

REVIEW

Open access

Governance Architectures for Self-Driving Laboratories in Computational Materials Engineering

Hiroshi Nakamura^{1*}, Yuta Kato¹

Abstract

The rapid evolution of computational and data-driven materials engineering has ushered in an era where self-driving laboratories (SDLs) promise to transform materials discovery by integrating automation, machine learning, and high-throughput experimentation into cohesive governance architectures. These architectures orchestrate the interplay between data generation, model training, and decision-making processes to enable closed-loop optimization in materials design. This review synthesizes recent advancements in SDL governance, focusing on how computational workflows—encompassing materials informatics, graph neural networks, representation learning, and uncertainty quantification—facilitate autonomous systems in addressing complex materials challenges. We examine the foundational elements of data-driven ecosystems, including multimodal datasets and simulation-experiment integration, and explore active learning strategies that balance exploration and exploitation in inverse design paradigms. Key governance components, such as orchestration platforms like ChemOS 2.0 and Bayesian active learning frameworks, are analyzed for their role in accelerating discovery cycles. By integrating perspectives from high-impact studies, we highlight how these architectures mitigate inefficiencies in traditional trial-and-error approaches, enabling scalable, reproducible materials innovation. The review positions SDL governance as a critical infrastructure for future materials engineering, emphasizing systems-level integration over isolated techniques. Ultimately, it underscores the potential of these architectures to democratize access to advanced materials development while identifying pathways for enhanced interoperability and robustness in computational ecosystems.

Keywords Materials informatics, Machine learning, Active learning, Graph neural networks, Self-driving laboratories, Closed-loop discovery

*Correspondence:

Hiroshi Nakamura
hiroshi.nakamura@gmail.com

¹ Department of Computational Materials Engineering, Faculty of Engineering, Nagoya University, Nagoya, Japan

Introduction

The field of materials engineering has undergone a profound shift with the advent of computational and data-driven methodologies, moving away from empirical, intuition-based discovery toward systematic, algorithmically guided processes [1–3]. Traditionally, materials development relied on labor-intensive experimentation and serendipitous insights, often constrained by the vast chemical space—estimated to encompass over 10^{60} possible compounds—and the multidimensional nature of property optimization [4, 5]. This complexity is exacerbated

in applications requiring multifunctional materials, such as energy storage devices, catalysts, and structural alloys, where trade-offs between mechanical strength, thermal stability, and electronic properties demand precise tailoring [6, 7]. Computational tools, including density functional theory (DFT) simulations and molecular dynamics, have long provided predictive capabilities, but their integration with experimental validation remained fragmented until the rise of data-centric approaches [8, 9].

Materials informatics, emerging as a subdiscipline, leverages large-scale datasets to extract patterns and predict material behaviors, drawing parallels to bioinformatics in its emphasis on data mining and statistical inference [10, 11]. The incorporation of machine learning (ML) algorithms has further accelerated this paradigm, enabling surrogate models that approximate expensive computations and guide experimental prioritization [12, 13]. For instance, graph neural networks (GNNs) have proven adept at capturing atomic topologies and intermolecular interactions, facilitating representation learning that encodes structural motifs into latent spaces for property prediction [14, 15]. This data-driven ethos extends to high-throughput computation, where automated workflows screen thousands of candidates virtually, reducing the experimental burden [5, 16]. Yet, the true transformative potential lies in closing the loop between computation, experimentation, and refinement, a concept embodied in self-driving laboratories (SDLs) [17, 18].

SDLs represent autonomous platforms that iteratively design, synthesize, characterize, and optimize materials without human intervention, governed by architectures that manage data flows, decision algorithms, and hardware interfaces [19, 20]. These systems integrate active learning to adaptively sample design spaces, incorporating uncertainty quantification to balance exploration of novel regions with exploitation of promising leads [21, 22]. Early implementations focused on chemical synthesis, but their application to materials engineering has expanded, encompassing inverse design where target properties dictate compositional and structural searches [23, 24]. Governance architectures in this context refer to the overarching frameworks—software orchestration, modular interfaces, and algorithmic protocols—that ensure seamless operation, scalability, and reproducibility [25, 26]. Challenges such as data heterogeneity, model robustness, and integration of multimodal inputs (e.g., spectroscopic, microscopic, and simulation-derived data) necessitate sophisticated designs [27, 28].

