ADVANCED INTELLIGENT SYSTEMS Open Access
www.advintellsyst.com

Research Acceleration in Self-Driving Labs: Technological Roadmap toward Accelerated Materials and Molecular Discovery

*Fernando Delgado-Licona and Milad Abolhasani**

The urgency of finding solutions to global energy, sustainability, and healthcare challenges has motivated rethinking of the conventional chemistry and material science workflows. Self-driving labs, emerged through integration of disruptive physical and digital technologies, including robotics, additive manufacturing, reaction miniaturization, and artificial intelligence, have the potential to accelerate the pace of materials and molecular discovery by 10–100X. Using autonomous robotic experimentation workflows, self-driving labs enable access to a larger part of the chemical universe and reduce the time-to-solution through an iterative hypothesis formulation, intelligent experiment selection, and automated testing. By providing a data-centric abstraction to the accelerated discovery cycle, in this perspective article, the required hardware and software technological infrastructure to unlock the true potential of self-driving labs is discussed. In particular, process intensification as an accelerator mechanism for reaction modules of self-driving labs and digitalization strategies to further accelerate the discovery cycle in chemical and materials sciences are discussed.

1. Introduction

Exponential growth is characterized by improvements in a key aspect of a process or device that follows a power law over a fixed time interval. The most famous example is Moore's Law stating that the number of transistors within an integrated circuit will double every 2 years.[1] A similar acceleration has been extended to other fields (accelerated returns),[2] for instance, the number of parameters of neural network models,[3] the Taniguchi chart for maximum achievable machining accuracy,[4] and the Carlson curve in biotechnology.[5] There is a growing expectation that advanced materials and specialty/fine chemicals will soon experience an exponential acceleration in their pace of discovery and sustainable manufacturing. This promise has been fueled by the recent integration of rapidly emerging technologies, including robotics,[6,7] advanced manufacturing,[8–10] and data science[11–15] with chemical and materials sciences.

Global challenges that require our immediate attention, such as climate change and rapidly rising energy demands, further amplify the need for research acceleration in chemical and materials sciences. The need to fast-track finding a solution to a global problem presented itself during the discovery and manufacturing of the vaccine against the COVID-19 virus, which, among other mitigating efforts, became a priority to restore normality and decrease global mortality rates.[16] Hence, time-to-solution is a critical measure to assess damages that global challenges could incur on society, the environment, and the global economy. To achieve the shortest time-to-solution, research acceleration is imperative. Despite the challenges, in the case of COVID-19 pandemic, the pharmaceutical industry was able to accelerate vaccine research by 4X, when compared to the fastest vaccine discovery effort to date,[17] thanks to a strong technological infrastructure (e.g., DNA sequencing, mRNA technology, bio-manufacturing) and an impressive global collaboration.[16,17]

To achieve the targeted 1.5–2 °C average global temperature rise by the year 2100[18] and avoid the impending catastrophic consequences in the near future, we need a similar global collaboration using an accelerated research infrastructure. Beyond policy agreements and international pledges, high-performing functional materials with cost-effective and energy efficient devices are needed as a gateway to a carbon-neutral future.[19] For example, grid decarbonization requires new high-performing clean energy materials and sustainable processes for affordable and scalable renewable energy applications.[20]

Accelerated materials and molecular discovery (AMMD) promises 10–100X reduction of energy and chemical consumption, processing time, risk of operation, and capital investment within a typical discovery cycle that involves 1) specification of a research question, 2) automated collection of relevant existing data, 3) formation of a hypothesis, 4) autonomous (closed-loop) experimentation and testing of hypothesis, and 5) communication of results and generated knowledge.[21] Confronted with the most decisive environmental crisis in recent years, climate

F. Delgado-Licona, M. Abolhasani
Department of Chemical and Biomolecular Engineering
North Carolina State University
911 Partners Way, Raleigh, NC 27695-7905, USA
E-mail: abolhasani@ncsu.edu

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aisy.202200331.

DOI: 10.1002/aisy.202200331

change has driven most of the materials discovery initiatives worldwide.[22] Accelerated materials discovery could also contribute to addressing worldwide environmental concerns such as resource scarcity and plastic pollution, by rethinking materials management, conventional chemical processing, and linear production paradigms.

Research acceleration in chemical and materials sciences has been recognized as an emerging field in the past 5 years.[23–25] Research acceleration is achieved through the use of artificial intelligence (AI), robotics, lab automation, and parallel batch or flow reactors in a closed-loop format.[26,27] The resulting closed-loop platform which enables accelerated experimentation and hypothesis testing is called a "self-driving lab (SDL)." SDLs involve streamlined and autonomous robotic experimentation using an integrated experimental workflow, operated continuously with faster time-to-solution when compared to a human researcher in a conventional chemistry and materials science research setting (**Figure 1**).[23,28] Two aspects of automation in the research acceleration by SDLs include 1) automation of experiments (physical unit operations and characterization units) and 2) automation of experiment–selection process (with the help of machine-learning algorithms). Physical automation of experiments using the same existing hardware found in conventional experimental workflows in chemical and material sciences (e.g., using a robotic arm instead of a human scientist to perform different experimental tasks) could effectively reduce experimentation time by at least 2X. Such automation translates to a 2X increase in the data-generation throughput compared to manual experimentation, furthermore, it allows the highly trained scientist to focus on other discovery steps, such as formulating research inquiries and hypothesis. While process automation and combinatorial screening provide reproducibility, process standardization, cost reduction, and parallelization capabilities, it has been argued that automation alone is not sufficient to support AMMD at a pace required to address the global challenges.[29] Relying solely on physical automation provides no flexibility for the system to adapt to unexpected process conditions, therefore, all preparations should be performed prior to the start of automated experiments by the expert scientist. However, SDLs, by integrating physical and digital lab automation, provide a framework for the development of synchronous, interoperable, and standard autonomous systems that could facilitate common processing tasks—error and exception handling, fault corrections, cleaning protocols, failure, and anomaly detection—by removing the necessity of "hardcoding" precise action sequences, and allowing the SDL learn these and other related tasks in an autonomous fashion, with reduced human intervention before, during, and after an experiment is performed. Consequently, SDLs are coupled with advanced AI strategies, such as active learning, to provide an additional 5–20X acceleration factor in addition to the 2X obtained by physical automation in chemical and materials sciences by intelligently navigating the exponentially increasing reaction and processing parameter spaces, avoiding exhaustive combinatorial searches, integrating cleaning, failure, error, and exception protocols, while significantly reducing reagent consumption.[30,31] For instance, Burger et al. reported an autonomous mobile robot performing photocatalyst optimization with 1000X faster throughput compared to manual methods, and 10X when compared to semiautomated workflows.[32]

It should also be noted that optimizing the automation process for both the physical and digital aspects of SDLs is crucial to reduce the reconfiguration time needed for exploring a new class of materials/molecules by the same SDL. Such optimization involves approaching automation from modular and flexible perspectives. Currently, it takes 3–5 years to build a reliable SDL for complex chemistries, and an additional 1–2 years to upgrade an SDL for another class of materials/molecules. This long development timeframe of SDLs and their high cost are considered as two of the most important SDL entry barriers for researchers.

