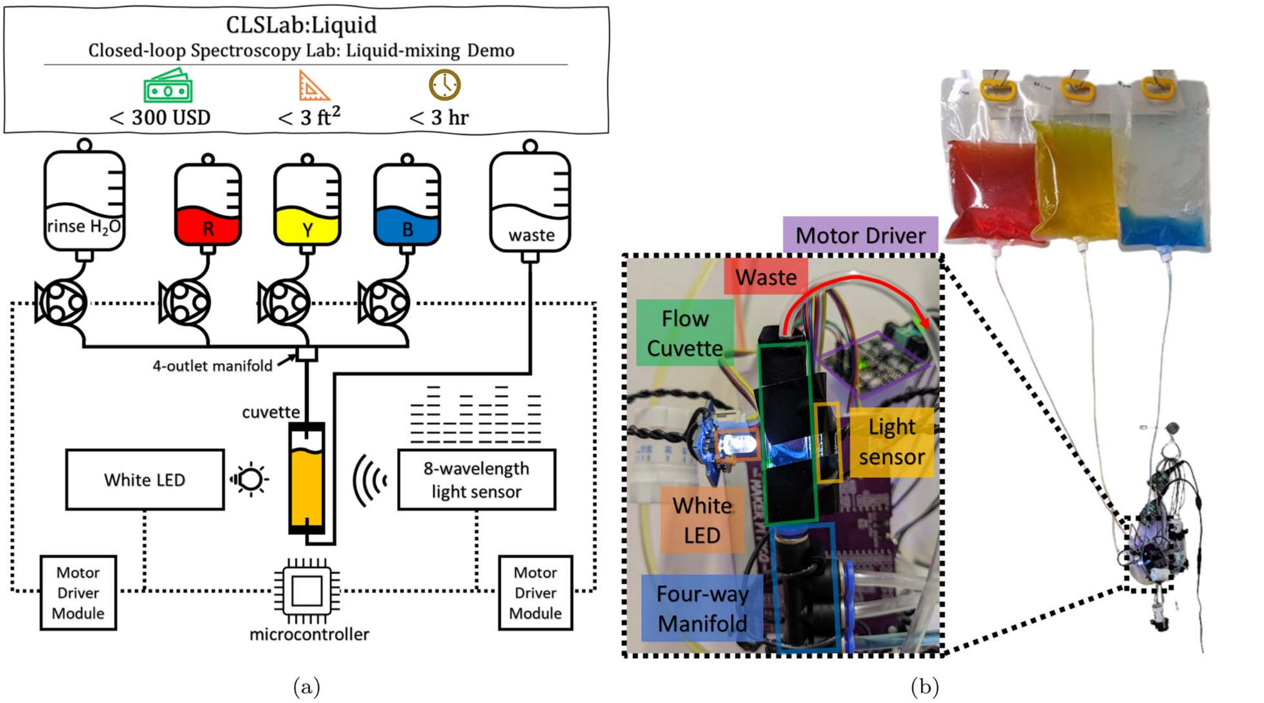


Fig. 7 The CLSLab:Liquid demo. (a) A summary schematic of CLSLab:Liquid. (b) An annotated image of the CLSLab:Liquid.[55]

lessons from CLSLab:Light to other domains. For example, using the modular software and hardware components, Baird and Sparks[83] extend the platform to a liquid-based color-matching task (Closed-loop Spectroscopy Lab: Liquid-mixing (CLSLab:Liquid)) which uses the prototypical example of mixing red, yellow, and blue food coloring dyes (Fig. 7).

The inherent simplicity of the color matching application as demonstrated in CLSLab:Light has also inspired others to employ it in other settings. For example, Ginsburg *et al.*[84] have implemented a color matching application in the context of their workcell execution interface science factory architecture.[93] It is designed with modular instrument interfaces and workflow specifications used to implement an application that connects an Opentrons OT-2 liquid handler, liquid replenishment robot, and camera station (see Fig. 6). The Globus platform is employed[45] to link optimization algorithms running on remote computers and to publish results to a remote data portal.

In sharp contrast to chemistry applications, low-cost examples of SDLs for solid-state materials science are effectively non-existent. To address this gap, an idea for a solid-state materials science extension involving the melting and mixing of colored wax powders is described in Section 6.1.

3.2 Designed for research

Typically, low-cost setups are not regarded as research tools because of their lack of accuracy, precision, and capabilities. However, many research groups are developing low-cost SDLs for reasons such as full control over the end-to-end design (Section 3.2.1), and ease of parallelization (Section 3.2.2). For example, the Sidekick liquid dispenser[94] was designed around the liquid dispensing requirements associated with automated exploratory synthesis of halide perovskites,[99] and only later used for teaching an introductory chemistry laboratory on automation.[100] Similarly, the Jubilee system[59] was originally demonstrated in the context of nanocrystal synthesis research,[101] and later used for education.[60]

3.2.1 End-to-end design. Instead of purchasing expensive and inflexible commercial systems to produce an SDL, building a low-cost SDL from scratch gives the researcher full control over the system. This concept of building a complete system from beginning to end is referred to as end-to-end design. Salley *et al.*[102] demonstrate this process through several examples over the last decade. With the wider availability of 3D printers and low-cost development kits, growing supply chains, better tutorials, and greater access to internet of things in the last two decades, custom scientific apparatus can be built at low costs. However, although low-cost electronic and hardware components offer a wide range of unique capabilities compared to fully developed systems, they generally require significant time and effort to design, engineer, and test.

Nevertheless, with a specific, unique, and focused research problem, Gutierrez *et al.*[68] take advantage of the full control over the end-to-end design of a novel, custom-built chemorobotic platform. This system is capable of exploring a diverse range of oil-droplet formulations which was designed to improve the understanding of evolutionary dynamics. Many low-cost components such as a RepRap 3D printer, camera, Arduino microcontroller, and 3D printed parts are used to gain the desired functionality for this specific experimental task.[68] Later, this robot was redesigned with a 3D printed arena for droplet mixing which could be easily transformed into different environments, adding a new independent variable to experimentation.[69] With high-throughput experimentation and automation, it is not crucial for the robot to be extremely accurate or precise, due to the ease of performing multiple replicates to reduce the uncertainty of results. In this oil-droplet system, several replicates are performed and the uncertainty of each measurement is accounted for before drawing conclusions from general trends.[69] Full control over the design of the experimental apparatus is invaluable for niche research problems.

The modular Geneva wheel platform engineered by Salley *et al.*[103,104] is another example of a low-cost SDL designed end-to-end to leverage the advantages of low-cost components and custom parts. The Geneva wheel platform is capable of rotating 24 reactors using the Geneva mechanism and a stepper motor which enables it to run 24 parallel reactions. For every rotation, the necessary reagents are dispensed serially into a reactor with peristaltic pumps. Each reactor has a magnetic stirring module which stirs 24 reactions in parallel. In addition, the sampling and cleaning modules can move in x, y, and z directions along the platform frame which enables in-line measurements, sample extraction, transfer between vials, and cleaning of reactors to prevent cross-contamination.[103] Due to its modular nature, it can be easily reconfigured for the synthesis of gold nanoparticles, polyoxometalates, or other coordination compounds.[71,103–106] From this system, an important takeaway is that "automation can only be so cheap before significant frustration is experienced".[102] In this example, Salley *et al.*[102] replace cheap aquarium pumps with motor-controlled stepper pumps, which offer better control and accuracy over liquid dispensing while still remaining affordable.

Although the "Chemputer" is not as low-cost as our other considerations, it is worth mentioning because of its end-to-end design for universal chemical synthesis. The Chemputer not only has custom 3D printed parts and low-level electronic components such as syringe pumps, but also interfaces with existing chemistry instruments that may already be in the lab such as hotplates, photoreactors, flow reactors, a rotary evaporator, benchtop NMR spectrometers, and in-line spectrometers (UV-Vis, infrared spectroscopy (IR) and electrospray ionization-MS) to perform organic synthesis and characterization.[107–116] Given its wide range of research capabilities, the "Chemputer" can cost over 30 000 USD with a setup time of 1 week. Manzano *et al.*[74] develop the "mini-Chemputer," which reduces the barrier of entry from 30 000 USD to 10 000 USD, and 1 week to 1 day of reported setup time. Having full control over the end-to-end design of this system enabled the Cronin group to develop both the Chemputer, and the low-cost, portable mini-Chemputer.

Another example of end-to-end design is the Jubilee platform created by Vasquez *et al.*[60] at the University of Washington.[60] Originally, Jubilee was designed for multi-tool fabrication tasks and more. Some examples of its intended application ranged from multi-head 3D printing to multi-pen plotting, and simple liquid handling through syringes. Jubilee presents a modular tool-changing design that accommodates user-created tools and beds (Fig. 8a).[60] Politi *et al.*[59] have demonstrated the use of this versatile, multi-tool platform configured for automated ultrasound application (Fig. 8c), along with an Opentrons OT-2 liquid-handling robot and a well-plate spectrometer for the synthesis of CdSe nanocrystals. In this example, the authors were able to test 625 unique sample conditions, in triplicate, in less than two months, ensuring repeatability and reducing uncertainty on the results. The components to build the Jubilee platform can be individually sourced from readily available and 3D printed materials or even purchased as a kit, for a total cost of ≤2000 USD. Furthermore, the project is fully open-hardware and open-source, resulting in a series of resources, from build instructions to an active Discord channel for informal communication, and requires no previous building skills, which significantly lowers the barrier to its implementation in materials research spaces. No modification of the off-the-shelf, commercially available sonicator was required and simple electronics allowed for instrument interfacing. There is currently no commercially available solution for automating single-point sonochemical processing, making this example a great demonstration of how SOTA technology can be easily democratized through "maker skills" (3D design and fabrication, electronics, and programming) and cheaper electronics. While successful, the study by Politi *et al.*[59] relied on three different instruments to conduct the workflow. It is however possible to integrate all the synthesis, processing, and characterization tools onto the same Jubilee platform, given its automatic tool-changing capabilities, creating a closed-loop experimental system. Finally, it should also be noted that systems like Jubilee, which originated from the digital fabrication space, might require additional hardening and possible small materials adjustments before they can be fully trusted as science tools (Fig. 8).

3.2.2 Ease of parallelization. With lower costs per duplicate of the system, several duplicates can be linked together for a high degree of parallelization offering benefits of decentralization, high-throughput, and batch optimization. Caramelli *et al.*[73] built a network of robots from a series of simple chemical robots that use several peristaltic pumps for liquid handling, a glass reaction vial, a webcam for reaction analysis, and a pcDuino board for electronic control. Due to its simplicity and low cost, the hardware is easily replicated, which enables parallelization of experiments. The following experiments described below exploit some of the advantages of building a network of robots: collaborative azo dye chemical space exploration, real-time control of an oscillating reaction, a reproducibility assessment of inorganic cluster crystallization, and gameplay-driven chemical discovery.[73]

First, the robots were able to communicate by uploading results to the cloud and screening for results from other robots *via* Twitter. This system prevents robots from duplicating others' reactions and allows them to explore more efficiently as a team. Using a network connection, multiple physically separated robots can be synchronized in real time. Caramelli *et al.*[73] use a chemical oscillator based on the Belousov–Zhabotinsky (BZ) reaction to showcase real-time control performance. The oscillation period is synchronized in real time between robots with an uncertainty of 2 s.

Reproducibility in the context of parallelization is necessary for accurate data acquisition. In one experiment, the network of

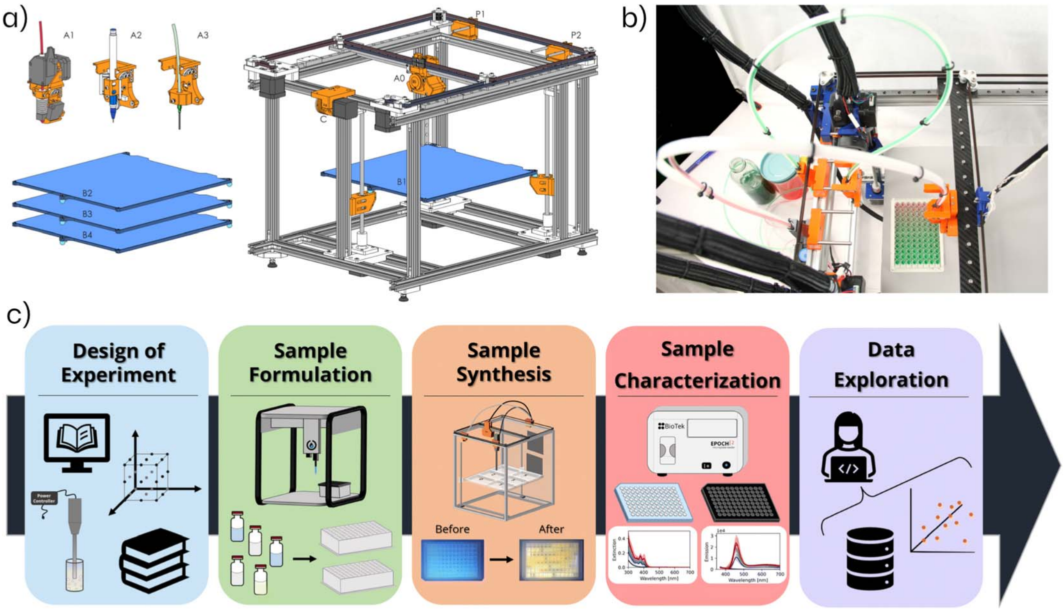


Fig. 8 (a) The blueprint design of the Jubilee system which can equip modular multi-headed tools. (b) Example of the Jubilee system dispensing liquids into a 96-well plate. (c) The workflow of adapting Jubilee into the automated sonochemical synthesis of nanocrystals. Adapted from ref. 60 with written permission from the authors under the Creative Commons Attribution license (CC-BY). Adapted from ref. 59 with permission from the Royal Society of Chemistry.

robots collaboratively explored the conditions for the crystallization of tungsten POM clusters. Crystallization is a stochastic process, which makes it challenging to determine its ideal conditions, particularly on small scale. Nevertheless, the network of robots found six sets of conditions that offered reproducibility between 11.8 and 50%, which may be deemed acceptable for a stochastic process on a small scale.