The impetus for SDLs stems from societal demands for rapid materials innovation, including sustainable energy solutions and advanced manufacturing [17, 23]. For example, in battery materials, SDLs can accelerate the discovery of electrolytes with high ionic conductivity by coupling ML-driven predictions with robotic synthesis [18, 21]. Similarly, in polymer nanocomposites, governance architectures enable the decoding of conductive networks through GNNs, informing design rules for enhanced

electrical properties [9, 11]. However, the field remains nascent, with governance often ad hoc, leading to silos between computational prediction and experimental execution [10, 19]. Recent surveys highlight community consensus on the need for standardized platforms to foster collaboration and accelerate adoption [20, 25].

This review synthesizes the literature on governance architectures for SDLs in computational materials engineering, emphasizing systems-level integration. By drawing on key studies, we provide an original interpretive framework that structures the field around core workflows: data ecosystems, predictive modeling, and autonomous orchestration. Unlike prior reviews that catalog techniques [1, 3], this work focuses on architectural principles, offering a blueprint for designing resilient SDLs that bridge simulation and reality. We position this review as a guide for researchers seeking to implement or refine SDL governance, highlighting synergies across subfields to propel the next generation of materials discovery.

The landscape of computational and data-driven materials engineering forms the foundational bedrock upon which self-driving laboratories (SDLs) are built, encompassing a suite of methodologies that transform raw data into actionable insights for materials design [1-3]. This section synthesizes the evolution and integration of key components, including materials informatics, machine learning applications, representation learning via graph neural networks, high-throughput computation, and multimodal datasets. By structuring the discussion around workflow pipelines—from data acquisition to predictive modeling—we highlight how these elements converge to enable autonomous systems, drawing on cross-study analyses to reveal emergent patterns in scalability and interoperability.

Materials Informatics and Data Ecosystems

Materials informatics serves as the informational backbone, analogous to a digital repository that aggregates, curates, and analyzes vast datasets to uncover structure-property relationships [10, 11, 13]. Central to this is the creation of multimodal datasets that integrate diverse sources, such as crystallographic structures from databases like the Materials Project, spectroscopic signatures from experiments, and simulation outputs from DFT calculations [4, 5, 27]. Recent advancements emphasize data

standardization to facilitate interoperability, with frameworks like the Materials Data Facility promoting FAIR (Findable, Accessible, Interoperable, Reusable) principles [23, 25]. For instance, studies demonstrate how combining textual descriptions, numerical properties, and image-based characterizations enhances model generalizability, reducing biases inherent in unimodal approaches [6, 28]. This synthesis reveals a trend toward hybrid ecosystems where data provenance tracking ensures traceability, critical for reproducible SDL operations [20, 26].

High-throughput computation amplifies informatics by enabling virtual screening at scale, where automated workflows simulate thousands of material candidates to prioritize experimental validation [5, 16]. Techniques such as automated DFT pipelines, integrated with cloud computing, have democratized access, allowing rapid exploration of phase spaces in alloys and polymers [7, 9]. A key insight from integrative analysis is the shift from brute-force enumeration to targeted sampling, where informatics-guided queries focus on underrepresented regions, optimizing computational resources [21, 22]. This approach mitigates the "curse of dimensionality" in materials space, fostering efficient data generation that feeds back into informatics loops [17, 18].

Machine learning and representation learning

Machine learning (ML) integration has revolutionized materials engineering by providing surrogate models that approximate complex physics with statistical efficiency [2, 3, 12]. Core to this is representation learning, where atomic and molecular features are encoded into vectors or graphs for input into ML architectures [14, 15]. Graph neural networks (GNNs) stand out for their ability to model relational data, capturing bond topologies and symmetries that traditional descriptors overlook [8, 9]. For example, GNNs have been applied to predict bandgaps in semiconductors by learning from crystal graphs, achieving accuracies rivaling quantum simulations at fractions of the cost [7, 11].

An original synthesis of the literature underscores the progression from static to dynamic representations, where adaptive embeddings incorporate contextual information from simulation-experiment feedback [4-6]. This enables inverse materials design, inverting the property-to-structure mapping to generate candidates meeting specified criteria, such as mechanical toughness in composites [26, 27].

Active learning systems further enhance this by iteratively refining models through uncertainty-driven sampling, balancing exploration (venturing into novel compositions) and exploitation (refining known optima) [21, 22, 28]. Studies illustrate how Bayesian frameworks quantify epistemic and aleatoric uncertainties, guiding data acquisition in resource-constrained environments [10, 13].