Early proof-of-concept demonstrations of SDLs in chemical and materials sciences have been successful in illustrating the research acceleration by employing closed-loop autonomous experimentation strategies enabled by integration of automated experimentation and automated experiment-selection.[33,34] While most of the focus in the emerging field of SDLs has been centered around robotic automation and AI, to have a better understanding of the evolution and potential of AMMD, it is important to understand its core technologies. Innovation and synergy over a range of technologies usually coincide with system-wide improvements and subsequent exponential breakthroughs.[35] Thus, a true research acceleration in chemical and materials sciences could only be achieved when the technological infrastructure to support the five iterative discovery steps are developed synergistically.

Understanding the technological infrastructure that powers AMMD efforts will offer a valuable perspective toward building an ideal SDL for accelerating finding solutions to global challenges in energy, sustainability, and healthcare. With this goal, in this perspective article, we discuss AMMD by using SDLs and

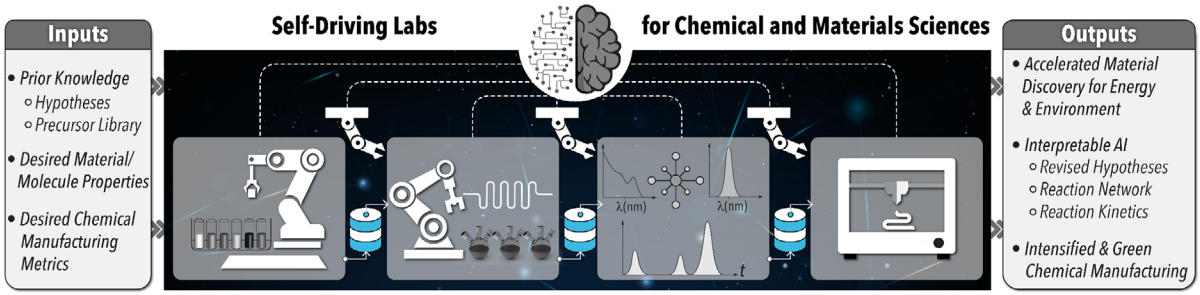

Figure 1. Schematic representation of a modular self-driving lab with inputs/outputs as well as the data flow between different modules for closed-loop hypothesis formulation and testing.

their role to facilitate data flow and processes, ultimately guiding the current technological infrastructure that we believe realizes these processes at each discovery step. We highlight the advantages and limitations of each core technology of SDLs and detail their larger role in academic and industrial research settings. We then discuss a roadmap to the realization of an accelerated SDL through the integration of the discussed technologies.

2. AMMD: Technological Infrastructure

Data processes are the foundation of AMMD. When the discovery cycle is analyzed under this abstraction, the technological roadmap toward continuous improvements becomes evident. A summary of the data processes for each discovery step is summarized in **Table 1**.

Every piece of hardware and software that facilitates the tasks listed in Table 1 belongs to the technological infrastructure that has the potential to complement human knowledge and intuition, thus, accelerating the pace and throughput of research efforts in chemical and materials sciences. Furthermore, strategies that strengthen the interconnection and data flow between two or more steps of the discovery cycle, a process described earlier as "closing the loop," increase the acceleration factor in discovering new materials and molecules. Self-driving labs play a central role in AMMD; they function as the bridge between bits and atoms, from software and data to hardware and materials. Over the past 5 years, SDLs have successfully demonstrated the benefit of closing the loop between hypothesis formation and autonomous experimentation.[36,37] The closed-loop experimentation has been possible thanks to the coordination of software and hardware technologies contributing to the aforementioned data processes (**Figure 2**). For example, most of the hypothesis formulations performed autonomously by SDLs consist of suggesting new parameters for testing by the experimental modules of the SDLs. In human language, the formulated hypothesis might take the form of "Does parameter x generate the target material with the highest discovered characteristic a?". Such hypotheses are automatically generated by data-driven predictive models trained and built using previous data generated by the SDL, followed by proposing new experiments given the best-predicted outcome to test and thus validate or invalidate the generated hypothesis in an iterative manner.[33]

Table 1. Data processes of each discovery step.

Discovery Step	Data Processes
I – Specification of research questions	Storage/Retrieval/knowledge modeling
II – Automated collection of relevant existing data	Extraction/scraping
III – Formation of a hypothesis	Analysis/modeling/prediction
IV – Autonomous experimentation and hypothesis testing	Scheduling/generation/acquisition/processing/analysis/modeling
V – Communication of results and generated knowledge	Aggregation/representation/standardization interoperability/democratization/reporting

Technologies that contribute to data-generation processes related to AMMD could be divided into 1) reactive strategies and 2) in silico strategies. In silico strategies are experiments that develop virtually using any computation machine, for example, digital computers.[38] Quantum chemistry simulations offer highly accurate models; however, even moderate-sized molecules (100 electrons) surpass current digital computational resources rendering them infeasible to compute.[39] To circumvent these limitations, other computational strategies have been developed, most noticeably AI methods, for example, by introducing graph-based neural network architectures, Qiao et al. have predicted energy solutions to the Schrödinger equation matching the accuracy of density-functional theory (DFT) while reducing the computational cost by 1000X.[40] Technologies surrounding in silico experimentation continue to advance by providing computationally affordable algorithms for molecule generation such as generative-adversarial-, transformer-, and diffusion-based models[41–43] or by developing innovative ways to perform computations such as analog and quantum computers.[44,45] A thorough review of other strategies that aim to intensify time-to-solution in this context can be found elsewhere.[46,47]

Intensification of data processes is expected to result in sustained research acceleration in chemical and materials sciences using SDLs. While operating in a completely digital world would offer enhanced acceleration, materializing such discoveries would imply having SDLs that provide precise control over the required chemical environment, while being reliable, reproducible, flexible, and universal. Ultimately, AMMD is a framework to solve physical worldwide problems (and beyond, e.g., outer space).[48] Thus, most state-of-the-art SDLs have bypassed the feasibility challenge of translating digital results into measurable physical data, through iterating over an automated experimental platform. In the context of reactive systems, the reaction module of SDLs could be considered their most important component due to its role in generating the data required for AMMD. SDLs have access to the technological infrastructure that allows for intensification of their data-generation processes similar to in silico strategies, but with the added physicality constraints such as material scarcity and toxic or harmful side products. The leading strategy to intensify the data-generation process of the reaction modules of SDLs is called process intensification (PI). We believe that PI is a core enabler for AMMD efforts, however, a formal rethinking of its fundamental approaches to better fit subsequent data processes is required. In the following subsection, we introduce PI and discuss how it can enhance research productivity of SDLs.