Lastly, success in gameplaying offers the insight that large amounts of data enabled by powerful computation can push ML models to reach superhuman performance.[117] Highly robust and reproducible materials chemistry SDLs can generate large amounts of data with low-cost experimentation and parallelization. Caramelli *et al.*[73] demonstrated that two robots can compete against each other in a well-defined game to discover novel colors in the context of an azo coupling reaction. The rules are simple: novel results are rewarded, and common results are punished. Each time that a loser emerges at the completion of a game, the loser can change strategies by redefining their reaction space. The goal of the gamification of such an experiment is for the model to develop an optimal strategy to maximize the objective without human guidance. The success of this simple experiment provides the groundwork for similar SDLs to solve more complex problems through a low-cost and parallelized approach.

4 How do we make frugal twins better?

We describe ways to continue improving and leveraging the strengths of frugal twins in terms of hardware and software modularity (Section 4.1), human-inspired *vs.* hardware-centric *vs.* human-in-the-loop design approaches (Section 4.2), and synergizing frugal twins with SOTA software tools and algorithms (Section 4.3).

4.1 Modularity

Modularity refers to the assembly of a cohesive system or device that has discrete, self-contained modules which can be easily interconnected and replaced. Each module performs a specific function or task, and they can be combined or modified independently. This approach allows for flexibility, scalability, and ease of maintenance, as well as facilitating the reuse of components in different applications. In this section, we explore modularity in the context of both low-cost hardware (Section 4.1.1) and open-source software (Section 4.1.2).

4.1.1 Hardware. MacLeod *et al.*[36] emphasize "the characteristic features of modern robots that make them useful for flexible automation [which] include large working areas, many degrees of freedom, high positioning accuracy and repeatability, intrinsic safety, and easy programming. Versatile multiaxis robots that can interact with both liquids and solids offer the flexibility to automate a wide range of experiments" (Fig. 9). Although low-cost SDLs cannot generally afford such characteristics, the emphasis is on leveraging cost-effective and creative strategies to automate a diverse range of experiments within their limitations. Gutierrez *et al.*[68] demonstrated their use of modular design for simple reconfiguration where parts can be easily redesigned, replaced, and tested. Their oil-water droplet robot can be readily reconfigured for adding new chemicals and other formulation-based studies in a variety of simple ways.[24,68,70] For example, the 3D printed polypropylene evolutionary arena can be interchanged with different designs

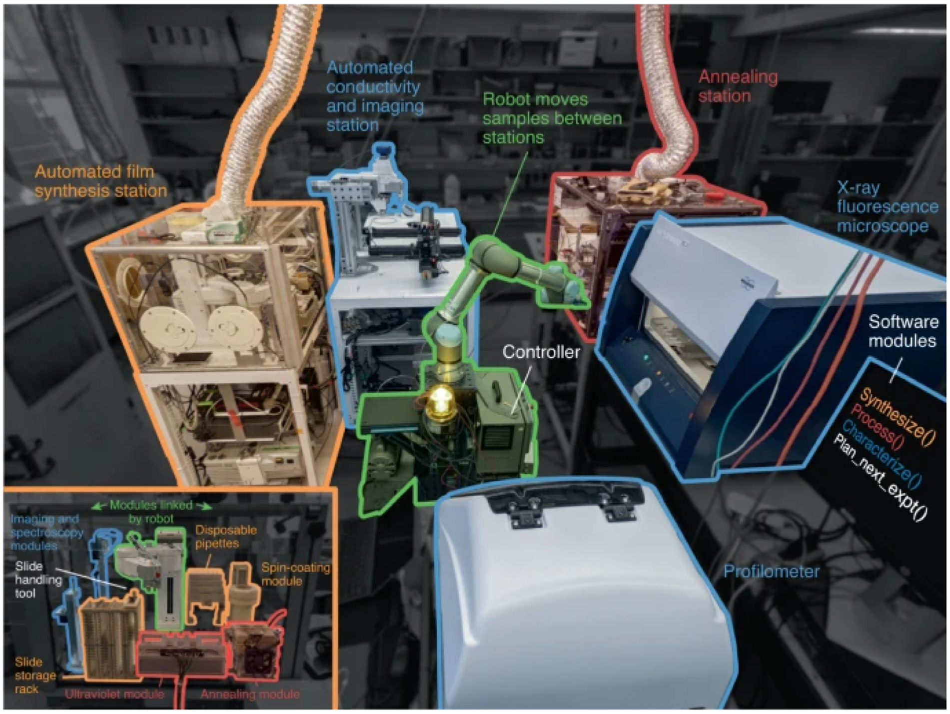


Fig. 9 Flexible automation for SDLs where components can be rearranged and replaced around the central robot arm. From L to R (main image): automated film synthesis station, automated conductivity and imaging station, robot moving samples between stations (center; controller labeled in white), profilometer (front), annealing station (back), X-ray fluorescence microscope, software modules. From L to R (inset image): imaging and spectroscopy modules (back), slide storage rack (front), modules linked by robot (slide handling tool labeled in white), ultraviolet module (front), disposable pipettes (back), spin-coating module (back), annealing module (front). MacLeod *et al.*, flexible automation accelerates materials discovery, *Nature Materials*, published 2022, Copyright © 2021, Springer Nature Limited.

that have pillars, caves, or other arrangements.[69] The well-plate array for sample preparation can also be switched with a Geneva wheel that automates drying and cleaning, increasing experimental throughput.[24] Another flexible concept for simple reconfigurations is Reactionware, which refers to low-cost 3D printable reactors for custom reactions and volumes.[113,118,119]

Given that devices inevitably break down at times, incorporating modularity into SDLs reduces the time and cost of maintenance. If one component breaks, then only that small portion of the instrument needs to be repaired or replaced. In addition, with smaller modular parts, debugging is simplified since each individual component can be tested separately, quickly determining the points of failure.

An SDL should be composed of a core infrastructure capable of interchangeably adapting to domain-specific requirements such as but not limited to liquid handling, solid dispensing, and thin-film manufacturing. This is more cost-effective than building a fixed, domain-specific system capable of performing all the desired tasks for only one given type of experimentation. After the first discovery campaign is completed, the cost of redesigning an inflexible SDL for further work could be much higher than for a modular system. To reduce the redesign cost for future systems, we need to incorporate modularity at the early conception stage of building any SDL.

Sometimes even small design choices can provide significant advantages and flexibility for an automation platform. In this context, the Jubilee[60] platform is a great example of hardware modularity. In fact, the platform was designed in an application-agnostic fashion where tools can be interchangeably loaded on the platform, which can then automatically pick them up and return them after their task is complete. All of this is accomplished through a locking mechanism that allows the tool to lock onto the central carriage and a tool template pattern which ensures constant tool location. Another advantage of Jubilee is its ability to host not only simple sample transfer tools, such as a liquid handling pipette or syringe, but also tools for processing or manipulation and subsequent characterization such as a sonicator.[59] This is not possible with commercially available liquid-handling robotic platforms, which can only complete a limited set of tasks before the labware needs to be moved onto a different automation instrumentation. The flexibility of Jubilee, in fact, allows for rapid reconfiguration of the platform for various applications, such as the nanomaterials synthesis shown by Politi *et al.*[59]

4.1.2 Software. While existing efforts to enforce SDL hardware modularity are valuable, in practice, it is still in its infancy. Some lessons can be taken from modern software development, such as functional and object-oriented programming (*i.e.*, organized use of functions and classes), the single responsibility principle (each module has a single, well-defined responsibility), and related concepts like version control (semantic versioning, commit history, backups, and rolling back to previous versions). These principles are applied out of necessity to optimization and workflow orchestration software ecosystems with large user bases such as Meta's Adaptive Experimentation (Ax) Platform (**https://ax.dev/**) and Agnostic's Covalent workflow orchestration platform (**https://www.covalent.xyz/**).

In some scenarios, software development best practices have been applied to chemistry and materials informatics optimization and workflow orchestration packages. As a set of computer instructions (codebase) evolves and matures, it often involves organizing lines of code into distinct blocks (functions) that perform specific tasks, and then further organizing these blocks into categories or groups (classes and modules) to create a more structured and manageable system.

A practical example of this is Gryffin,[120] a Bayesian optimization tool that supports continuous and categorical variables, physicochemical descriptors, and batch optimization. Gryffin is written in Python and uses a common structure called a class to organize its code using "object-oriented programming." Object-oriented programming is a style of coding involving the creation and use of 'objects', which are self-contained pieces of code that can store information and perform tasks.

In the case of Gryffin, an "instance" (*i.e.*, copy) of an object is created based on the Gryffin class, which is referred to as "object instantiation" in programming terms. This object can be customized by supplying information about the variables to be tuned and the objectives to be optimized. Once this object has been created, you can use its built-in functions (class methods) to perform various operations. For example, you can use the recommend function to get recommendations from Gryffin, or the build_surrogate function to build a surrogate model—a simplified representation of a more complex system.

Likewise, alab_management and Bluesky utilize classes. For example, alab_management offers base classes for devices and tasks. A user only needs to create a custom class for a specific device or task once that can be reused, making it unnecessary to copy-paste "boilerplate" code. Bluesky, designed with synchrotron facilities in mind, uses "motors" and "detectors" to clarify the difference between hardware that performs tasks based on inputs (*e.g.*, temperature controllers, sample changers) and characterization hardware that produces research data (*e.g.*, photodiodes, CCD cameras, spectrometers).[121]

While the hardware associated with low-cost SDLs may not be as performant as high-cost examples, the same SOTA software that is deployed on a high-cost SDL can be deployed to a low-cost SDL with minimal effort. This enables both rapid, low-risk prototyping (Section 2.2) and opportunities to integrate low-cost and high-cost experiments *via* multi-fidelity optimization (Section 4.3.2). A more general discussion of SOTA optimization with workflow orchestration tools and algorithms is given in Section 4.3.

4.2 Design approaches

In this section, we describe three different design approaches for SDLs. The most common of these for automation is the

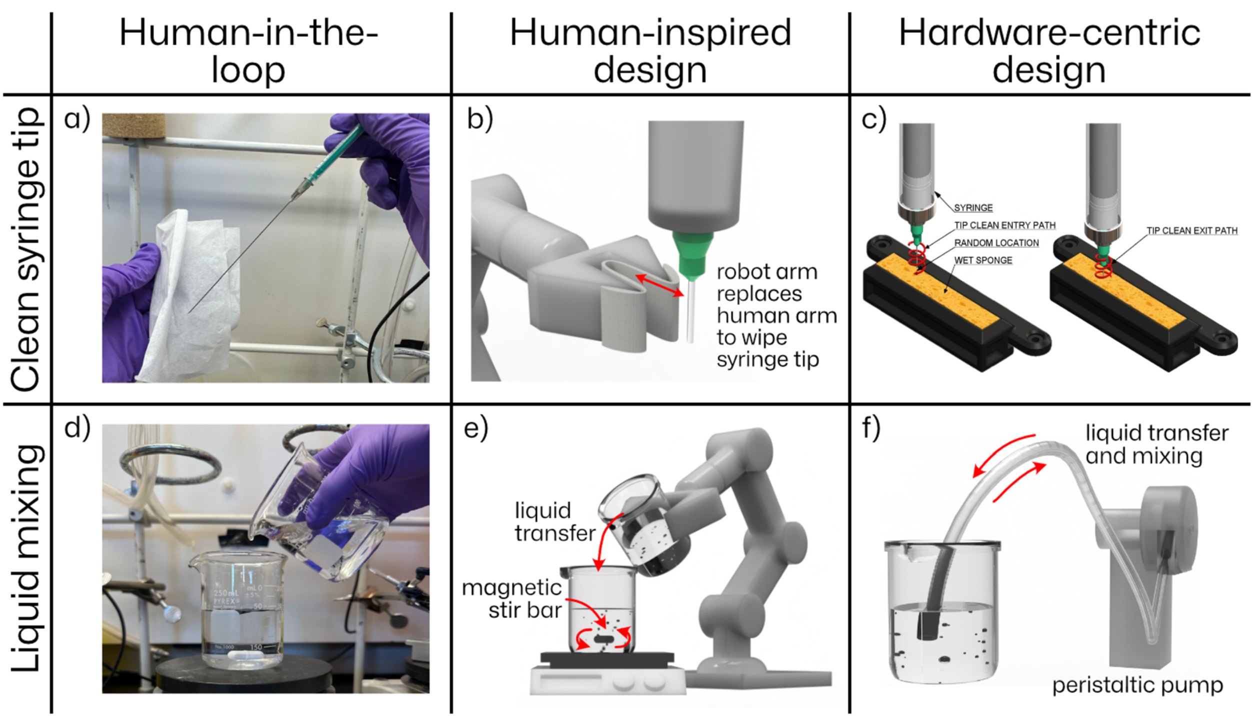


Fig. 10 Examples of human-in-the-loop *vs.* human-inspired *vs.* hardware-centric design. (a) Wiping a needle by hand *vs.* (b) wiping a needle using a cloth attached to a robot arm *vs.* (c) helical insertion into a sponge. Adapted from ref. 17 under the Creative Commons Attribution license (CC-BY). Copyright © 2021, this is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply. (d) Mixing liquids together in a traditional lab setting using manual pouring *vs.* (e) using a peristaltic with a digitally controlled stir plate *vs.* (f) leveraging a bidirectional peristaltic pump to perform both liquid transfer and mixing.

human-inspired approach (Section 4.2.1) because of the intuitive translation between human and robotic motion. Alternatively, hardware-centric design (Section 4.2.2) is becoming more prevalent due to taking better advantage of the potential of hardware components. However, at times, it is more cost-effective and practical to keep the human in the loop (Section 4.2.3) for the main objective of accelerating scientific discovery. Each of these approaches is conceptually summarized in Fig. 10. At the end of Section 4.2.3, we describe the role of frugal twins in bridging gaps between these seemingly disparate design philosophies.