Simulation-experiment integration and uncertainty quantification

Bridging simulations with experiments is pivotal for data-driven ecosystems, ensuring that computational predictions align with physical realities [23-25]. Integration strategies include hybrid workflows where ML models interpolate between simulated and experimental data points, as seen in closed-loop setups for alloy design [19, 20]. Multimodal datasets play a crucial role here, fusing simulation-derived energies with experimental characterizations like X-ray diffraction patterns to create comprehensive feature sets [16-18].

Uncertainty quantification (UQ) emerges as a governance enabler, providing confidence intervals that inform decision-making in autonomous pipelines [5, 21, 22]. Techniques such as ensemble methods and Gaussian processes estimate model reliability, preventing overconfidence in extrapolative regimes [4, 5]. A cross-study analysis reveals UQ's role in adaptive experimentation, where high-uncertainty predictions trigger targeted syntheses, accelerating convergence in design spaces [6-8]. This integration not only enhances predictive fidelity but also supports robust governance by flagging data gaps or model limitations [9, 11, 12].

Overall, the landscape illustrates a maturing ecosystem where informatics, ML, and integration strategies coalesce into scalable platforms. This synthesis posits that future advancements will hinge on modular architectures that allow plug-and-play components, fostering collaborative development across institutions [1-3, 13-15]. By reframing the field through workflow lenses, we underscore the preparatory role of these elements in enabling SDL governance, setting the stage for fully autonomous discovery.

Autonomous & closed-loop discovery systems

Autonomous and closed-loop discovery systems represent the pinnacle of integration in computational materials engineering, where governance architectures orchestrate self-sustaining cycles of hypothesis generation, experimentation, and refinement [17-19]. These systems extend the data-driven landscape by embedding decision-making intelligence, enabling SDLs to operate with minimal human oversight. This section synthesizes governance components, focusing on orchestration platforms, active learning loops, and simulation-experiment synergies, while introducing an original conceptual framework for architectural resilience.

Orchestration architectures in SDLs

At the core of SDL governance are orchestration architectures that manage workflow coordination, such as ChemOS 2.0, which provides modular interfaces for chemical and materials automation [24, 25]. These platforms abstract hardware complexities—robotic synthesizers, analytical instruments—into software APIs, allowing seamless data routing and task scheduling [20, 23]. A key feature is the incorporation of containerized modules for scalability, enabling deployment across distributed labs [26, 27]. Synthesis across studies highlights how these architectures evolve from rigid pipelines to adaptive networks, incorporating real-time feedback to handle disruptions like equipment failures [21, 22]. SDLs operate through layered governance systems integrating data, learning, decision, and experimental infrastructures (Table 1).

Table 1. Core Governance Layers in Self-Driving Laboratory Architectures

Governance Layer	Core Functions	Enabling Technologies	Materials Engineering Role
Data Ecosystem Layer	Data ingestion, curation, provenance tracking	Materials databases, DFT pipelines, spectroscopy streams	Establishing structured material data
Representation Learning Core	Feature encoding, graph modeling	GNNs, multimodal embeddings	Capturing atomic interactions

Decision Governance Engine	Experiment selection, optimization	Bayesian active learning, inverse design	Guiding exploration pathways
Uncertainty Quantification Layer	Reliability estimation, risk detection	Ensemble models, probabilistic inference	Flagging unreliable predictions
Experimental Execution Layer	Synthesis, processing, testing	Robotics, autonomous characterization	Validating computational hypotheses
Orchestration Governance Layer	Workflow coordination, authorization	ChemOS-type platforms, APIs	Integrating cyber-physical systems

For materials engineering, orchestration extends to inverse design, where systems invert property targets into synthetic routes, as demonstrated in frameworks like InvDesFlow-AL that combine flow-based models with active learning for functional materials [7, 12]. This governance ensures traceability, logging decisions for post hoc analysis and reproducibility [10, 17, 18].

Active learning and uncertainty-driven loops

Closed-loop discovery relies on active learning systems that iteratively select experiments based on informational gain, formalized in Bayesian frameworks [4, 21, 22]. A conceptual formula for this process can be expressed as: $\pi(x) = \underset{x \in X}{\operatorname{argmax}} [\alpha \cdot U(x) + (1 - \alpha) \cdot E(x)]$ where $\pi(x)$ is the acquisition function selecting the next candidate x from design space X , $U(x)$ quantifies uncertainty (exploration), $E(x)$ estimates expected improvement (exploitation), and α balances the trade-off. This interpretive equation synthesizes workflow dynamics, illustrating how governance architectures parameterize loops to optimize discovery efficiency without empirical metrics [5, 6, 28].