2.1. Process Intensification

PI is defined as disruptive methods that drastically improve how chemicals are produced in terms of processing time, energy, material consumption, and waste generation as well as environmental and operational safety. PI strategies consist of four principles 1) maximizing the effectiveness of inter-, and intramolecular events; 2) providing homogeneous physical and chemical environment; 3) optimizing the driving forces of the process and enhancing the specific surface area; and 4) maximizing the synergy of unit operations.[49] When implemented collectively, the resulting intensified reaction module of SDLs becomes

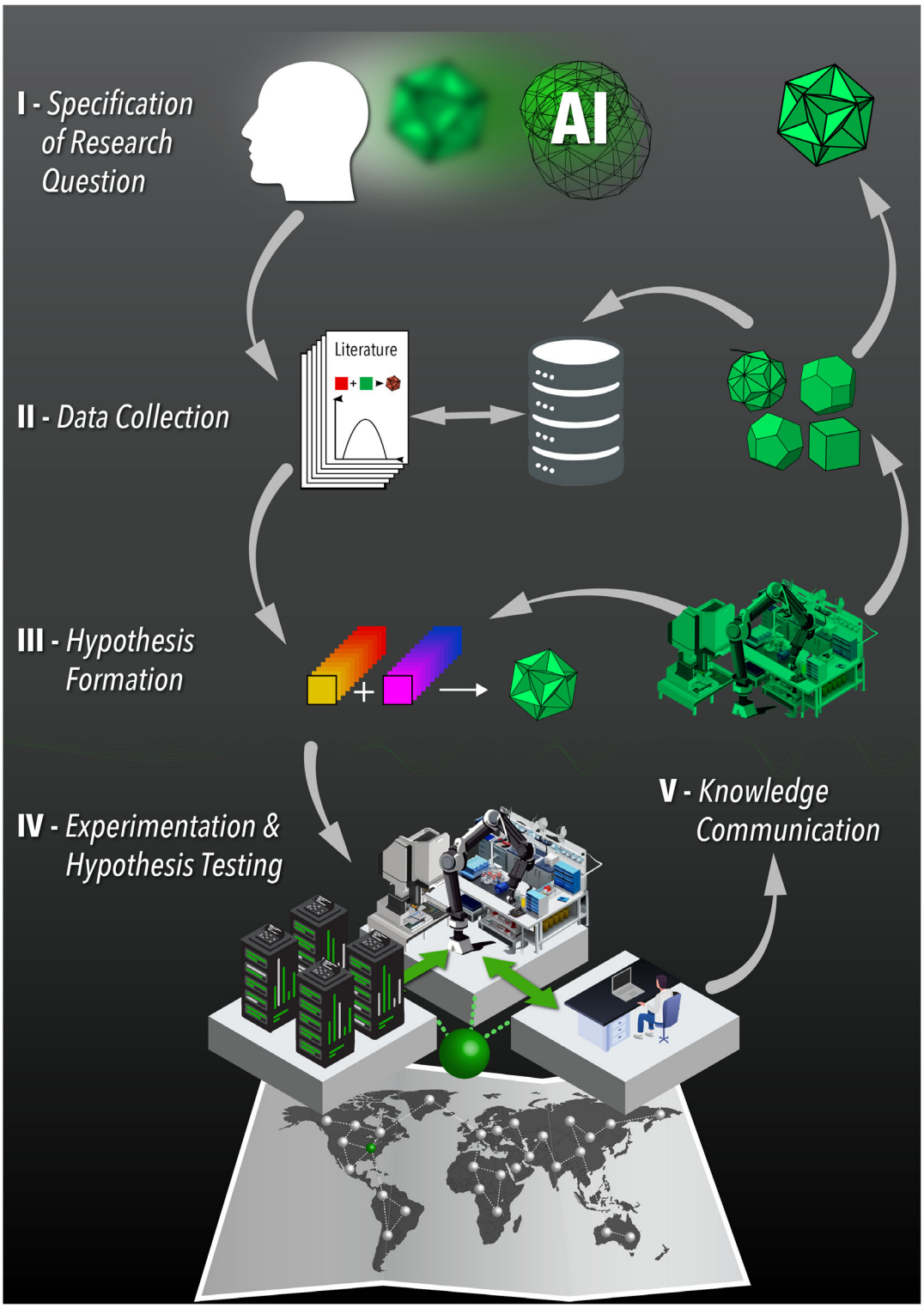

Figure 2. Schematic illustration of a self-driving lab (SDL) that has closed the loop on all discovery stages (I–V), further accelerating the research throughput and pace of discovery.

closer to an optimal data-generator for subsequent AI-assisted modeling and decision-making steps. The four principles of PI have been successfully demonstrated in isolation (i.e., only for a reactive process without closing the loop) using various fundamental approaches (structure, energy, time, synergy) at the molecular-, meso-, and macro-scales.[49]

The best-known instance for intensifying a reactive system through the "structure" approach is miniaturization. Milli- and micro-fluidic reactors benefit from superior heat, mass, and photon transfer rates, due to enhanced surface-to-volume ratios when compared to conventional batch reactors with 10–100X larger characteristic length scales. Such enhanced transport rates in flow reactors are extremely beneficial for AMMD applications, by converging the relevant reactions to their true kinetics (via removing mass, heat, or photon transfer rate limitations), further closing the gap toward an optimal data-generation scheme for SDLs. However, given the reduced characteristic length scale of flow reactors (<1 mm), they are prone to fouling and clogging when handling solids.[50] Small-scale fluid transport is dominated by interfacial and viscous forces (instead of inertial and gravitational forces), leading to possible long diffusion times and unwanted Taylor dispersion. To reduce such challenges in small-scale flow reactors, various reaction miniaturization implementations have integrated strategies such as multiphase flow,[51] static micromixers,[52] and intensified reactor designs including Taylor–Couette,[53] spinning disk reactor,[54] cascade of continuous miniaturized stirred tank reactors,[55] among others,[56] to enhance the performance of the reactive modules up to 100X in terms of chemical consumption and total processing time when compared to batch reactors.