4.2.1 Human-inspired. When most people think of robots, they think of human-inspired robotic design (Fig. 10b and e), where robots perform tasks as a human would approach the problem. For example, robotic arm setups[23,36] are often used to mimic human behavior. While there are benefits, this design approach exhibits its own set of trade-offs. We define human-inspired design as mimicking human behavior to accommodate traditional experiments.

For example, robots can be made to use existing, human-centric lab equipment without modification.[23] However, without complex sensing capabilities such as computer vision, a hard-coded system is sensitive to slight perturbations in absolute positions and orientations. This often requires extensive routine calibration and is tedious to implement when integrating new scientific instrumentation. The introduction of computer vision to recognize particular objects can introduce greater flexibility but suffers from the larger startup cost of the vision algorithm and may not elegantly handle all possible situations. Additionally, glassware is an essential component of any chemistry lab, but it is incredibly challenging for computer vision to recognize transparent objects.[122]

An alternative that combines the benefits of hard-coded routines and complex computer vision decisions is to use fiducial systems such as AprilTags,[123,124] which are used by Wang *et al.*[125] and Xu *et al.*[122] (Fig. 11). These can be thought of as QR codes or bar codes attached to pieces of equipment to help with relative positioning. However, the true value is not simply to identify hardware with unique IDs; the AprilTag detection software allows for computation of "the precise 3D position, orientation, and identity of the tags relative to the camera." More recent work also enables flexible fiducial markers to be placed on circular, annular, and other shaped objects[126] such as vials. Likewise, Krogius *et al.*[126] demonstrate the use of nested, recursive layouts for high dynamic range. While there are challenges associated with mimicking human behavior, there remain excellent use cases for the human-inspired approach.

4.2.2 Hardware-centric. Replicating human behavior is often a difficult task such as computer vision using cameras or sample transfer between modules, which are tasks that humans excel at but robots do not. An effective alternative to the human-inspired design approach exists which we refer to as hardware-centric design where existing hardware is leveraged to carry out experiments without mimicking human actions. This has been previously noted. For example, Seifrid *et al.*[3] state: "[It] is critical to understand that adapting experimental procedures that were designed for human experimenters is not as simple as transferring those same actions to an automated system, and there may be more efficient ways to achieve the same goal in an automated fashion." Similarly, Abolhasani and Kumacheva[4] discuss the nuances between using a mobile robot arm, a stationary robot arm, and fluidic sample transfer, each with varying levels of human-likeness and difficulty.

In terms of low-cost SDLs, Deneault *et al.*[17] provide a prudent example of leveraging the existing robotic setup (a 3-axis printer) and moving the syringe into and against a fixed sponge with a helical motion to clean the external surface of the syringe (Fig. 10c). When cleaning a syringe, a human might run it under water, wipe it with a cloth (Fig. 10a), put it in an ultrasonic cleaner, or replace the tip entirely. A robotic arm with human-inspired design could be equipped with a cloth to wipe the syringe tip (Fig. 10b), or remove the tip and place it in an ultrasonic cleaner. However, helical insertion into a sponge

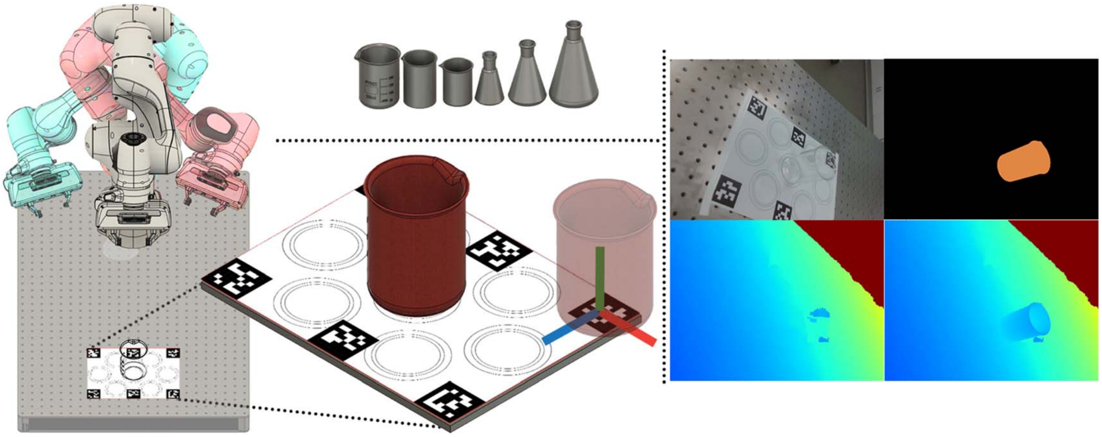


Fig. 11 AprilTags, a type of fiducial marker, are affixed to a base plate to allow for accurate detection of its position and orientation (six degrees of freedom) relative to the camera. Reproduced from ref. 122 with permission from H. Xu, Y. R. Wang, S. Eppel, A. Aspuru-Guzik, F. Shkurti and A. Garg, *arXiv*, 2021, https://doi.org/10.48550/arXiv.2110.00087.

leverages existing equipment at a low cost. While it has limitations (*e.g.*, how well is the syringe tip cleaned relative to more standard procedures; cross-contamination), it is an informative example of hardware-centric design. Another example is liquid handling that is dual-purposed for both dispensing and mixing, where mixing occurs by cycles of forward and reverse pumping to agitate the solution (Fig. 10f) instead of using a magnetic stir bar and stir plate (Fig. 10d and e).

By designing equipment with desired material states and processing conditions in mind, we create hardware that is time- and cost-efficient for autonomous experimentation. Especially in low-cost settings, we should try to do as much hardware-centric design as possible. This will both lower cost and require less equipment.

4.2.3 Human-in-the-loop. However, it can be easy to over-automate, whether in hardware-centric or human-inspired design. Sometimes, we need humans to be "in the loop" for tasks where robots do not excel. We have evidence from Amazon, Tesla, Carnegie Mellon University cloud labs, and personal experience, where robots do not perform well on certain tasks. We define human-in-the-loop design as systems that require manual human intervention during an experiment.

Here, we draw from the "Pareto principle," described by Jana and Tiwari[127] as a commonplace case where "80% of the outcomes are controlled or decided by 20% of the activities or factors. For example, 80% of the total profit is generated by 20% of the product categories, or 80% of the maintenance expenses are incurred by 20% of the machines." Applying the Pareto principle, the last 20% of automation may require 80% of the total effort towards bringing full autonomy to an experiment. A common example is sample transfer between automated experimental modules, especially of solid materials or sample containers. For example, samples often need to be moved between synthesis and characterization equipment, such as the transfer of wellplates between an OT-2 robot and a plate reader in Vaddi *et al.*[101]

In the low-cost automation literature, there are many examples which incorporate automated modules while leaving experimental step(s) as human-in-the-loop because of high opportunity cost (*i.e.*, the benefits that are lost when one makes a decision over an alternative – such as the lost opportunity for students to learn hands-on from running an experiment manually when it is automated), time constraints, and tasks where humans are naturally better than robots. Xie *et al.*[67] automate the design and synthesis of metal–organic frameworks (MOFs) using Bayesian optimization (BO) and a RepRap 3D printer but leave humans to transfer the sample from the robot to the X-ray diffraction instrument. Since many of these complex characterization techniques are costly and designed for humans, the time and cost of building another robot to perform sample transfer exceed the benefits gained from automating every single task in the workflow for greater efficiency. Rodriguez *et al.*[128] provide an excellent example of automating the most effective process steps such as synthesis (with an Opentrons OT-2 liquid handling robot), melting point determination, and electrochemical characterization for discovering new deep eutectic solvent electrolytes. Rodriguez *et al.*[128] did not automate the processes of sample transfer or handling of existing equipment such as a dehydrator and vacuum oven because of the great opportunity cost.

In a similar vein, most of the experimentation in Salley *et al.*,[103] Cao *et al.*,[129] and Lachowski *et al.*[130] is automated except for the characterization tools which include XRD, viscosity analysis, and UV-Vis spectroscopy, respectively. Conversely, Chen *et al.*[131] develop a new low-cost system, RAMSAY-2, for automating the burdensome task of sample preparation for mass spectroscopy. It involves two robot arms which aliquot solutions, incubate the samples with the reagents, deliver the samples to the ion source of the mass spectrometer, and initiate data acquisition.[131] This approach significantly accelerates the characterization workflow but is a non-trivial solution that requires substantial time and effort. It is also important to consider the opportunity cost of automating tasks that are trivial for humans but challenging for robots due to the consequential researcher time spent. Automation is most profoundly effective when researchers are freed from performing tedious, time-consuming, and repetitive tasks. Another opportunity cost is the amount of money required to acquire instruments that are already automated. For example, an automated differential scanning calorimetry (DSC) instrument can be purchased for ∼50 000 USD.[132] However, Rodriguez *et al.*[133] automate DSC with a low-cost system of 1080 USD, which can run samples in 15 minutes, with up to 96 samples at a time.[133] A cost/benefit analysis of the different design approaches and associated opportunity costs remains necessary to automate any solution.

4.2.4 Role of frugal twins. While the implementation cost of robotic solutions can currently be prohibitive, the exploration of low-cost sample transfer, especially of solid materials and across modules remains important and robotic solutions remain a warranted goal. To push the agenda with a future-looking vision, we need to put low-cost frugal twins in the hands of the community.

Rather than polarizing the community between fully autonomous *vs.* human-in-the-loop generalist setups, we believe it is wiser to meet in the middle and pair the tool to the task. This type of experimentation and exploration, enabled by low-cost frugal twins, can form a rich test bed in classroom settings. For example, students could be tasked with a design problem and divided into three groups: human-in-the-loop, human-inspired robotic design, and hardware-centric design. The students can present their experiences, learn from other groups, and discuss trade-offs between each approach: how many experiments could be performed within the first day for each group? Within the first week? This can be replicated for different experiments to solidify best practices related to autonomous system design and cross-pollinate seemingly disparate design approaches.

4.3 State-of-the-art software

Seifrid *et al.*[3] present challenges of setting up a SDL, such as the need for algorithms that can handle constraints and unexpected outcomes, and difficulties surrounding software control and integration (stemming from instrument manufacturers generally not designing with SDLs in mind). Here, we highlight key places where SDLs can benefit from leveraging and integrating frugal twins with SOTA software. This includes topics such as batch and asynchronous optimization (Section 4.3.1), multi-fidelity optimization (Section 4.3.2), workflow orchestration (Section 4.3.3), and cloud experimentation (Section 4.3.4).

4.3.1 Batch and asynchronous optimization. Fundamental to optimizing efficiency in the lab is the parallelization of experiments, which reduces the time to obtain results and allows more efficient experimental design. Using lower-cost hardware, even with an initial potential for loss of accuracy, facilitates parallelization of SDLs. This democratizes access to cutting-edge research tools, such that geographically distant labs can build clones of the same low-cost SDL. These SDLs can then network to execute high-throughput and parallel materials discovery campaigns. Caramelli *et al.*[73] demonstrate the advantages of low-cost parallelization of SDLs with their network of identical autonomous research systems (Fig. 12). The systems can evaluate the variability across different instances of the robot with four different experimental tasks in a financially reasonable manner (*i.e.*, the hardware components of their SDLs are low-cost (≤500 USD)). Similarly to adding more cores to a CPU, adding more instances of an SDL (which need not be in the same location or even operating on the same step at a given point in time) increases throughput for an optimization campaign at the cost of additional hardware. However, it is important to acknowledge the trade-off between parallelization and the total number of trials in an optimization campaign. There is an adaptivity gap between the parallel and the sequential approach for optimization models. In the parallel approach, the model is required to make decisions in advance of having all of the information. If time is not a limiting factor and/or cost is a limiting factor, it is ideal to prioritize the sequential approach. Conversely, if time is a limiting factor and/or cost is not a limiting factor, it is more efficient to prioritize the parallel approach. For additional discussion, see "Tradeoff between parallelism and total number of trials".