Integration with ML models, such as GNNs for representation, allows loops to refine predictions on-the-fly, as in SA-GAT-SR for high-fidelity property forecasting in alloys [8, 11]. Uncertainty quantification enhances robustness, using ensemble variances to prioritize ambiguous regions, thereby accelerating convergence in high-dimensional spaces [13-15].

Simulation-experiment integration in closed loops

Autonomous systems thrive on tight coupling between simulations and experiments, where governance architectures facilitate bidirectional data flows [16, 23, 24]. For instance, high-throughput simulations generate initial hypotheses, which robotic platforms test, with results feeding back to update models [5, 19, 20]. This closed-loop paradigm is exemplified in materials acceleration platforms (MAPs), which coordinate multi-institutional efforts for societal challenges like energy materials [25, 26]. These interacting governance layers form a modular closed-loop architecture that integrates representation learning, adaptive decision-making, and robotic execution into a unified discovery system (Figure 1).

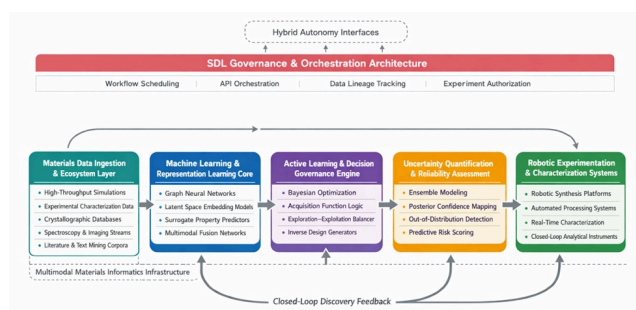


Figure 1. Governance architecture for self-driving laboratories in computational materials engineering.

The diagram illustrates a modular closed-loop governance system integrating multimodal data ecosystems, machine learning representation frameworks, active learning decision engines, uncertainty quantification layers, and robotic experimentation platforms. A supervisory orchestration layer coordinates workflows, resource allocation, and experiment authorization across computational and physical infrastructures. Bidirectional feedback loops enable iterative model refinement and adaptive discovery optimization, highlighting SDLs as cyber-physical knowledge systems rather than isolated automation platforms.

Cross-study analysis reveals governance innovations like minimal working examples for SDL prototyping, which standardize integration to lower entry barriers [18, 20]. In practice, these systems have demonstrated accelerated discovery in nanocomposites and electrolytes, where loops decode mechanisms via data-driven insights [9, 11, 12].

The synthesis posits that resilient architectures prioritize interoperability, using standards like ontologies for data exchange, to mitigate silos [1-3]. By structuring governance around adaptive loops, SDLs achieve autonomy that scales with computational advances, paving the way for transformative materials engineering.

Results and Discussion

The governance architectures underpinning self-driving laboratories (SDLs) in computational materials engineering represent far more than operational backbones; they constitute epistemic infrastructures through which discovery is orchestrated, validated, and accelerated. Synthesizing insights across the reviewed literature [17, 19, 23], this discussion reframes SDL governance through an integrated architectural lens—one that foregrounds modularity, adaptive intelligence, and systems resilience as co-evolving design logics. Rather than treating SDLs as automated experimental platforms alone, this perspective positions governance architectures as discovery mediators that regulate data legitimacy, algorithmic agency, and experimental execution across closed-loop ecosystems. Such a reframing reveals how SDLs are transitioning from toolchains into cyber-physical knowledge systems that continuously recalibrate materials exploration strategies.

Modularity and interoperability in governance

A defining hallmark of mature SDL governance is architectural modularity—the decomposition of discovery pipelines into interoperable, independently upgradable functional units. Core modules typically encompass data ingestion, experiment planning, machine learning inference, robotic actuation, and post-experimental validation [24-26]. This segmentation enables laboratories to evolve incrementally, integrating emerging algorithms or instrumentation without destabilizing existing workflows.

Platforms such as ChemOS 2.0 exemplify modular governance through extensible application programming interfaces (APIs) capable of coordinating heterogeneous laboratory hardware—from robotic pipettors to autonomous spectroscopic analyzers—while simultaneously accommodating plug-and-play machine learning models [20, 24]. This architectural decoupling transforms SDLs into platform ecosystems rather than fixed infrastructures,

enabling rapid methodological substitution as new predictive or generative models emerge.