It is important to note that gravitational forces start to become negligible as reactor channels become smaller than 1 mm in microfluidic devices.[51] Thus, microfluidic systems have been promoted as a prime tool for engineering complex space and exoplanetary reactive systems,[57] further supporting PI as a promising framework for the development of universal reactive modules for SDLs. Further intensification in the spatial domain (structure approach) could maximize the effectiveness of molecular events by providing a homogenous physical and chemical environment in the reactive modules of SDLs. For example, the interplay between microreactor geometry and its resulting hydrodynamic profile has been identified as a key parameter for tuning nanoparticle size, morphology, and composition distributions.[58]

The experimental acceleration involves two strategies: 1) performing multiple experiments simultaneously (parallelization), or 2) performing experiments in a shorter amount of time than the conventional experiments by kinetic acceleration enabled via reaction and reactor engineering. Miniaturized and parallelized material synthesis strategies fall under the first category, while reaction intensification through enhanced heat and mass transfer rates of flow reactors fall under the second category of experimental intensification strategies. Beyond fluidic devices, miniaturization has accelerated the fabrication and exploration of solid-state materials for electronic devices, that is, semiconductors, ferroelectric, and ferromagnetic materials, through high-throughput experiments, and combinatorial synthesis in small areas ($<1\,mm^2$).[59] Similar miniaturization concept has also been utilized for co-deposition of multinary thin films, where uniform nanoscale reactor volumes were formed to leverage fast reactions due to small diffusion length scales.[60] Intensification of reactive systems for sustainable chemical synthesis has taken advantage of the rest of the fundamental PI approaches. For example, localized and uniform photon transport has intensified the photo-induced production of organic compounds and nanoparticles up to 100X when compared to conventional batch reactors.[61,62] The energy approach to PI has accelerated other processes in material sciences by precisely and optimally delivering electric, magnetic (e.g., microwaves), acoustic (e.g., ultrasound), and gravitational fields.[63] For example, ternary alloy thin-film materials have been successfully synthesized using pulsed laser deposition, by precisely targeting the laser position at the surface of the thin film.[64] Similarly, a reactive system can be intensified by manipulating the time scales at which different processes occur. These timescale reductions are not limited to the already introduced shortening of mixing times or highly precise temperature–time profile tuning in multistage syntheses,[65] but include pulsations (periodicity) and the paradigm of shifting processes from batch to continuous flow. An intensified continuous process has the potential to minimize reactor down times, regularly associated with batch reactions (e.g., reactor start-up, shut down, and cleaning times) further accelerating the experiment process by 2–3X, while providing the possibility of an uninterrupted stream of reliable and reproducible data if integrated with high-throughput characterization techniques for subsequent steps of the AMMD cycle.

2.2. Process Intensification for AMMD

The successful demonstration of PI strategies for different reactive systems illustrates the role of reactor and reaction engineering in future SDLs. Correct implementation of PI strategies in the reaction module of SDLs could lead to an additional 100X research acceleration factor for AMMD efforts. Thus, data-generation acceleration could be considered as another fundamental approach to PI strategies. Such implementation would eliminate the missing link between PI and subsequent data-processing steps underlying AMMD efforts enabled by SDLs. Similar to other PI strategies, the new fundamental approach to PI should be capable of fulfilling and realizing all aforementioned PI principles.

Lopez-Guajardo et al. proposed a fifth approach for process intensification: data, that corresponds to the Knowledge domain.[66] This approach explicitly frames PI as a data-generation process, meaning that the reaction module of SDLs should extend its data-generation capabilities beyond carrying out an intensified reaction by facilitating in situ or online characterizations techniques and providing self-describing data of the network of intensified reaction modules used in the SDLs. Examples of such intensified process descriptors include reactor design/geometry, material, time- and space-resolved process parameters, classification, and graph representation that could later augment a data-driven representation of the system (digital twin).[66] Furthermore, the described data approach establishes the principles that the generated data should hold, including well-documented Industry 4.0 data principles.[67]

In situ or online product characterization provides an inherent acceleration factor as described earlier due to the self-optimization capability of SDLs. However, it is not reasonable to assert that intensification of reaction modules is attained only through in situ or online product characterization, due to lack of improvement over the four principles of PI. The realization of

research intensification in SDLs through the data approach occurs when there is an enhancement of the reaction module or network of modules, through any synergistic combination of the PI approaches described earlier (e.g., restructuring the reactor geometry, choosing a better-suited intensified reactor, introducing an energy field). This process could be defined as self-intensification, which in combination with the self-optimization capability, could have a compounding effect to achieve an exponential improvement of SDLs.[68]

We extend the importance of the data-based PI approach in SDLs by contextualizing it with further steps in the AMMD cycle and identifying it as a key complement for the digitalization strategy that has the potential to close the loop over the entire set of data processes described in Table 1.

3. Data-Centric AMMD

By effectively closing the loop between hypothesis formulation and testing, promising research acceleration factors have been achieved in both academia and industrial settings.[10,22,28,29,32,33,36] However, some of these advancements are being perceived as difficult and expensive to adopt, thereby challenging the goals of AMMD. Commercial SDLs often employ a user-friendly software to control their hardware composed of multiple robots for handling starting reactants and sample transfer, high-precision sensors, a single or parallel batch reactors, and online sampling modules for automated product characterization.[69,70] Apart from the out-of-reach adoption costs for most researchers, such fixed ecosystems with proprietary software and hardware limit the flexibility and adaptability to existing workflow architectures in chemistry and materials science labs.[71]