While the batch optimization described earlier implies that all experiments within the batch need to be completed before moving on to the next one, the complementary topic of asynchronous optimization uses resources as soon as they become available. This is important when experimental runtimes can vary depending on the input parameters: thereby, equipment downtime is reduced. Whether using batch or asynchronous optimization, care must be taken so that redundant or low-value experiments are not suggested by considering either completed or in-progress experiments. Examples of methods that factor in-progress experiments into the optimization scheme include Monte Carlo-based joint acquisition optimization and models where predictions for in-progress experiments are sequentially added as "fantasy datapoints" before suggesting the next experiment in the batch (see Appendix F2 of Balandat *et al.*[134]).

4.3.2 Multi-fidelity optimization. Another use of building low-cost SDLs is to have them work in tandem with high-cost SDLs on the same discovery campaign through multi-fidelity optimization. Multi-fidelity optimization refers to leveraging multiple information sources with varying accuracy and cost. In chemistry and materials science, many optimization problems involve finding the best set of parameters or conditions that maximize a certain objective function, such as the yield of a reaction or the strength of a material. However, obtaining accurate predictions for these systems often requires robust, reproducible, and expensive experimental setups. In the case for SDLs, multi-fidelity optimization seeks to balance the trade-off between accuracy and cost by using multiple SDLs of varying levels of fidelity, where fidelity refers to the degree to which an SDL accurately represents the true system. One approach is to start with a low-fidelity instrument which could be a low-cost

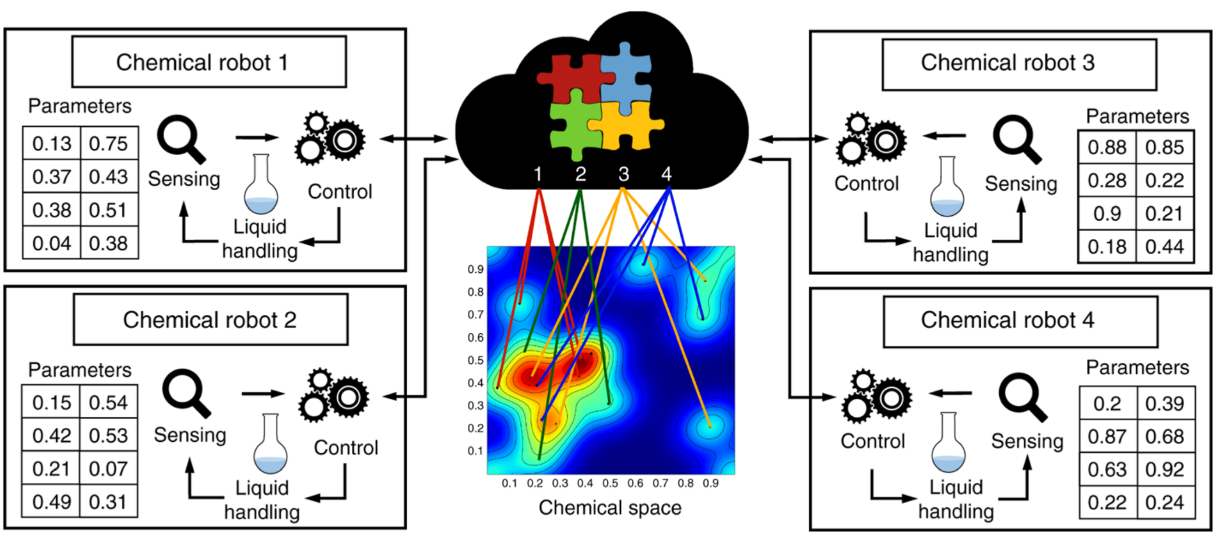


Fig. 12 Illustration of a network of parallel chemical synthesis robots working towards a common optimization goal.[73] Reproduced from ref. 73 with permission under the Creative Commons Attribution license (CC-BY). Copyright © 2018, Caramelli *et al.*

SDL, to explore the parameter space and identify promising regions, and then employ a higher-fidelity SDL which is generally higher in costs to refine the optimization in those regions. This can reduce the overall cost of the optimization while still achieving high accuracy in the final result. Multi-fidelity optimization can also involve the incorporation of different types of data, including both simulations and experiments or multiple types of experiments.[135,136] For example, as mentioned in Section 2.1, in-line IR or benchtop NMR is a low-fidelity yet high-throughput approach compared to the gold standard NMR instrument which is high-fidelity but single-throughput. By coarsely exploring the search space with in-line IR or benchtop NMR, only the experiments for finer optimizations in promising regions are directed to the gold standard NMR instruments which can reduce the time and cost of operating the high-cost NMR instruments.

4.3.3 Workflow orchestration. When experiments contain multiple steps, workflow orchestration software should ideally be used. While custom code can be written to manage workflows, it is preferable to use existing packages that are fully-featured, modular (see software modularity in Section 4.1.2), and well-maintained to streamline orchestration efforts. Examples of workflow orchestration platforms include Covalent, BlueSky, alab-management, and HELAO. A curated list of workflow orchestration platforms applicable to SDLs is available in **https://github.com/AccelerationConsortium/awesome-self-driving-labs** under the "Workflow Orchestration" section.

4.3.4 Cloud experimentation. "Cloud experimentation" allows users to be geographically distant from experimental hardware, in analogy to cloud computing, where software programs can be executed remotely. One of the key benefits of removing geographic barriers is the decentralization of expertise.[137] For example, domain specialists, roboticists, and software developers can collaborate across continents on the same experiments.

Several examples of cloud-based SDLs exist.[65,73,138–144] Many commercial solutions have a heavy focus on biology applications such as Emerald Cloud Lab,[139] the former Lilly-Strateos lab,[140] Culture Biosciences,[141] and Arctoris.[142] On the other hand, solid-state materials science cloud laboratories are effectively non-existent except for some minor capabilities of biology- and chemistry-focused labs. While existing cloud labs have primarily targeted industry users, a noteworthy example beginning to target academic users is CMU Cloud Lab.[145–150] This is a partnership between Carnegie Mellon University and Emerald Cloud Labs to build a subscription-based, 40 million USD facility with over 200 types of scientific instrument. Unlike typical user research facilities, academic and industry users can conduct an end-to-end experimental workflow and acquire the results from anywhere around the world, 24/7, 365 days a year.[145–150] Typically, a research group needs to secure funding for the reagents, cost of the instrument, and upkeep costs to perform an experiment. Armer *et al.*[151] outline several systemic reasons for the lack of adoption of cloud-based science, such as the lack of initial cloud access to gain preliminary data for grant applications, the lack of cloud science grants in general, the lack of academic training, and the costs for a cloud lab subscription in addition to university facility expenses. To tackle some of these concerns, having an academic institution such as CMU build its own cloud labs will reduce some of the barriers of entry for academics to access high-cost scientific equipment.[151] In addition, CMU Cloud Lab promotes open science, a recent movement that aims to enhance the transparency, accessibility, inclusivity, and credibility of scientific knowledge,[152] where problems and results can be shared easily.

A platform such as CMU Cloud Lab typically requires extensive capital and expertise to develop onboarding, security, access restriction, priority queuing, and workflow orchestration protocols. It also relies on human-in-the-loop sample transfer between modules, necessitating full-time technicians to perform menial tasks. The costs associated with these infrastructure components inevitably get passed onto the user which can be prohibitive for educational settings and citizen science. Since low-cost SDLs operate at a smaller scale and the risks associated with data leakage and malicious threats are lower, they are a great platform for prototyping SDL infrastructure with low operational costs. For example, free, open-source tools may be implemented into low-cost SDLs, such as Bluesky for workflow orchestration,[121] secure, encrypted IoT-style communication through platforms such as HiveMQ,[83] and the Google Authentication application programming interface for security measures.[121] By leveraging the advantages of rapid, low-risk prototyping benefits of SDL frugal twins described in Section 2.2, we envision a low-cost SDL cloud lab that can act as a test bed for research-grade cloud experimentation ecosystems, but with dramatically lower operational costs. See Discussion #62 and Discussion #91 from Section 7.

5 Ethical benefits and risks

With any new technology, there are several ethical benefits and risks to consider, especially if low-cost SDLs can be put into the hands of many without regulation or guidelines, due to their low cost. In this section, we attempt to highlight why low-cost SDLs should overcome societal barriers to enable citizen science (Section 5.1), and address the concerns around democratizing this technology which is capable of discovering novel substances (Section 5.2).

5.1 Citizen science

Access to research facilities has historically been limited to universities, government, and industry laboratories, and their personnel. This limitation reduces access for non-professional, citizen scientists, many of whom could contribute greatly to the body of scientific understanding.[153] The lack of gender, racial, ethnic, and socioeconomic diversity, equity, and inclusion in science hinders a truly representative citizen science.[154] We hope that by making SDLs low-cost, accessible, and open source, it will be easier to build equity and inclusion into the educational system.

Additive manufacturing (*i.e.*, 3D printing) is a natural place for citizen science, as it is low-cost, operationally fairly safe, easy

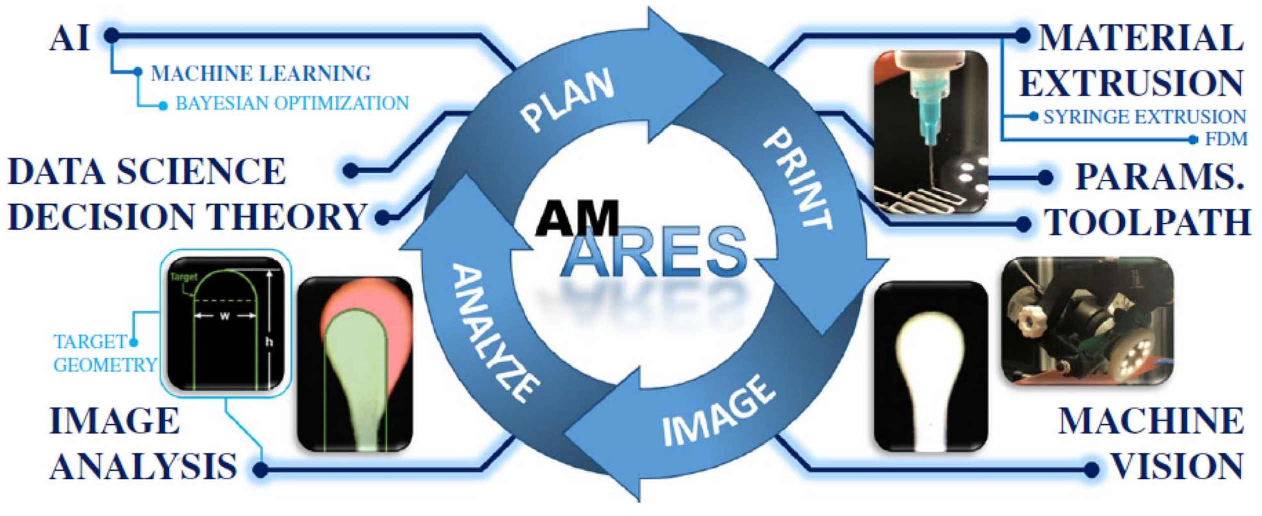


Fig. 13 A simplified closed-loop workflow of the AM ARES platform. Reproduced from ref. 17 with permission under the Creative Commons Attribution license (CC-BY). Copyright © 2021. This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply.

to learn with the abundance of online resources, and adaptable to many different objectives. For example, Deneault *et al.*[17] developed an SDL known as Additive Manufacturing Autonomous REsearch System (AM ARES) for optimizing the print parameters of several materials for additive manufacturing. This is a low-cost additive manufacturing SDL that uses a 300 USD commercial 3D printer with a custom syringe extruder, Raspberry Pi controllers and webcams, and software that will be released as open-source (Fig. 13). The authors use BO to guide the selection of 3D print parameters for latex caulk with silicone additives, attaining excellent extrusion properties after 100 iterations. In addition, AM ARES performed self-calibration for three different unknown source filaments, which resulted in better performance than default manufacturer specifications in an average of 15 experimental iterations. Although this system is robust, low-cost, and a stepping-stone for many to learn about SDLs, there is yet to be widespread adoption due to the lack of educational infrastructure such as open-source software, course materials, and a step-by-step build guide.

To address this problem, the project was extended between the US Air Force Research Laboratory and Airship Consulting to create ATHENA, an affordable AM ARES system with open-source software (ARES OS 2.0) and off-the-shelf hardware. This initiative aims to make SDLs and autonomous experimentation systems widely accessible in grade schools, trade schools, and universities. ARES OS 2.0 is a platform-agnostic, web-facing software framework for autonomous experimentation SDLs which takes much of the software development burden from the researcher. The goal is to provide a library of open-source modules for all to use and contribute back to the growing community, with the intent that "Anyone Can Download an Autonomous 'Research Robot'".[155] ATHENA is an example of the movement towards low-cost autonomous experimentation systems/SDLs to improve access to citizen scientists and especially under-served communities through open-source software and low-cost systems.