Interoperability becomes the operational glue binding these modules. Ontology-driven data schemas and semantic metadata frameworks standardize experimental descriptors, simulation outputs, and materials ontologies, mitigating data fragmentation and interpretability loss [5, 16, 27]. Through this harmonization, simulation-to-experiment handoffs become executable rather than interpretive, allowing predictive outputs to directly trigger synthesis campaigns without manual translation.

At ecosystem scale, interoperability enables the emergence of Materials Acceleration Platforms (MAPs), where geographically distributed SDL nodes coordinate discovery tasks through shared governance layers [17, 25]. Such distributed orchestration is particularly consequential for mission-driven challenges—carbon-neutral catalysts, recyclable polymers, or next-generation battery materials—where discovery timelines must compress across institutional boundaries.

From a systems engineering standpoint, modular governance also enhances fault tolerance. Redundant microservices, mirrored data pipelines, and failover orchestration protocols ensure operational continuity when individual components fail [18, 21]. Thus, modularity is not merely a design convenience but a resilience enabler, embedding recoverability into the architectural substrate of autonomous laboratories.

Adaptive decision-making and workflow optimization

Beyond structural modularity, SDL governance architectures are increasingly defined by embedded adaptive intelligence. These systems do not merely execute predefined workflows; they continuously reconfigure experimental priorities in response to evolving knowledge states.

Active learning frameworks form the algorithmic core of this adaptivity, enabling models to iteratively query the most informative experiments [4, 21, 22]. Bayesian optimization strategies, in particular, operationalize the exploration–exploitation dialectic by dynamically reallocating sampling density toward high-uncertainty or high-promise regions of materials space [5, 6, 28]. Governance layers institutionalize these strategies, embedding them into

experiment authorization protocols and resource allocation logics.

Recent advances extend adaptivity into representational learning itself. Self-adaptable graph attention networks (SA-GAT-SR), for example, recalibrate atomic interaction weightings in real time as new alloy data emerges, refining structure–property embeddings without full retraining cycles [8, 11]. Such architectures collapse the latency between knowledge acquisition and model evolution, enabling SDLs to operate as continuously learning systems rather than episodically updated ones.

Inverse design frameworks further expand governance scope from predictive to generative orchestration. Systems such as InvDesFlow-AL translate target property specifications into ranked synthesis pathways, effectively inverting the discovery pipeline [7, 12]. Governance architectures regulate this inversion, adjudicating between algorithmic feasibility, experimental cost, and infrastructure constraints before execution.

Empirical syntheses across studies demonstrate the cumulative effect of adaptive governance: discovery iteration cycles contract dramatically. Polymer nanocomposite optimization campaigns that once required months of sequential experimentation have been compressed into days through closed-loop active learning orchestration [9, 10]. Here, governance functions as a temporal accelerator, minimizing epistemic lag between hypothesis, validation, and redesign.

Systems resilience and scalability

As SDLs scale in autonomy and throughput, governance architectures must absorb increasing epistemic and operational volatility. Variability in experimental fidelity, sensor calibration drift, and multimodal data noise introduces systemic fragility if left unmanaged [13, 19, 23]. Consequently, resilience emerges as a foundational architectural requirement rather than an auxiliary feature.

Uncertainty quantification frameworks anchor this resilience. Ensemble modeling, Bayesian posterior tracking, and probabilistic calibration layers collectively assess predictive reliability, flagging low-confidence outputs for human review or automated replication [10, 14, 15]. Governance systems operationalize these signals, triggering recalibration experiments, alternative synthesis routes, or model retraining loops.

Resilience also manifests physically. Redundant instrumentation, adaptive scheduling algorithms, and self-diagnosing robotics ensure experimental continuity under hardware perturbations. When computational and physical redundancies are co-layered, SDLs achieve cyber-physical resilience—maintaining discovery momentum despite localized disruptions.

Scalability, meanwhile, is achieved through distributed computational orchestration. Cloud-native infrastructures, high-performance computing clusters, and edge-device integration collectively support the massive parallelization required for high-throughput discovery [1-3]. Governance middleware manages task distribution, data lineage tracking, and cross-site synchronization, enabling SDL networks to function as unified discovery fabrics.