While initial efforts revolving SDLs in academia were focused on using only modular, off-the-shelf[72] or digitally fabricated hardware[73] with open-source software (e.g., Python),[74] reproducibility issues proved to be a key challenge. Such issues arise from the lack of effective data recording and communication surrounding the different physical and chemical environments of the reaction modules in such custom-built SDLs.[8] By utilizing modular 3D-printed reactionwares, it has been demonstrated that digitizing and communicating unit operations such as stirring, temperature profile, filtering, and sample transfer can help to close the lab-to-lab reproducibility gap in organic reactions.[8] Digitalization of chemical processes could be improved by employing intensified reaction modules. Intensified reactive modules provide a homogeneous physical and chemical environment, that maximize inter- and intramolecular events, driving forces, and synergy between unit operations. Therefore, communicating intensified process descriptors could provide valuable insights at molecular and meso- and macroscales while increasing the volume, reliability, and variety of data.[75] For example, synthesis and processing of colloidal nanocrystals, with an experimental space ranging from 10^5 to 10^{10} possible experimental conditions, are time- and material-intensive tasks, using manual experimentation and batch reactors. In one example, Abdel-Latif et al.[76] reported intensification of the synthesis and bandgap tuning of inorganic metal halide perovskite nanocrystals using two integrated microfluidic reactors. In this example, process intensification was achieved not only by reducing the chemical consumption during the synthetic route discovery campaigns, but also by bypassing the need for an intermediate nanocrystal washing step after the first reaction. An AI-assisted experiment selection strategy provided another acceleration factor to the intensified experimental platform by reducing the total experimental cost (i.e., the number of experiments to perform) for synthetic route discovery of inorganic metal halide perovskite nanocrystals. Furthermore, development and integration of in situ and online characterization techniques (sensors) such as absorption and photoluminescence spectroscopy, and chromatography, as well as small-angle and wide-angle X-ray scattering with automated experimental platforms have been critical technological infrastructure advancements to enable high-throughput data generation for SDL. Access to such a large amount of data generated by in situ characterization tools has accelerated integration of data science tools with chemical and materials sciences. Building digital twins of chemical and materials processes is one of the natural evolutions of the integration of data science tools with automated experimentation strategies equipped with in situ or online characterization tools. By providing the metadata of such intensified process parameters and descriptors, the generated knowledge is expected to be reproducible and relevant to build digital twins for fundamental and applied queries.

Data aggregation, standardization, interoperability, and democratization are captured by the last step in the discovery cycle, that is communication of results and generated knowledge; this step precedes all previous data processes and plays a defining role for a closed-loop AMMD cycle. We outline the synergistic interplay between data, machines, and humans for a true research acceleration enabled by the use of SDLs in Figure 2. Furthermore, to harness the research acceleration for AMMD derived from global collaborations, it is necessary to aggregate data from all possible data sources (data fusion), including SDLs, and provide a standard for effectively communicating data to humans and machine collaborators. A recent review on this topic by Bai et al. details the landscape of current data representations and exchange protocols that could unite isolated AMMD efforts.[71] Semantic web technologies are highlighted as a promising technological infrastructure to achieve a fully interconnected web, hence, universal interoperability across scales and domains.[71]

There are two main components in a closed-loop AMMD cycle, shown in Figure 2: 1) the physical world (Figure 2, bottom panel) where data generation takes place, and 2) the virtual world where data processes are interconnected (Figure 2, top panel). It is important to note that the virtual world, despite being presented as different levels for clarity, is a single universal entity similar to a dynamic knowledge graph, where various steps of the AMMD cycle take place. This framework represents all aspects of the real world, including an ontological representation of the physical space and interoperable agents (automated software) that are responsible for all data processes (Figure 2I–V).[77] More generally, this virtual world falls within the technological infrastructure of digital twins and the encompassing metaverse technological infrastructure.[78] While some digital twin implementations consist of building accurate process models for property predictions,[34] a thorough implementation should consider the digital representation of both hardware and software components, and a facile control/interaction between them. In one

example, Li et al. presented an integration of various physical and digital technologies, such as virtual reality, collaborative robots, and AI to accelerate optimization of the optical properties of semiconductor nanocrystals.[79] In this example, the various components, including the solid and liquid handling robots communicated through 5 G network, and the generated data was stored in a structured query language (SQL) database hosted in the cloud.[79] The key enabling factor of a closed-loop AMMD cycle is the fully interconnected, interoperable, and distributed network of SDLs. In the case of SDLs following a reactive strategy, intensified process descriptors could mitigate data representation and interoperability issues that arise when two different groups share data generated by different reactive platforms. Since intensified process descriptors should in principle capture a detailed representation of physical and chemical environment, reproducibility, and most importantly an interoperable data representation, could evolve from data-based PI.

An interesting aspect of future SDLs is automated research question formulation that is currently underdeveloped. Graph representation of potentially missing data in the virtual world of SDLs could potentially guide the specification of research questions. Although several proof-of-concept strategies in graph-based knowledge modeling have demonstrated automated agents that address inconsistent thermodynamic data and send requests to perform DFT calculations,[80] such graph representations could be generalized to agents that guide the attention of researchers. Another example arises from analyzing relationships among the knowledge graph structures to find potentially new properties or structures of chemicals, therefore, guiding research efforts.[81] For example, through transfer learning, a graph neural network accurately predicted conductivity on more than 10^4 potential candidates of lithium-conducting polymer electrolytes, narrowing the search of high-performing materials for energy storage technologies.[82] Building graph representation of data requires access to large data libraries, which are usually scattered across multiple databases, from patents and research articles[83] to personal lab notebooks.[84] Aggregating such data into useful query-able knowledge graphs could be achieved by leveraging cloud-based computation resources. RoboRXN exemplifies this workflow through the orchestration of different cloud-based technologies that range automated data collection, automated hypothesis making, and autonomous experimentation.[1]

4. AMMD: Redefining the Role of Human in Experimental Sciences

The presented data processes mark a technological roadmap toward an AMMD. While a strong technological infrastructure is imperative for the advancement of research acceleration efforts, its integration into the existing chemistry and material science communities will be a crucial factor for its eventual success. As it is true for any new framework or technology, its broad adoption will depend on various factors, including society awareness, identifying tangible benefits, and having access to a large enough pool of highly trained workforce with multidisciplinary skills. We believe that the broad interest to AMMD continues to grow, as evidenced by various efforts that embrace data-driven materials development in academia and industrial workforce by introducing new educational programs.[85] Furthermore, there is a common perception that AMMD is not bound to replace human researchers, but to redefine their role in the entire discovery cycle, enhancing the discovery capability of researchers by allowing them to focus on creative and new endeavors within the chemical universe.

From an emerging technology adoption perspective, implementation challenges naturally arise for data-intensive research. This implementation challenge could be analogously studied by how knowledge diffuses among the academic and industrial research communities. The leading effort to ensure democratization of access and communication of molecular and material data is through Open Data efforts. These data management infrastructures follow FAIR (Findable, Accessible, Interoperable, Reusable) data principles to facilitate the integration of various AMMD efforts. For a comprehensive review of different implementation challenges such as financing, ownership, and longevity, we refer the reader to an excellent review by Himanen et al.[13] We expect that further development in the technological infrastructure of AMMD will continue to improve the underlying data processes of each discovery step, lowering the entry barrier, and thereby facilitating both adoption and implementation of AMMD strategies. For instance, advances in blockchain technology could impact transparency, accessibility, and reliability in SDLs data generation and retrieval.[86]

It is well-known from other exponential technologies that sustained improvements have physical limits, for example, Moore's Law is expected to no longer follow the doubling trend after 2025, and thus potential strategies to keep the pace of technological advancements are already in motion. In the context of AMMD, we foresee a similar roadblock and draw similarities between the prototypical strategies, such as creating processes application specific, while keeping the data flows intact. Another example is rethinking the entire chemistry workflow. Alternative chemical paradigms such as Circular Chemistry, that go beyond discovering better materials, can further improve the goals of minimizing chemical consumption and create a safer working environment. SDLs can facilitate material flow and life cycle analysis of emerging advanced materials and molecules by providing precise (meta) data throughout the overall process. Furthermore, circularity and sustainability indices, as well as carbon footprint of different materials and molecular discovery campaigns, can be readily integrated into the decision-making process of SDLs to further enhance the overall goals of AMMD strategies.