5.2 Risks

While we have focused on how SDL systems accelerate the discovery of beneficial materials, autonomy can be a double-edged sword if it leads to the creation of dangerous substances, whether by accident or design. As with any technology, there are risks for people or organizations to engage in actions that are harmful, illegal, or morally wrong.‡ We recognize that this is a polarizing topic. On one hand, there will always be some people with malicious intent; people will figure out a way. For example, the widespread adoption of low-cost 3D printers resulted in an increase in 3D-printed guns. Updated legislation regarding firearm manufacturing and use plays a key role in regulating this increase. However, the large majority of gun-related incidents do not seem to involve so-called "ghost guns" (*i.e.*, 3D printed guns). In another example, explosives can be created from commonly obtained materials, and safeguards have been put in place, such as limiting purchase amounts or requiring licenses, permits, and certifications. Naturally, regulations are also region-dependent. Recently, concerns have been raised about the potential for large language models and autonomous platforms (*e.g.*, cloud laboratories) to be used for nefarious purposes such as the synthesis of illicit drugs or chemical weapons.[137,156–159]

We do not have the solution for safeguarding SDLs, but methods exist to make it harder for ill-intentioned people and organizations to engage in harmful behavior and easier for researchers to implement preventive strategies against the (accidental) synthesis of harmful substances. The key is to address this problem early, quickly, and judiciously through governance, regulations, standards, education, awareness, and self-adherence to ethical use.

There are valuable open source practices that can be learned and adapted to low-cost SDLs because there are potential risks associated with open sourcing, such as open access to

‡ See "bad actor" definition in the Cambridge Dictionary.

hazardous information or datasets and the potential misuse of research tools. To mitigate these risks, a cultural shift towards open methodology and open review may help regulate the dissemination of malicious code, data, or materials.[160] Creators of SDLs should also consider designs which mitigate misuses or failure modes which would endanger lives or property. For example, incorporating steps to assess the toxicity of autonomously generated substances can prevent the release of unknown toxic chemicals into the environment.[161]

6 Future work

In this section, we describe ideas for new frugal twins (Section 6.1), suggested educational course content (Section 6.2), and classifying levels of autonomy (Section 6.3).

6.1 Ideas for new frugal twins

As mentioned in Section 3.1.2 there are several examples of low-cost SDLs involving liquid handling; however, low-cost SDLs involving the transfer and processing of solid matter are practically non-existent. This largely stems from the relative ease of transferring liquids using, *e.g.*, diaphragm or peristaltic pumps and tubes compared with solids using, *e.g.*, powder feeders and robotic arms (see Discussion #92). For perspective, autonomous powder dispensers such as Trajan's CHRONECT series cost significantly more (100k+ USD) than liquid handlers of similar resolution. Liquid transfer issues such as viscosity, density, and surface tension are largely solved problems. With powder handling, variable particle sizes, consistencies, and electrostatic interactions make it difficult to robustly dispense powders of different types using the same type of equipment. One workaround to transferring solids is to dissolve or disperse them in liquids (*i.e.*, as solutions or slurries); however, this approach is not feasible for many materials science scenarios where suitable solvents are unavailable or unwanted chemical reactions may occur. To complicate matters further, substrates and sample holders may be required to accommodate high temperatures, high pressures, or state changes (*e.g.*, solid to liquid).

To address the lack of solid-state materials science SDL demos, we propose a solid-based color-matching demo extension (Closed-loop Spectroscopy Lab: Solid-mixing (CLSLab:Solid)) that uses a low-cost mobile robot arm, mixtures of granulated colored wax powders (Fig. 14), and a halogen lamp. Similarly to moving from a light-mixing to a liquid-mixing demo (Section 3.1.2), the solid-mixing demo requires hardware and workflow changes. At the start of the experiment, a robotic arm will pick and place one tealight candle in a holder from a stacked array of holders in a storage array onto a motorized turntable. The turntable will then move the candle holder to a position beneath a funnel connected to red, yellow, and blue wax powder dispensers. The candle will then be positioned beneath a heat source (*e.g.*, halogen lamp) to melt and convectively mix the wax, followed by color sensing using the same sensor as CLSLab:Light and CLSLab:Liquid. When the candle holder returns to its original position on the turntable, the robotic arm will pick it up and place it into a separate storage/waste area.

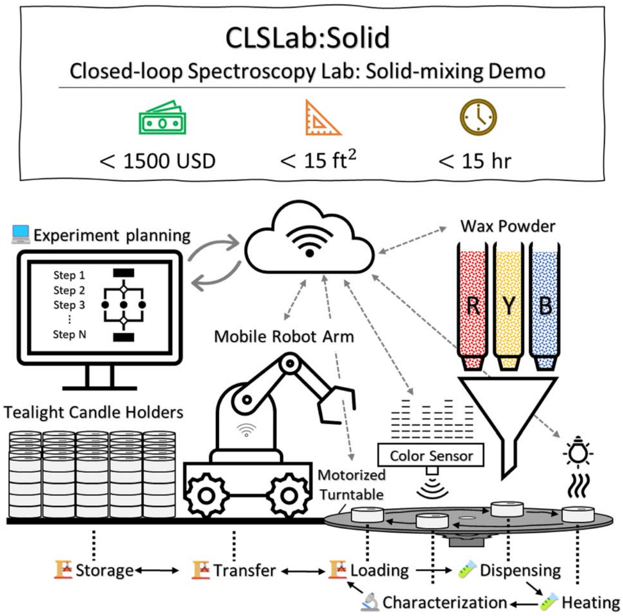


Fig. 14 A summary schematic of the CLSLab:Solid demo, which is envisioned as a minimal working example for an inorganic solid-state SDL. Taking from the light- and liquid-based color-matching demos, the task is to find the optimal mixture of wax powders and processing conditions to reach a desired, solidified wax color. This demo incorporates more advanced features than other demos due to need to handle and characterize solid samples. The demo is intended to be have reasonable trade-offs between the monetary cost, the time required for setup, and the device footprint.[55]

Moving one step further is the idea of a "robot chocolatier." Chocolate captures key materials science principles such as liquid phase transformations, bulk material characterization (as opposed to thin-film), and processing–structure–property (PSP) relationships. This robot chocolatier (RoboChocolatier) will reuse many components from CLSLab:Solid and add a do-it-yourself (DIY) tensile tester and a chocolate 3D printer such as the highly customizable Cocoa Press. Both CLSLab:Solid and RoboChocolatier act as toy examples for the more industry-relevant materials discovery task of additively manufactured metal alloys for aerospace and automotive applications. Again, as a recurring theme, they can serve as proofs of concept that can be used during prototyping and the preparation of grant proposals (Section 2). For a continuing discussion of solid-state materials science SDL demos, see Discussion #153.

Other topics that the community may consider exploring in the context of SDL frugal twins include other types of inorganic synthesis, battery formulations,[162–164] batch chemical synthesis, semiconductor fabrication, polymer synthesis,[165] artificial organ compatibility, mobile and fixed robotic arms, autonomous multi-agent systems,[166] microfluidic devices,[44] and closed-loop microscopy.[167,168]

6.2 Suggested course outcomes

Educators may be wondering how to incorporate SDL concepts into existing and new curricula. To streamline efforts to democratize SDLs, it is important to define course structures

Table 3 Suggested learning outcomes of a course covering SDL topics. For a continuing discussion, see Discussion #186

Topic	Potential learning outcome
Experience	Familiarize the concept of SDLs (hardware, algorithms, orchestration)
	Acquire hands-on and software development experience by setting up a toy demo
	Propose a design for a research-oriented SDL *via* a white paper
Best practices	Identify SDL best practices (*e.g.*, modularity, reproducibility, safety, documentation)
	Identify best practices for "cloud experimentation" (*e.g.*, data transfer, storage)
	Identify best practices for ML (*e.g.*, validation, prevention of data leakage)
Algorithms	Compare and contrast three forms of experiment planning algorithms
	Test the complexity/efficiency trade-offs for advanced optimization
	Identify methods for incorporating domain knowledge

and outcomes that can be tailored to meet the individual needs and disciplines of each student. Ideally, this would begin as early as middle- or high school and continue throughout associate- and bachelor-level undergraduate degree programs, including programming, data handling, physical "maker skills" (3D design and fabrication, electronics, and programming), automation, and the associated core science disciplines.[28]

We present in Table 3 suggestions for possible educational outcomes for hands-on experience, learning best practices, and using algorithms. Hands-on hardware and software development experience, brainstorming designs, and expertise in applying optimization algorithms are emphasized. We encourage the community to weigh in on and converge on a set of desired outcomes and skills necessary for successful SDL implementations. In future work, we plan to flesh out the details for creating a syllabus, course outline, and course content along with practical examples for teaching SDLs to students. Eventually, as the ecosystem matures, we envision higher education programs and degrees specific to SDLs for chemistry and materials science.

Once again, it is inevitable to mention the multi-tool motion platform developed at the University of Washington.[60] The platform was designed with community development and customization as one of the project's aims. Its original design was inspired by the RepRap and maker movements, which have already generated an array of open-source hardware toolkits enabling flexible and extensible technologies for laboratory automation. This connection anticipates the co-development of tools configured for platforms such as Jubilee. These features also make the platform a great educational tool, as it provides a solution with a low-cost barrier and allows students, from most disciplines, to obtain skills for all steps of an experimental campaign in a single SDL platform. A successful example of this is the implementation of Jubilee into engineering design courses at the University of Hawai'i at Mānoa.

6.3 Classifying levels of autonomy

In this work, we have focused on fully autonomous low-cost examples but also pointed out several partially autonomous examples that are equally important in accelerating the discovery of new materials and teaching the next generation of data-driven scientists. However, there are no established standards to define the levels of autonomy for SDLs. To better categorize levels of automated chemical design, Goldman *et al.*[33] proposed a set of definitions in the context of ideation (finding non-obvious trends) and decision making in chemical design, similar to those for self-driving vehicles.[169,170] They define the highest level of autonomy (level 5) as systems where these two processes are handled without human intervention over multiple iterations. Beal and Rogers[171] propose levels of autonomy for synthetic biology engineering which are also very similar to those for self-driving vehicles. They define the highest level of autonomy (level 5) as biology workflows where all of the protocol executions, data analysis, and interpretation are done by a machine, while the human only sets goals and receives results. The same levels described by Beal and Rogers[171] that focus primarily on synthetic biology systems, can also be closely described for the levels of autonomy of SDLs. The SDL community will benefit from collectively determining a set of classifications or standards. One possibility is to classify autonomy levels on a per-category basis: synthesis, characterization, sample transfer, and experiment planning.

To make these categories conceptually and visually easy to understand, emoji can be used to represent whether a process is fully autonomous *vs.* one that requires manual intervention (Fig. 15). This type of classification is utilized in **https://github.com/AccelerationConsortium/awesome-self-driving-labs** as of 2022-08-08. For a discussion centered on these representations, see **https://github.com/AccelerationConsortium/awesome-self-driving-labs/discussions/15**. Autonomy levels could also include failure rate/tolerance, number of iterations without manual intervention, or use of physics-based simulations to supplement experiments.

6.4 Frugal twins in biology

Autonomous experimentation has also captured the attention of biologists. Exciting examples in biology include autonomous experimentation for genome engineering,[172–174] and optimal growth of cell cultures.[175] Si *et al.*[173] and HamediRad *et al.*[174] utilize iBioFab, a delocalized biofoundry which is similar to the concept of delocalized experimentation with cloud labs. iBioFab can produce one gene sequence for <3 USD, so it is inexpensive from the user's perspective. However, it is expensive to build because it uses a Fanuc F5 robotic arm on a 5-meter track, Tecan Evo200 liquid handling robot, TECAN M1000 microplate reader, and more.[176,177] Kanda *et al.*[175] use an

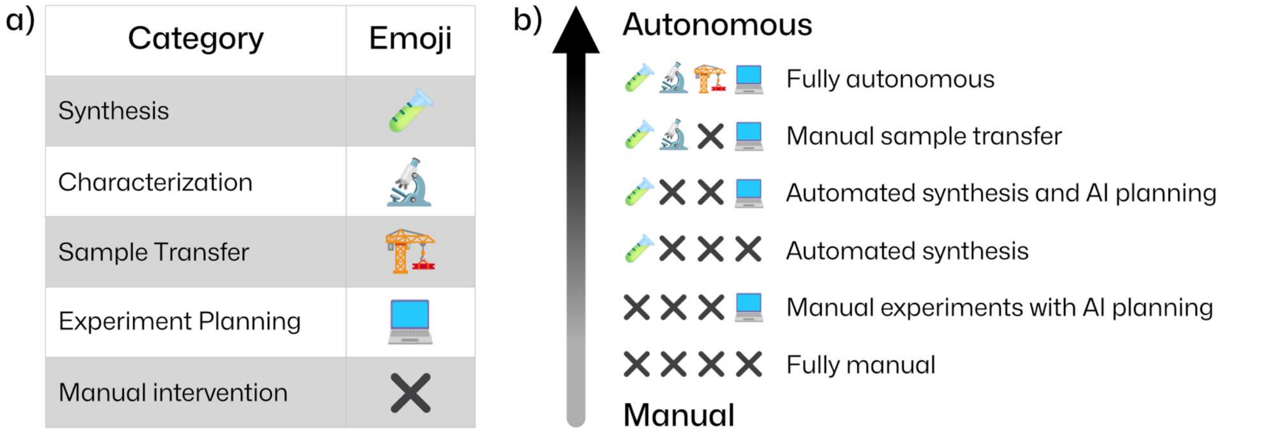


Fig. 15 (a) Legend for the emoji classification. (b) Classifying levels of autonomy in SDLs through multi-emoji classification. Emoji and their names and unicode values are given. Synthesis (🧪 "test tube": U+1F9EA); characterization (🔬 "microscope": U+1F52C); sample transfer (🏗 "building construction": U+1F3D7); experiment planning (💻 "personal computer": U+1F4BB); manual intervention (✖ "heavy multiplication X": U+2716). Please note that the exact symbols may appear differently on different systems. Alternatively, the symbols may be copy-pasted directly from https://github.com/AccelerationConsortium/awesome-self-driving-labs/blob/main/contributing.md.

Table 4 Self-driving-lab-demo GitHub discussions and awesome-self-driving-labs GitHub discussions for various topics related to SDLs

Topic	Repository	Link
All discussions	Self-driving-lab-demo	All discussions
Data and access management	Self-driving-lab-demo	Category
Demo extensions and design	Self-driving-lab-demo	Category
Examples and tutorials	Self-driving-lab-demo	Category
Scaling up SDLs	Self-driving-lab-demo	Category
Packaging open-source hardware as commercial kits	Self-driving-lab-demo	Discussion #124
Experimental orchestration software	Self-driving-lab-demo	Discussion #64
Educational outcomes and homework problems	Self-driving-lab-demo	Discussion #186
Solid-state materials science demo	Self-driving-lab-demo	Discussion #153
Low-cost powder handling	Self-driving-lab-demo	Discussion #153
Roadmap for demo extensions	Self-driving-lab-demo	Discussion #77
A network of cloud-based experiments	Self-driving-lab-demo	Discussion #62
Classifying level of autonomy	Self-driving-lab-demo	Discussion #15
What is a self-driving lab?	Awesome-self-driving-labs	Discussion #32

industrial life science robotic system, Maholo LabDroid, which costs approximately 890 000 USD.[178] There are not yet demonstrations of frugal twins for high-cost SDLs in biology applications. However, we are aware of two frugal twin systems for cell growth. PioReactor is <300 USD open-source bioreactor which can control peristaltic pumps and temperature and monitor realtime optical density to optimize yeast, bacteria, and algae growth.[64] Gerber *et al.*[179] described the use of a LEGO Mindstorm EV3 kit (<400 USD) to build a liquid handling robot with a light sensor, and its use in a K-12 afterschool setting to perform experiments related to sterile transfer and determine optimal sucrose concentrations for yeast growth.[179] Despite these early examples, frugal twins in biology remain an underexplored research direction.

7 A continuing discussion

While a review article represents a fixed snapshot, there is a benefit to allowing a continuing discussion of these important topics in a less rigid environment[180] that is amenable to the fast-paced evolution of SDLs. While this can also take on many forms such as social media and informal communication, we provide a public, organized, and persistent set of public, ongoing discussions hosted on GitHub, as summarized in Table 4. Anyone can access up-to-date dialogue relevant to low-cost SDLs, and SDLs in general. GitHub accounts are free, and users may contribute to existing threads or open entirely new discussions. We hope that the content in this article spurs further dialogue in the community around democratizing SDLs, defining best practices, and gaining hands-on experience with advanced ML algorithms.

8 Conclusion

SDL frugal twins can equip the next generation with the necessary skills, provide a low-risk environment for prototyping and hands-on learning, and help to create a more equitable, global ecosystem through decentralized equipment, software, and expertise. SDL frugal twins are being used for both education and research, and there is much room for improvement. Modularity for both hardware and software is an effective design principle for reducing redesign and maintenance costs, and care must be taken when considering human-inspired *vs.* hardware-centric *vs.* human-in-the-loop design approaches. The true value of these low-cost systems can be realized when SOTA software implementations such as batch and multi-fidelity optimization, workflow orchestration, and cloud experimentation are combined with SDL frugal twins across the spectrum. With the ethical and responsible use of this technology, frugal twins are poised to accelerate the discovery of society-benefiting materials within the SDL community.

Abbreviations

AM ARES	Additive Manufacturing Autonomous REsearch System[28,29]
CLSLab:Light	Closed-loop Spectroscopy Lab: Light-mixing[8,10–12,30]
CLSLab:Liquid	Closed-loop Spectroscopy Lab: Liquid-mixing[10,14,30]
CLSLab:Solid	Closed-loop Spectroscopy Lab: Solid-mixing[30,31]
HPLC-MS	High-performance liquid chromatography coupled with mass spectrometry[6]
ML	Machine learning[8,17,34]
SDL	Self-driving laboratory[1–3,5–8,10,12,14,17–20,23–25,27,28,30–36]
SOTA	State-of-the-art[1,6,7,15,17,20,24,34]

Data availability

As this is a review article, no primary research results, data, software or code have been included. More information and resources can be found at **https://github.com/sparks-baird/self-driving-lab-demo/discussions** and **https://github.com/AccelerationConsortium/awesome-self-driving-labs**.

Author contributions

Andrés Aguilar-Granda: writing – review & editing, visualization. Alán Aspuru-Guzik: supervision, project administration, funding acquisition, conceptualization, writing – review & editing. Sterling G. Baird: project administration, conceptualization, data curation, writing – original draft, writing – review & editing, visualization. Ben Blaiszik: validation, resources, writing – review & editing. Nessa Carson: writing – review & editing. Ian Foster: writing – review & editing, visualization. Sergei V. Kalinin: conceptualization, writing – original draft. Stanley Lo: project administration, conceptualization, data curation, writing – original draft, writing – review & editing, visualization. Benji Maruyama: writing – review & editing. Maria Politi: writing – original draft, writing – review & editing. Joshua Schrier: writing – review & editing. Taylor D. Sparks: supervision, project administration, funding acquisition, conceptualization, writing – review & editing, visualization. Helen Tran: supervision, funding acquisition, writing – review & editing.

Conflicts of interest

A. A.-G. is the chief visionary officer and a board member of Kebotix, Inc., a company that carries out closed-loop molecular materials discovery. A. A.-G. is also the founder of Intrepid Labs, Inc., a company that builds closed-loop machine learning algorithms for drug discovery. J. S. is a scientific advisory board member of Atinary Technologies Inc., a company that builds machine learning algorithms and integrations for SDLs.

Acknowledgements

We acknowledge Keith Brown and Raymundo Arroyave for suggestions and discussion related to the use of chocolate as a solid-state material teaching demonstration for SDLs. We acknowledge Michelle Murphy for an insightful discussion on the ethical benefits and risks of low-cost SDLs. We acknowledge Lilo Pozzo, Tonio Buonassisi, Leroy Cronin, Edward Dunlea, and Zhe Liu for asynchronous discussions, suggestions, and feedback. We acknowledge Graig S. Ganitano, and Gilbert L. Peterson for their contribution and discussions about AM ARES, ARES 2.0, and ATHENA. ChatGPT and GitHub Copilot Chat were used sparingly and judiciously, always with verification. ChatGPT 4.0 was used for generating AI renderings of lab equipment. This research was undertaken thanks in part to funding provided to the University of Toronto's Acceleration Consortium from the Canada First Research Excellence Fund. S. Lo acknowledges the support from the Natural Sciences and Engineering Research Council (NSERC) under the "NSERC discovery grant: RGPIN-2020-05634" project (Fund number: 508893) and the Ministry of Colleges and Universities under the Ontario Graduate Scholarship. S. Baird acknowledges the support from the National Science Foundation Division of Materials Research (DMR-1651668). J. Schrier acknowledges the support from the National Science Foundation Division of Physics Research (NSF PHY-2226511). S. V. Kalinin acknowledges the support from the US Department of Energy, Office of Science, Office of Basic Energy Sciences, as part of the Energy Frontier Research Centers program: CSSAS – The Center for the Science of Synthesis Across Scales (DE-SC0019288). H. Tran acknowledges the support from NSERC (RGPIN2021-03554) and the Canada Foundation for Innovation (JELF-41743). M. Politi acknowledges the support from the National Science Foundation (DMR-2121848 and TI-2229018). T. Sparks acknowledges the support from the National Science Foundation (NSF 2334411). A. A.-G. acknowledges the generous support of the Canada 150 Research Chairs program as well as Dr Anders G. Frøseth. A.-A.-G. refers to Alán Aspuru-Guzik.

References

1 M. Seifrid, J. Hattrick-Simpers, A. Aspuru-Guzik, T. Kalil and S. Cranford, *Matter*, 2022, **5**, 1972–1976.

2 *Reproducibility and Replicability in Science*, National Academies Press technical report, 2019.

3 M. Seifrid, R. Pollice, A. Aguilar-Granda, Z. Morgan Chan, K. Hotta, C. T. Ser, J. Vestfrid, T. C. Wu and A. Aspuru-Guzik, *Acc. Chem. Res.*, 2022, **55**, 2454–2466.
4 M. Abolhasani and E. Kumacheva, *Nat. Synth.*, 2023, **2**, 483–492.
5 J. A. Bennett and M. Abolhasani, *Curr. Opin. Chem. Eng.*, 2022, **36**, 100831.
6 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.
7 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**, eaaz8867.
8 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.
9 M. M. Flores-Leonar, L. M. Mejía-Mendoza, A. Aguilar-Granda, B. Sanchez-Lengeling, H. Tribukait, C. Amador-Bedolla and A. Aspuru-Guzik, *Curr. Opin. Green Sustainable Chem.*, 2020, **25**, 100370.
10 Automata, *What Is Lab 4.0 & Digital Transformation in the Lab?*.
11 Labmate, *What Is Laboratory 4.0? Labmate Online*, **https://www.labmate-online.com/news/laboratory-products/3/breaking-news/what-is-laboratory-40/57352**.
12 S. Peschisolido, *Lab 4.0: Making Digital Transformation Work for Your Laboratory/Technology Networks*, **https://www.technologynetworks.com/informatics/articles/lab-40-making-digital-transformation-work-for-your-laboratory-358583**.
13 *Laboratory Internet of Things (IoT) Devices/Labcompare.Com*, **https://www.labcompare.com/General-Laboratory-Equipment/26012-Laboratory-Internet-of-Things-IoT-Devices/**.
14 *Leverage the Internet of Things (IoT) within the Laboratory*, **https://www.pharmaceuticalonline.com/doc/leverage-the-internet-of-things-iot-within-the-laboratory-0001**.
15 T. Perraudin, *Internet of Laboratory Things Makes Life Better at Work*, 2020.
16 K. Williams, E. Bilsland, A. Sparkes, W. Aubrey, M. Young, L. N. Soldatova, K. De Grave, J. Ramon, M. De Clare, W. Sirawaraporn, S. G. Oliver and R. D. King, *J. R. Soc., Interface*, 2015, **12**, 20141289.
17 J. R. Deneault, J. Chang, J. Myung, D. Hooper, A. Armstrong, M. Pitt and B. Maruyama, *MRS Bull.*, 2021, **46**, 566–575.
18 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.
19 T. Isenhour, *J. Chem. Inf. Comput. Sci.*, 1985, **25**, 292–295.
20 P. Nikolaev, D. Hooper, F. Webber, R. Rao, K. Decker, M. Krein, J. Poleski, R. Barto and B. Maruyama, *npj Comput. Mater.*, 2016, **2**, 16031.
21 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.
22 A.-C. Bédard, A. Adamo, K. C. Aroh, M. G. Russell, A. A. Bedermann, J. Torosian, B. Yue, K. F. Jensen and T. F. Jamison, *Science*, 2018, **361**, 1220–1225.
23 B. Burger, P. M. Maffettone, V. 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.
24 J. Grizou, L. J. Points, A. Sharma and L. Cronin, *Sci. Adv.*, 2020, **6**, eaay4237.
25 A. E. Gongora, B. Xu, W. Perry, C. Okoye, P. Riley, K. G. Reyes, E. F. Morgan and K. A. Brown, *Sci. Adv.*, 2020, **6**, eaaz1708.
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**, 2001626.
27 A. G. Kusne, H. Yu, C. Wu, H. Zhang, J. Hattrick-Simpers, B. DeCost, S. Sarker, C. Oses, C. Toher, S. Curtarolo, A. V. Davydov, R. Agarwal, L. A. Bendersky, M. Li, A. Mehta and I. Takeuchi, *Nat. Commun.*, 2020, **11**, 5966.
28 M. Abolhasani, K. A. Brown and G. Editors, *MRS Bull.*, 2023, **48**, 134–141.
29 A. Choudhury, *Archives of Computational Methods in Engineering*, 2021, **28**, 3361–3381.
30 C. W. Coley, N. S. Eyke and K. F. Jensen, *Angew. Chem., Int. Ed.*, 2020, **59**, 22858–22893.
31 C. W. Coley, N. S. Eyke and K. F. Jensen, *Angew. Chem., Int. Ed.*, 2020, **59**, 23414–23436.
32 F. Delgado-Licona and M. Abolhasani, *Advanced Intelligent Systems*, 2023, **5**, 2200331.
33 B. Goldman, S. Kearnes, T. Kramer, P. Riley and W. P. Walters, *J. Med. Chem.*, 2022, **65**, 7073–7087.
34 M. L. Green, B. Maruyama and J. Schrier, *Applied Physics Reviews*, 2022, **9**, 030401.
35 C. J. Leong, K. Y. A. Low, J. Recatala-Gomez, P. Q. Velasco, E. Vissol-Gaudin, J. D. Tan, B. Ramalingam, R. I. Made, S. D. Pethe, S. Sebastian, Y.-F. Lim, Z. H. J. Khoo, Y. Bai, J. J. W. Cheng and K. Hippalgaonkar, *Matter*, 2022, **5**, 3124–3134.
36 B. P. MacLeod, F. G. L. Parlane, A. K. Brown, J. E. Hein and C. P. Berlinguette, *Nat. Mater.*, 2022, **21**, 722–726.
37 H. G. Martin, T. Radivojevic, J. Zucker, K. Bouchard, J. Sustarich, S. Peisert, D. Arnold, N. Hillson, G. Babnigg, J. M. Marti, C. J. Mungall, G. T. Beckham, L. Waldburger, J. Carothers, S. Sundaram, D. Agarwal, B. A. Simmons, T. Backman, D. Banerjee, D. Tanjore, L. Ramakrishnan and A. Singh, *Curr. Opin. Biotechnol.*, 2023, **79**, 102881.
38 B. Maruyama, J. Hattrick-Simpers, W. Musinski, L. Graham-Brady, K. Li, J. Hollenbach, A. Singh and M. L. Taheri, *MRS Bull.*, 2022, **47**, 1154–1164.