In summary, SDLs will not reach their full potential for AMMD without open science and collaborations within academia and industry. It is early to decide, for example, what the correct technological strategies are to attain the shortest time-to-solution to several worldwide problems. However, we believe that by providing a data-centric framework for AMMD, the call could reach a larger audience, and interdisciplinary global collaboration could be further facilitated.

5. Conclusion

Motivated by the growing expectation that advanced materials and specialty/fine chemicals will experience an exponential acceleration in their pace of discovery, in this perspective article, we discussed the otherwise ignored role of chemical engineering, in particular, reactor and reaction engineering as a fundamental technological infrastructure for intensifying the research output of SDLs. We identified PI as an enabling factor in elevating the performance of currently existing SDLs. Specifically, we discussed how treating PI as a data-enabler mechanism of the reaction modules of SDLs could result in sustained acceleration of AMMD efforts. We believe the overall technological infrastructures of SDLs should be reevaluated through the lens of PI to unleash their full potential for much-needed acceleration of materials and molecular discovery. 1) We laid a data-centric framework to identify, address, and advance the hardware and software strategies in SDLs from an AMMD perspective to optimally "close the discovery loop" in chemical and materials sciences. In addition, we discussed the importance of digitalization strategies, specifically the representation of intensified reaction modules, in addition to data standardization, data interoperability, and data democratization as requirements for AMMD efforts. 2) We envision an ideal and universal SDL should have access to the prior PI knowledge, such as intensified reactors or processes suited for different chemistries. Such knowledge can then be leveraged in a digital format to identify if using certain PI strategies can be applied anywhere in the SDL workflow to further intensify research productivity and sustainability for AMMD. In addition, continuous intensification of different aspects of SDLs will result in a significant reduction of the environmental impact of AMMD.

Acknowledgements

The financial support from the Dreyfus Program for Machine Learning in the Chemical Sciences and Engineering (Award # ML-21-064) and National Science Foundation (Award # 1940959) is gratefully acknowledged by MA.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

accelerated discovery, autonomous experimentation, digital labs, process intensification, self-driving labs

Received: October 1, 2022
Revised: December 6, 2022
Published online: December 23, 2022

[1] J. Shalf, *Philos. Trans. R. Soc. A* **2020**, *378*, 20190061.
[2] R. Kurzweil, in *Alan Turing: Life and Legacy of a Great Thinker* (Ed: C. Teuscher), Springer, Berlin, Heidelberg **2004**, pp. 381–416.
[3] L. Bernstein, A. Sludds, R. Hamerly, V. Sze, J. Emer, D. Englund, *Sci. Rep.* **2021**, *11*, 3144.
[4] N. Taniguchi, *CIRP Ann.* **1983**, *32*, 573.
[5] Z. Yin, H. Lan, G. Tan, M. Lu, A. V. Vasilakos, W. Liu, *Comput. Struct. Biotechnol. J.* **2017**, *15*, 403.
[6] J. Zhong, J. Riordon, T. C. Wu, H. Edwards, A. R. Wheeler, K. Pardee, A. Aspuru-Guzik, D. Sinton, *Lab Chip* **2020**, *20*, 709.
[7] S. O'Neill, *Engineering* **2021**, *7*, 1351.
[8] A. Bubliauskas, D. J. Blair, H. Powell-Davies, P. J. Kitson, M. D. Burke, L. Cronin, *Angew. Chem., Int. Ed.* **2022**, *61*, e202116108.
[9] W. Hou, A. Bubliauskas, P. J. Kitson, J.-P. Francoia, H. Powell-Davies, J. M. P. Gutierrez, P. Frei, J. S. Manzano, L. Cronin, *ACS Cent. Sci.* **2021**, *7*, 212.
[10] J. M. Parrilla-Gutierrez, S. Tsuda, J. Grizou, J. Taylor, A. Henson, L. Cronin, *Nat. Commun.* **2017**, *8*, 1144.
[11] J. Yano, K. J. Gaffney, J. Gregoire, L. Hung, A. Ourmazd, J. Schrier, J. A. Sethian, F. M. Toma, *Nat. Rev. Chem.* **2022**, *6*, 357.
[12] H. J. Kulik, M. S. Sigman, *Acc. Chem. Res.* **2021**, *54*, 2335.
[13] L. Himanen, A. Geurts, A. S. Foster, P. Rinke, *Adv. Sci.* **2019**, *6*, 1900808.
[14] M. E. Deagen, L. C. Brinson, R. A. Vaia, L. S. Schadler, *MRS Bull.* **2022**, *47*, 379.
[15] N. J. Szymanski, Y. Zeng, H. Huo, C. J. Bartel, H. Kim, G. Ceder, *Mater. Horiz.* **2021**, *8*, 2169.
[16] M. Li, H. Wang, L. Tian, Z. Pang, Q. Yang, T. Huang, J. Fan, L. Song, Y. Tong, H. Fan, *Signal Transduction Targeted Ther.* **2022**, *7*, 1.
[17] P. Ball, *Nature* **2020**, *589*, 16.
[18] M. T. Dvorak, K. C. Armour, D. M. W. Frierson, C. Proistosescu, M. B. Baker, C. J. Smith, *Nat. Clim. Change* **2022**, *12*, 547.
[19] F. Wang, J. D. Harindintwali, Z. Yuan, M. Wang, F. Wang, S. Li, Z. Yin, L. Huang, Y. Fu, L. Li, S. X. Chang, L. Zhang, J. Rinklebe, Z. Yuan, Q. Zhu, L. Xiang, D. C. W. Tsang, L. Xu, X. Jiang, J. Liu, N. Wei, M. Kästner, Y. Zou, Y. S. Ok, J. Shen, D. Peng, W. Zhang, D. Barceló, Y. Zhou, Z. Bai, et al., *Innovation* **2021**, *2*, 100180.
[20] A. A. Kebede, T. Kalogiannis, J. Van Mierlo, M. Berecibar, *Renewable Sustainable Energy Rev.* **2022**, *159*, 112213.
[21] E. O. Pyzer-Knapp, J. W. Pitera, P. W. J. Staar, S. Takeda, T. Laino, D. P. Sanders, J. Sexton, J. R. Smith, A. Curioni, *npj Comput. Mater.* **2022**, *8*, 1.
[22] R. Pollice, G. dos Passos Gomes, M. Aldeghi, R. J. Hickman, M. Krenn, C. Lavigne, M. Lindner-D'Addario, A. Nigam, C. T. Ser, Z. Yao, A. Aspuru-Guzik, *Acc. Chem. Res.* **2021**, *54*, 849.
[23] M. M. Flores-Leonar, L. M. Mejía-Mendoza, A. Aguilar-Granda, B. Sanchez-Lengeling, H. Tribukait, C. Amador-Bedolla, A. Aspuru-Guzik, *Curr. Opin. Green Sustainable Chem.* **2020**, *25*, 100370.
[24] Z. S. Campbell, F. Bateni, A. A. Volk, K. Abdel-Latif, M. Abolhasani, *Part. Part. Syst. Charact.* **2020**, *37*, 2000256.
[25] K. Abdel-Latif, F. Bateni, S. Crouse, M. Abolhasani, *Matter* **2020**, *3*, 1053.
[26] C. W. Coley, N. S. Eyke, K. F. Jensen, *Angew. Chem., Int. Ed.* **2020**, *59*, 22858.
[27] C. W. Coley, N. S. Eyke, K. F. Jensen, *Angew. Chem., Int. Ed.* **2020**, *59*, 23414.
[28] R. W. Epps, M. S. Bowen, A. A. Volk, K. Abdel-Latif, S. Han, K. G. Reyes, A. Amassian, M. Abolhasani, *Adv. Mater.* **2020**, *32*, 2001626.
[29] M. Seifrid, R. Pollice, A. Aguilar-Granda, Z. Morgan Chan, K. Hotta, C. T. Ser, J. Vestfrid, T. C. Wu, A. Aspuru-Guzik, *Acc. Chem. Res.* **2022**, *55*, 2454.
[30] R. W. Epps, A. A. Volk, K. G. Reyes, M. Abolhasani, *Chem. Sci.* **2021**, *12*, 6025.
[31] P. Sagmeister, F. F. Ort, C. E. Jusner, D. Hebrault, T. Tampone, F. G. Buono, J. D. Williams, C. O. Kappe, *Adv. Sci.* **2022**, *9*, 2105547.
[32] 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, A. I. Cooper, *Nature* **2020**, *583*, 237.