39 M. May, *Nature*, 2019, **569**, 587–588.
40 J. H. Montoya, M. Aykol, A. Anapolsky, C. B. Gopal, P. K. Herring, J. S. Hummelshøj, L. Hung, H.-K. Kwon, D. Schweigert, S. Sun, S. K. Suram, S. B. Torrisi, A. Trewartha and B. D. Storey, *Applied Physics Reviews*, 2022, **9**, 011405.
41 X. Peng and X. Wang, *MRS Bull.*, 2023, **48**, 179–185.
42 J. M. Perkel, *Nature*, 2017, **542**, 125–126.
43 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.
44 M. A. Soldatov, V. V. Butova, D. Pashkov, M. A. Butakova, P. V. Medvedev, A. V. Chernov and A. V. Soldatov, *Nanomaterials*, 2021, **11**, 619.
45 R. Vescovi, R. Chard, N. D. Saint, B. Blaiszik, J. Pruyne, T. Bicer, A. Lavens, Z. Liu, M. E. Papka, S. Narayanan, N. Schwarz, K. Chard and I. T. Foster, *Patterns*, 2022, **3**, 100606.
46 Y. Xie, K. Sattari, C. Zhang and J. Lin, *Prog. Mater. Sci.*, 2023, **132**, 101043.
47 T. Fuhrmann, D. I. Ahmed, L. Arikson, M. Wirth, M. L. Miller, E. Li, A. Lam, P. Blikstein and I. Riedel-Kruse, *Interaction Design and Children*, Athens, Greece, 2021, pp. 444–449.
48 L. Saar, H. Liang, A. Wang, A. McDannald, E. Rodriguez, I. Takeuchi and A. G. Kusne, *MRS Bull.*, 2022, **47**, 881–885.
49 S. Vargas, S. Zamirpour, S. Menon, A. Rothman, F. Häse, T. Tamayo-Mendoza, J. Romero, S. Sim, T. Menke and A. Aspuru-Guzik, *J. Chem. Educ.*, 2020, **97**, 689–694.
50 P. B. Vose and A. Cervellini, *IAEA Bull.*, 1983, **25**(2), 37–40.
51 A. Dhai, in *Culture Matters: International Research Collaboration in a Changing World—Summary of a Workshop*, National Academies Press, USA, 2014.
52 P. van Helden, *EMBO Rep.*, 2012, **13**, 395.
53 D. Jones, C. Snider, A. Nassehi, J. Yon and B. Hicks, *CIRP Journal of Manufacturing Science and Technology*, 2020, **29**, 36–52.
54 R. Saracco, *Computer*, 2019, **52**, 58–64.
55 S. G. Baird, *Sparks-Baird/Self-Driving-Lab-Demo: V0.8.2*, Zenodo, 2023.
56 C. Tools, *3D Printshow 2014 London - ZMorph - Cake and Chocolate 3D Printer V01*, 2014.
57 Cjp24, *English: Three-Point Flexural Test on a Composite Beam at Speed = 10 mm/min. Instron Universal Testing Machine with a 300 kN Dynamometer*, 2009.
58 F. C. S. G. CNR-ICMATE, Giovanna Canu, *English: Vacuum Arc-Melting Is a Rapid and Efficient Melting Process for Production of Metallic Small Samples with a Relevant Chemical Homogeneity. A Vacuum Arc Remelting (VAR) Furnace (Edmund Buhler) is Used, Probably a Compact Arc Melter MAM-1*, 19 December 2023, 15:53:37.
59 M. Politi, F. Baum, K. Vaddi, E. Antonio, J. Vasquez, B. P. Bishop, N. Peek, V. C. Holmberg and L. D. Pozzo, *Digital Discovery*, 2023, **2**, 1042–1057.
60 J. Vasquez, H. Twigg-Smith, J. Tran O'Leary and N. Peek, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, USA, 2020, pp. 1–13.
61 A. K. Chaurasiya, *English: ELISA Reader with Microtiter Plate*, 26 December 2021, 12:17:20.
62 K. N. Martin, M. S. Rubsamen, N. P. Kaplan and M. P. Hendricks, *Method for Interfacing a Plate Reader Spectrometer Directly with an OT-2 Liquid Handling Robot*, 2022.
63 G. S. Ganitano, S. G. Wallace, B. Maruyama and G. L. Peterson, *Prog. Addit. Manuf.*, 2023, DOI: 10.1007/s40964-023-00480-1.
64 *Pioreactor*, https://pioreactor.com/.
65 S. G. Baird and T. D. Sparks, *Matter*, 2022, **5**, 4170–4178.
66 L. M. Roch, F. Häse, C. Kreisbeck, T. Tamayo-Mendoza, L. P. E. Yunker, J. E. Hein and A. Aspuru-Guzik, *PLoS One*, 2020, **15**, e0229862.
67 Y. Xie, C. Zhang, H. Deng, B. Zheng, J.-W. Su, K. Shutt and J. Lin, *ACS Appl. Mater. Interfaces*, 2021, **13**, 53485–53491.
68 J. M. P. Gutierrez, T. Hinkley, J. W. Taylor, K. Yanev and L. Cronin, *Nat. Commun.*, 2014, **5**, 5571.
69 J. M. Parrilla-Gutierrez, S. Tsuda, J. Grizou, J. Taylor, A. Henson and L. Cronin, *Nat. Commun.*, 2017, **8**, 1144.
70 D. Doran, M. Rodriguez-Garcia, R. Turk-MacLeod, G. J. T. Cooper and L. Cronin, *Beilstein J. Org. Chem.*, 2017, **13**, 1702–1709.
71 Y. Jiang, D. Salley, A. Sharma, G. Keenan, M. Mullin and L. Cronin, *Sci. Adv.*, 2022, **8**, eabo2626.
72 M. A. Flores-Ortiz, R. A. Gutiérrez-Márquez, R. S. Mier-Jiménez, M. M. Flores-Leonar and A. A. Granda, *Building C-C Bonds Using a Cheap Automated Synthesis Platform*, 2023.
73 D. Caramelli, D. Salley, A. Henson, G. A. Camarasa, S. Sharabi, G. Keenan and L. Cronin, *Nat. Commun.*, 2018, **9**, 3406.
74 J. S. Manzano, W. Hou, S. S. Zalesskiy, P. Frei, H. Wang, P. J. Kitson and L. Cronin, *Nat. Chem.*, 2022, **14**, 1311–1318.
75 P. S. Gromski, J. M. Granda and L. Cronin, *Trends Chem.*, 2020, **2**, 4–12.
76 C.-E. Rosa, F.-N. Jorge, G.-M. Luis, C.-E. Juana and P.-D. Edgar, *HardwareX*, 2023, **14**, e00406.
77 E. C. Lee, J. M. Parrilla-Gutierrez, A. Henson, E. K. Brechin and L. Cronin, *Matter*, 2020, **2**, 649–657.
78 F. Häse, T. Tamayo-Mendoza, C. Boixo, J. Romero, L. Roch and A. Aspuru-Guzik, *Autonomous Titration for Chemistry Classrooms: Preparing Students for Digitized Chemistry Laboratories*, 2020.
79 A. Pomberger, N. Jose, D. Walz, J. Meissner, C. Holze, M. Kopczynski, P. Müller-Bischof and A. Lapkin, *Chem. Eng. J.*, 2023, **451**, 139099.
80 N. Famularo, Y. Kholod and D. Kosenkov, *J. Chem. Educ.*, 2016, **93**, 175–181.
81 F. Yang, V. Lai, K. Legard, S. Kozdras, P. L. Prieto, S. Grunert and J. E. Hein, *J. Chem. Educ.*, 2021, **98**, 876–881.
82 *Opentrons/Lab Automation/Lab Robots for Life Scientists*, https://opentrons.com/.
83 S. G. Baird and T. D. Sparks, *Build Instructions for Closed-loop Spectroscopy Lab: Light-mixing Demo*, 2023.
84 T. Ginsburg, K. Hippe, R. Lewis, D. Ozgulbas, A. Cleary, R. Butler, C. Stone, A. Stroka and I. Foster, *Exploring Benchmarks for Self-Driving Labs Using Color Matching*, 2023.