[33] J. A. Bennett, M. Abolhasani, *Curr. Opin. Chem. Eng.* **2022**, *36*, 100831.
[34] F. Bateni, R. W. Epps, K. Antami, R. Dargis, J. A. Bennett, K. G. Reyes, M. Abolhasani, *Adv. Intell. Syst.* **2022**, *4*, 2200017.
[35] S. V. Mohan, R. Katakojwala, *Curr. Opin. Green Sustainable Chem.* **2021**, *28*, 100434.
[36] R. W. Epps, M. Abolhasani, *Appl. Phys. Rev.* **2021**, *8*, 041316.
[37] A. A. Volk, M. Abolhasani, *Trends Chem.* **2021**, *3*, 519.
[38] H. Hayashi, H. Katsuyama, H. Takano, Y. Harabuchi, S. Maeda, T. Mita, *Nat. Synth.* **2022**, *1*, 804.
[39] T. Kim, C. Ri, H. Yun, R. An, G. Han, S. Chae, G. Kim, G. Jong, Y. Jon, *Sci. Rep.* **2019**, *9*, 20264.
[40] Z. Qiao, M. Welborn, A. Anandkumar, F. R. Manby, T. F. Miller, III, *Chem. Phys.* **2020**, *153*, 124111.
[41] R. Irwin, S. Dimitriadis, J. He, E. J. Bjerrum, *Mach. Learn.: Sci. Technol.* **2022**, *3*, 015022.
[42] E. Hoogeboom, V. G. Satorras, C. Vignac, M. Welling, in *Proceedings of the 39th International Conference on Machine Learning* (Eds: K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, S. Sabato), Proceedings of Machine Learning Research, Vol. *162*, PMLR 17–23 July, **2022**, pp. 8867–8887..
[43] Q. Vanhaelen, Y. Lin, A. Zhavoro, *ACS Med. Chem. Lett.* **2020**, *11*, 1496.
[44] F. Zangeneh-Nejad, D. L. Sounas, A. Alù, R. Fleury, *Nat. Rev. Mater.* **2021**, *6*, 207.
[45] P. W. K. Jensen, L. B. Kristensen, C. Lavigne, A. Aspuru-Guzik, *J. Chem. Theory Comput.* **2022**, *18*, 3318.
[46] V. Stanev, K. Choudhary, A. G. Kusne, J. Paglione, I. Takeuchi, *Commun. Mater.* **2021**, *2*, 1.
[47] A. Ajagekar, F. You, *Renewable Sustainable Energy Rev.* **2022**, *165*, 112493.
[48] S. Amselem, *Pharm. Res.* **2019**, *36*, 183.
[49] T. Van Gerven, A. Stankiewicz, *Ind. Eng. Chem. Res.* **2009**, *48*, 2465.
[50] R. L. Hartman, *Org. Process Res. Dev.* **2012**, *16*, 870.
[51] A. Günther, K. F. Jensen, *Lab Chip* **2006**, *6*, 1487.
[52] K.-J. Wu, L. Torrente-Murciano, *React. Chem. Eng.* **2018**, *3*, 267.
[53] J.-Y. Jung, J.-G. Lee, Y.-K. Baek, Y.-D. Kim, J.-P. Hong, Y.-K. Kim, *J. Alloys Compd.* **2019**, *784*, 816.
[54] G. Vilardi, M. Stoller, L. Di Palma, K. Boodhoo, N. Verdone, *Chem. Eng. Process.* **2019**, *146*, 107683.
[55] I. Lignos, Y. Mo, L. Carayannopoulos, M. Ginterseder, M. G. Bawendi, K. F. Jensen, *React. Chem. Eng.* **2021**, *6*, 459.
[56] A. I. Stankiewicz, P. Yan, *Ind. Eng. Chem. Res.* **2019**, *58*, 9212.
[57] V. Hessel, M. M. Sarafraz, N. N. Tran, *Chem. Eng. Sci.* **2020**, *225*, 115774.
[58] J. P. McMullen, K. F. Jensen, *Annu. Rev. Anal. Chem.* **2010**, *3*, 19.
[59] I. Takeuchi, J. Lauterbach, M. J. Fasolka, *Mater. Today* **2005**, *8*, 18.
[60] A. Ludwig, *npj Comput. Mater.* **2019**, *5*, 1.
[61] D. Heggo, S. Ookawara, *Chem. Eng. Sci.* **2017**, *169*, 67.
[62] D. Cambié, C. Bottecchia, N. J. W. Straathof, V. Hessel, T. Noël, *Chem. Rev.* **2016**, *116*, 10276.
[63] F. J. Keil, *Rev. Chem. Eng.* **2018**, *34*, 135.
[64] M. Kneiß, P. Storm, G. Benndorf, M. Grundmann, H. von Wenckstern, *ACS Comb. Sci.* **2018**, *20*, 643.
[65] J. Baek, Y. Shen, I. Lignos, M. G. Bawendi, K. F. Jensen, *Angew. Chem., Int. Ed.* **2018**, *57*, 10915.
[66] E. A. López-Guajardo, F. Delgado-Licona, A. J. Álvarez, K. D. P. Nigam, A. Montesinos-Castellanos, R. Morales-Menendez, *Chem. Eng. Process.* **2021**, *180*, 108671.
[67] M. Ghobakhloo, *J. Clean. Prod.* **2020**, *252*, 119869.
[68] A. A. Khan, A. A. Lapkin, *Chem. Eng. Process.* **2022**, *180*, 108885.
[69] K. Gilmore, D. Kopetzki, J. Weon Lee, Z. Horváth, D. Tyler McQuade, A. Seidel-Morgenstern, P. H. Seeberger, *Chem. Commun.* **2014**, *50*, 12652.
[70] R. L. Greenaway, V. Santolini, M. J. Bennison, B. M. Alston, C. J. Pugh, M. A. Little, M. Miklitz, E. G. B. Eden-Rump, R. Clowes, A. Shakil, H. J. Cuthbertson, H. Armstrong, M. E. Briggs, K. E. Jelfs, A. I. Cooper, *Nat. Commun.* **2018**, *9*, 2849.
[71] J. Bai, L. Cao, S. Mosbach, J. Akroyd, A. A. Lapkin, M. Kraft, *JACS Au* **2022**, *2*, 292.
[72] C. A. Hone, C. O. Kappe, *Chem.–Methods* **2021**, *1*, 454.
[73] S. S. Zalesskiy, P. J. Kitson, P. Frei, A. Bubliauskas, L. Cronin, *Nat. Commun.* **2019**, *10*, 5496.
[74] L. M. Roch, F. Häse, C. Kreisbeck, T. Tamayo-Mendoza, L. P. E. Yunker, J. E. Hein, A. Aspuru-Guzik, *Plos One* **2020**, *15*, 0229862.
[75] D. Iglesias, D. Haddad, V. Sans, *Chem.–Methods* **2022**, *2*, e202200031.
[76] K. Abdel-Latif, R. W. Epps, F. Bateni, S. Han, K. G. Reyes, M. Abolhasani, *Adv. Intell. Syst.* **2021**, *3*, 2000245.
[77] J. Akroyd, S. Mosbach, A. Bhave, M. Kraft, *Data-Centric Eng.* **2021**, *2*, e14.
[78] G. Wang, A. Badal, X. Jia, J. S. Maltz, K. Meuller, K. J. Myers, C. Niu, M. Vannier, P. Yan, Z. Yu, R. Zeng, *Nat. Mach. Intell.* **2022**, *4*, 922.
[79] J. Li, Y. Tu, R. Liu, Y. Lu, X. Zhu, *Adv. Sci.* **2020**, *7*, 1901957.
[80] S. Mosbach, A. Menon, F. Farazi, N. Krdzavac, X. Zhou, J. Akroyd, M. Kraft, *J. Chem. Inf. Model.* **2020**, *60*, 6155.
[81] S. Spangler, A. D. Wilkins, B. J. Bachman, M. Nagarajan, T. Dayaram, P. Haas, S. Regenbogen, C. R. Pickering, A. Comer, J. N. Myers, I. Stanoi, L. Kato, A. Lelescu, J. J. Labrie, N. Parikh, A. M. Lisewski, L. Donehower, Y. Chen, O. Lichtarge, in *Proc. 20th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, ACM, New York, NY, USA **2014**, pp. 1877–1886.
[82] K. Hatakeyama-Sato, T. Tezuka, M. Umeki, K. Oyaizu, *J. Am. Chem. Soc.* **2020**, *142*, 3301.
[83] J. Guo, A. S. Ibanez-Lopez, H. Gao, V. Quach, C. W. Coley, K. F. Jensen, R. Barzilay, *J. Chem. Inf. Model.* **2022**, *62*, 2035.
[84] P. Tremouilhac, C.-L. Lin, P.-C. Huang, Y.-C. Huang, A. Nguyen, N. Jung, F. Bach, R. Ulrich, B. Neumair, A. Streit, S. Bräse, *Angew. Chem., Int. Ed.* **2020**, *59*, 22771.
[85] E. Chen, O. Igbineweka, L. Kubie, J. S. Peerless, *MRS Bull.* **2022**, https://doi.org/10.1557/s43577-022-00410-6.
[86] M. W. D. Hanson-Heine, A. P. Ashmore, *Chem. Sci.* **2020**, *11*, 4644.

Fernando Delgado-Licona received a B.Sc. degree in chemical engineering with a minor in innovation and development at Tecnologico de Monterrey in 2020. During his undergraduate studies, Fernando developed an interest in process intensification and its role to attain sustainable processes. He then joined the Energy and Climate Change group at Tecnologico de Monterrey for his M.Sc. studies and worked on precision synthesis of metal oxides in flow reactors. Fernando is currently a Ph.D. student in the Department of Chemical & Biomolecular Engineering at North Carolina State University working at the intersection of process intensification and machine learning in the Abolhasani Lab.

Milad Abolhasani is an associate professor and a University Faculty Scholar in the Department of Chemical and Biomolecular Engineering at North Carolina State University. Abolhasani leads a diverse research group that studies self-driving labs tailored toward accelerated development and manufacturing of advanced functional materials and molecules using fluidic micro-processors. Abolhasani has received numerous awards and fellowships, including NSF CAREER Award, AIChE 35 Under 35, Dreyfus Award for Machine Learning in the Chemical Sciences & Engineering, I &EC Research 2021 Class of Influential Researchers, AIChE NSEF Young Investigator Award, AIChE Futures Scholar, and ACS-PRF Doctoral New Investigator Award.