85 C. F. Carter, H. Lange, S. V. Ley, I. R. Baxendale, B. Wittkamp, J. G. Goode and N. L. Gaunt, *Org. Process Res. Dev.*, 2010, **14**, 393–404.
86 T. Brodmann, P. Koos, A. Metzger, P. Knochel and S. V. Ley, *Org. Process Res. Dev.*, 2012, **16**, 1102–1113.
87 C. Tully, *How Much Does an NMR Cost/2021 Guide/Ansazai Instruments*, 2020, https://www.aiinmr.com/nmr-spectroscopy-q-a-blog/How-much-does-an-Eft-NMR-Cost/.
88 J. Li, S. G. Ballmer, E. P. Gillis, S. Fujii, M. J. Schmidt, A. M. E. Palazzolo, J. W. Lehmann, G. F. Morehouse and M. D. Burke, *Science*, 2015, **347**, 1221–1226.
89 D. Pierre, *Undergraduate Journal of Mathematical Modeling: One + Two*, 2019, **10**(1), 8.
90 R. Soong, K. Agmata, T. Doyle, A. Jenne, A. Adamo and A. J. Simpson, *J. Chem. Educ.*, 2019, **96**, 1497–1501.
91 Y. Kosenkov and D. Kosenkov, *J. Chem. Educ.*, 2021, **98**, 4067–4073.
92 F. Garcia-Loro, P. Plaza, B. Quintana, E. S. Cristobal, R. Gil, C. Perez, M. Fernandez and M. Castro, *2021 IEEE Global Engineering Education Conference (EDUCON)*, Vienna, Austria, 2021, pp. 903–909.
93 R. Vescovi, T. Ginsburg, K. Hippe, D. Y. Ozgulbas, C. Stone, A. Stroka, R. Butler, B. J. Blaiszik, T. Brettin, K. Chard, M. Hereld, A. Ramanathan, R. Stevens, A. Vriza, J. Xu, Q. Zhang and I. Foster, *Digital Discovery*, 2023, **2**, 1980–1998.
94 R. Keesey, R. LeSuer and J. Schrier, *HardwareX*, 2022, **12**, e00319.
95 A. Faiña, B. Nejati and K. Stoy, *Appl. Sci.*, 2020, **10**, 814.
96 miLAB IDC, *OpenLH: Open Liquid-Handling System for Creative Experimentation with Biology*, 2018.
97 *Open-Source Liquid Handler/OTTO*, https://openliquidhandler.com/.
98 S. Eggert, P. Mieszczanek, C. Meinert and D. W. Hutmacher, *HardwareX*, 2020, **8**, e00152.
99 Z. Li, M. A. Najeeb, L. Alves, A. Z. Sherman, V. Shekar, P. Cruz Parrilla, I. M. Pendleton, W. Wang, P. W. Nega, M. Zeller, J. Schrier, A. J. Norquist and E. M. Chan, *Chem. Mater.*, 2020, **32**, 5650–5663.
100 A. J. Norquist, G. Jones-Thomson, K. He, T. Egg and J. Schrier, *J. Chem. Educ.*, 2023, **100**, 3445–3453.
101 K. Vaddi, H. T. Chiang and L. D. Pozzo, *Digital Discovery*, 2022, **1**, 502–510.
102 D. Salley, J. S. Manzano, P. J. Kitson and L. Cronin, *ACS Cent. Sci.*, 2023, **9**, 1525–1537.
103 D. S. Salley, G. A. Keenan, D.-L. Long, N. L. Bell and L. Cronin, *ACS Cent. Sci.*, 2020, **6**, 1587–1593.
104 D. Salley, G. Keenan, J. Grizou, A. Sharma, S. Martín and L. Cronin, *Nat. Commun.*, 2020, **11**, 2771.
105 T. Minato, D. Salley, N. Mizuno, K. Yamaguchi, L. Cronin and K. Suzuki, *J. Am. Chem. Soc.*, 2021, **143**, 12809–12816.
106 D. J. Kowalski, C. M. MacGregor, D.-L. Long, N. L. Bell and L. Cronin, *J. Am. Chem. Soc.*, 2023, **145**, 2332–2341.
107 V. Sans, L. Porwol, V. Dragone and L. Cronin, *Chem. Sci.*, 2015, **6**, 1258–1264.
108 V. Dragone, V. Sans, A. B. Henson, J. M. Granda and L. Cronin, *Nat. Commun.*, 2017, **8**, 15733.
109 L. Porwol, D. J. Kowalski, A. Henson, D.-L. Long, N. L. Bell and L. Cronin, *Angew. Chem., Int. Ed.*, 2020, **59**, 11256–11261.
110 J. M. Granda, L. Donina, V. Dragone, D.-L. Long and L. Cronin, *Nature*, 2018, **559**, 377–381.
111 D. Caramelli, J. M. Granda, S. H. M. Mehr, D. Cambié, A. B. Henson and L. Cronin, *ACS Cent. Sci.*, 2021, **7**, 1821–1830.
112 P. J. Kitson, S. Glatzel and L. Cronin, *Beilstein J. Org. Chem.*, 2016, **12**, 2776–2783.
113 P. J. Kitson, S. Glatzel, W. Chen, C.-G. Lin, Y.-F. Song and L. Cronin, *Nat. Protoc.*, 2016, **11**, 920–936.
114 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.
115 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.
116 M. Bornemann-Pfeiffer, J. Wolf, K. Meyer, S. Kern, D. Angelone, A. Leonov, L. Cronin and F. Emmerling, *Angew. Chem., Int. Ed.*, 2021, **60**, 23202–23206.
117 K. Hippalgaonkar, Q. Li, X. Wang, J. W. Fisher, J. Kirkpatrick and T. Buonassisi, *Nat. Rev. Mater.*, 2023, **8**, 241–260.
118 P. J. Kitson, G. Marie, J.-P. Francoia, S. S. Zalesskiy, R. C. Sigerson, J. S. Mathieson and L. Cronin, *Science*, 2018, **359**, 314–319.
119 A. Bubliauskas, D. J. Blair, H. Powell-Davies, P. J. Kitson, M. D. Burke and L. Cronin, *Angew. Chem., Int. Ed.*, 2022, **61**, e202116108.
120 R. J. Hickman, M. Aldeghi, F. Häse and A. Aspuru-Guzik, *Digital Discovery*, 2022, **1**, 732–744.
121 *Bluesky — An Experiment Specification & Orchestration Engine*, 2023, https://github.com/bluesky/bluesky, accessed 2023-03-10.
122 H. Xu, Y. R. Wang, S. Eppel, A. Aspuru-Guzik, F. Shkurti and A. Garg, *Seeing Glass: Joint Point Cloud and Depth Completion for Transparent Objects*, 2021.
123 E. Olson, *2011 IEEE International Conference on Robotics and Automation*, Shanghai, China, 2011, pp. 3400–3407.
124 J. Wang and E. Olson, *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejeon, South Korea, 2016, pp. 4193–4198.
125 Y. R. Wang, Y. Zhao, H. Xu, S. Eppel, A. Aspuru-Guzik, F. Shkurti and A. Garg, *MVTrans: Multi-View Perception of Transparent Objects*, 2023.
126 M. Krogius, A. Haggenmiller and E. Olson, *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China, 2019, pp. 1898–1903.
127 P. Jana and M. Tiwari, in *Lean Tools in Apparel Manufacturing*, ed. P. Jana and M. Tiwari, Woodhead Publishing, 2021, pp. 17–45.
128 J. Rodriguez, M. Politi, S. Adler, D. Beck and L. Pozzo, *Mol. Syst. Des. Eng.*, 2022, **7**, 933–949.
129 L. Cao, D. Russo, K. Felton, D. Salley, A. Sharma, G. Keenan, W. Mauer, H. Gao, L. Cronin and A. A. Lapkin, *Cell Rep. Phys. Sci.*, 2021, **2**, 100295.

130 K. J. Lachowski, K. Vaddi, N. Y. Naser, F. Baneyx and L. D. Pozzo, *Digital Discovery*, 2022, **1**, 427–439.
131 C.-L. Chen, T.-R. Chen, S.-H. Chiu and P. L. Urban, *Sens. Actuators, B*, 2017, **239**, 608–616.
132 *METTLER TOLEDO High Temp Differential Scanning Calorimeter (DSC) - Specialty Lab Equipment, Instruments, and Apparatuses, Calorimetry*, https://www.fishersci.com/shop/products/high-temp-differential-scanning-calorimeter-dsc/p-7139360.
133 J. Rodriguez, M. Politi, S. Scheiwiller, S. Bonageri, S. Adler, D. Beck and L. D. Pozzo, *Journal of Open Hardware*, 2021, **5**, 6.
134 M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson and E. Bakshy, *BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization*, 2020.
135 A. Palizhati, S. B. Torrisi, M. Aykol, S. K. Suram, J. S. Hummelshøj and J. H. Montoya, *Sci. Rep.*, 2022, **12**, 4694.
136 Y. Jin and P. V. Kumar, *Nanoscale*, 2023, **15**, 10975–10984.
137 D. A. Boiko, R. MacKnight and G. Gomes, *Emergent Autonomous Scientific Research Capabilities of Large Language Models*, 2023.
138 *IBM RoboRXN/Science/IBM Research*, 2021, https://research.ibm.com/science/ibm-roborxn/.
139 *Emerald Cloud Lab: Remote Controlled Life Sciences Lab*, https://www.emeraldcloudlab.com/.
140 *Robotic Controlled Life Sciences Lab - Strateos Cloud Lab*, https://strateos.com/.
141 *Culture Biosciences Home/Bioreactors in the Cloud*, https://www.culturebiosciences.com/.
142 *Home - Arctoris*, https://dev.arctoris.com/.
143 *Kebotix*, https://www.kebotix.com.
144 *Atinary Technologies Inc.*, https://atinary.com/.
145 S. Castellanos, *Carnegie Mellon's Cloud Lab to Automate Labor-Intensive Science Experiments*, https://www.wsj.com/articles/carnegie-mellons-cloud-lab-to-automate-labor-intensive-science-experiments-11630348526.
146 M. Taylor, *Women in Science: The $40 M Cloud Investment That COVID-19 Proved Was Needed*, http://www.laboratoryequipment.com/582482-Women-in-Science-The-40-M-Cloud-Investment-That-COVID-19-Proved-Was-Needed/.
147 C. Arnold, *Nature*, 2022, **606**, 612–613.
148 T. Ireland, *The Observer*, 2022.
149 *Carnegie Mellon University and Emerald Cloud Lab to Build World's First University Cloud Lab - News - Carnegie Mellon University*, 2021, http://www.cmu.edu/news/stories/archives/2021/august/first-academic-cloud-lab.html.
150 M. Thomas, *Learn about Carnegie Mellon's $40 Million Life Science 'Gamble'*, 2021, https://www.ddw-online.com/learn-about-carnegie-mellons-40-million-life-science-gamble-14670-202111/.
151 C. Armer, F. Letronne and E. DeBenedictis, *PLoS Biol.*, 2023, **21**, e3001919.
152 A. A. Hunter, MA Thesis, Carleton University, 2021.
153 J. Silvertown, *Trends Ecol. Evol.*, 2009, **24**, 467–471.
154 R. F. Lei, E. R. Green, S.-J. Leslie and M. Rhodes, *Dev. Sci.*, 2019, **22**, e12837.
155 E. Tegler, *Forbes*, 2021.
156 Future of Life Institute, *Pause Giant AI Experiments: An Open Letter*, 2023.
157 A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White and P. Schwaller, *ChemCrow: Augmenting Large-Language Models with Chemistry Tools*, 2023.
158 R. Sohn, AI Drug Discovery Systems Might Be Repurposed to Make Chemical Weapons, Researchers Warn, *Sci. Am.*, 2022, https://www.scientificamerican.com/article/ai-drug-discovery-systems-might-be-repurposed-to-make-chemical-weapons-researchers-warn/, accessed 2022-11-19.
159 E. Falletti and C. Gallese, *4th European Conference on the Impact of Artificial Intelligence and Robotics*, Oxford, UK, 2022.
160 N. Chakravorty, C. S. Sharma, K. A. Molla and J. K. Pattanaik, *Proc. Indian Natl. Sci. Acad.*, 2022, **88**, 456–471.
161 J. A. Smith and J. B. Sandbrink, *PLoS Biol.*, 2022, **20**, e3001600.
162 A. Dave, J. Mitchell, K. Kandasamy, H. Wang, S. Burke, B. Paria, B. Póczos, J. Whitacre and V. Viswanathan, *Cell Rep. Phys. Sci.*, 2020, **1**, 100264.
163 S. Matsuda, G. Lambard and K. Sodeyama, *Cell Rep. Phys. Sci.*, 2022, **3**, 100832.
164 I. Oh, M. A. Pence, N. G. Lukhanin, O. Rodríguez, C. M. Schroeder and J. Rodríguez-López, *Device*, 2023, **1**(5), 1–11.
165 K. M. Jablonka, G. M. Jothiappan, S. Wang, B. Smit and B. Yoo, *Nat. Commun.*, 2021, **12**, 2312.
166 A. G. Kusne and A. McDannald, *Matter*, 2023, **6**, 1880–1893.
167 S. V. Kalinin, M. Ziatdinov, J. Hinkle, S. Jesse, A. Ghosh, K. P. Kelley, A. R. Lupini, B. G. Sumpter and R. K. Vasudevan, *ACS Nano*, 2021, **15**, 12604–12627.
168 S. V. Kalinin, D. Mukherjee, K. M. Roccapriore, B. Blaiszik, A. Ghosh, M. A. Ziatdinov, A. Al-Najjar, C. Doty, S. Akers, N. S. Rao, J. C. Agar and S. R. Spurgeon, *Deep Learning for Automated Experimentation in Scanning Transmission Electron Microscopy*, 2023.
169 C. Badue, R. Guidolini, R. V. Carneiro, P. Azevedo, V. B. Cardoso, A. Forechi, L. Jesus, R. Berriel, T. M. Paixão, F. Mutz, L. de Paula Veronese, T. Oliveira-Santos and A. F. De Souza, *Expert Systems with Applications*, 2021, **165**, 113816.
170 V. Shreyas, S. N. Bharadwaj, S. Srinidhi, K. U. Ankith and A. B. Rajendra, in *Advances in Data and Information Sciences*, ed. M. L. Kolhe, S. Tiwari, M. C. Trivedi and K. K. Mishra, Springer, Singapore, 2020, vol. 94 of Lecture Notes in Networks and Systems, pp. 361–371.
171 J. Beal and M. Rogers, *Mol. Syst. Biol.*, 2020, **16**, e10019.
172 R. D. King, K. E. Whelan, F. M. Jones, P. G. K. Reiser, C. H. Bryant, S. H. Muggleton, D. B. Kell and S. G. Oliver, *Nature*, 2004, **427**, 247–252.
173 T. Si, R. Chao, Y. Min, Y. Wu, W. Ren and H. Zhao, *Nat. Commun.*, 2017, **8**, 15187.

174 M. HamediRad, R. Chao, S. Weisberg, J. Lian, S. Sinha and H. Zhao, *Nat. Commun.*, 2019, **10**, 5150.

175 G. N. Kanda, T. Tsuzuki, M. Terada, N. Sakai, N. Motozawa, T. Masuda, M. Nishida, C. T. Watanabe, T. Higashi, S. A. Horiguchi, T. Kudo, M. Kamei, G. A. Sunagawa, K. Matsukuma, T. Sakurada, Y. Ozawa, M. Takahashi, K. Takahashi and T. Natsume, *eLife*, 2022, **11**, e77007.

176 N. Hillson, M. Caddick, Y. Cai, J. A. Carrasco, M. W. Chang, N. C. Curach, D. J. Bell, R. Le Feuvre, D. C. Friedman, X. Fu, N. D. Gold, M. J. Herrgård, M. B. Holowko, J. R. Johnson, R. A. Johnson, J. D. Keasling, R. I. Kitney, A. Kondo, C. Liu, V. J. J. Martin, F. Menolascina, C. Ogino, N. J. Patron, M. Pavan, C. L. Poh, I. S. Pretorius, S. J. Rosser, N. S. Scrutton, M. Storch, H. Tekotte, E. Travnik, C. E. Vickers, W. S. Yew, Y. Yuan, H. Zhao and P. S. Freemont, *Nat. Commun.*, 2019, **10**, 2040.

177 R. Chao, J. Liang, I. Tasan, T. Si, L. Ju and H. Zhao, *ACS Synth. Biol.*, 2017, **6**, 678–685.

178 *Robot Advantage: Drug Discovery Faster, Cheaper, Better & Error Free – Asian Robotics Review*, **https://asianroboticsreview.com/home77-html**.

179 L. C. Gerber, A. Calasanz-Kaiser, L. Hyman, K. Voitiuk, U. Patil and I. H. Riedel-Kruse, *PLoS Biol.*, 2017, **15**, e2001413.

180 A. R. Akbashev and S. V. Kalinin, *Nat. Mater.*, 2023, **22**, 270–271.