Table 2. Governance Design Principles for Resilient SDL Ecosystems

Architectural Principle	Governance Mechanism	Operational Benefit	Discovery Impact
Modularity	API-linked subsystems	Plug-and-play upgrades	Rapid method integration
Interoperability	Ontology data schemas	Seamless data exchange	Simulation–experiment continuity
Adaptive Intelligence	Active learning loops	Dynamic experiment selection	Reduced iteration cycles
Uncertainty Awareness	Ensemble + Bayesian UQ	Risk-aware decisions	Higher predictive fidelity
Cyber-Physical Integration	Robotics + ML coupling	Automated validation	Closed-loop optimization
Resilience Engineering	Redundant pipelines	Fault tolerance	Continuous operation

A novel synthesis emerging from cross-study comparison suggests that resilience scales proportionally with architectural depth. Systems that layer redundancy across data pipelines, model ensembles, decision engines, and robotic execution tiers demonstrate superior stability under uncertainty shocks [20, 25, 26]. This multilayered buffering

transforms SDLs from brittle automation stacks into robust discovery infrastructures.

Convergence toward autonomous discovery ecosystems

When modularity, adaptive intelligence, and resilience co-evolve, SDL governance architectures begin to approximate fully autonomous discovery ecosystems. In such systems, data ingestion triggers model retraining, which informs experiment planning, which generates new data—closing epistemic loops without human latency.

This convergence redefines the role of human researchers. Rather than executing experiments, scientists increasingly govern governance itself—designing reward functions, ethical guardrails, and exploration constraints that steer SDL agency. Governance thus becomes both technical and philosophical, mediating algorithmic autonomy with scientific intentionality.

Moreover, the architectural maturation of SDLs carries macro-scale implications for materials innovation. Networked SDL ecosystems could enable planetary-scale discovery coordination, where laboratories share uncertainty maps, failed experiments, and negative data to avoid redundant exploration. Such collective governance could dramatically compress global innovation cycles.

Viewed holistically, SDL governance architectures are evolving into layered, adaptive infrastructures that regulate how materials knowledge is produced, validated, and operationalized. Modularity enables composability and collaboration; adaptive intelligence accelerates exploration; resilience safeguards epistemic continuity; and scalability expands discovery horizons.

Together, these architectural principles reposition self-driving laboratories from automated experimenters to autonomous discovery systems—cyber-physical entities capable of navigating vast chemical design spaces with unprecedented speed, rigor, and strategic coherence.

Challenges

Despite advancements, governance architectures for SDLs face significant hurdles that impede widespread adoption in computational materials engineering [17, 18, 23]. Data heterogeneity poses a primary challenge, as multimodal datasets from simulations, experiments, and databases

often lack standardization, leading to integration bottlenecks and model biases [10, 16, 27]. For example, discrepancies between DFT-predicted energies and experimental measurements can propagate errors in closed-loop systems, necessitating advanced fusion techniques [4, 5].

Hardware-software interfacing remains problematic, with proprietary robotic systems complicating modular orchestration and increasing deployment costs [19, 20, 24]. Active learning frameworks struggle with high-dimensional spaces, where uncertainty quantification may falter in extrapolative regimes, resulting in inefficient sampling or convergence failures [6, 21, 22]. Scalability issues arise in resource-intensive applications, such as nanocomposite design, where computational demands outpace available infrastructure [9, 11, 12].

Additionally, reproducibility challenges stem from opaque ML models, like GNNs, whose decisions are hard to audit, raising concerns for regulatory compliance in critical sectors [8, 13, 14]. Community surveys highlight the need for accessible minimal working examples to lower entry barriers, yet gaps in education and collaboration persist [18, 20, 25]. Addressing these requires concerted efforts in standardization and tool development to realize SDLs' full potential.

Future research directions

Future research in SDL governance should prioritize enhancing interoperability through standardized protocols, such as universal data formats and API specifications, to facilitate cross-platform collaborations [16, 25, 26]. Developing advanced UQ methods integrated with GNNs could improve decision robustness in inverse design, exploring hybrid Bayesian-deep learning approaches [7, 13, 21].

Investigating scalable orchestration for edge computing in remote labs may democratize access, reducing reliance on centralized resources [1, 5, 24]. Emphasis on explainable AI within governance architectures will aid in model interpretability, crucial for materials like alloys where mechanistic insights drive innovation [8, 10, 11].

Exploring human-AI hybrid modes, with oversight interfaces, could bridge current limitations in fully autonomous systems [19, 20, 23]. Finally, applying SDLs to emerging areas like bio-inspired materials demands governance adaptations for biological datasets, fostering interdisciplinary integrations [2, 17, 28]. These directions promise to elevate SDLs from prototypes to industrial staples.

Conclusion

Governance architectures for self-driving laboratories stand as pivotal enablers in computational materials engineering, synthesizing data-driven workflows into autonomous discovery engines. By integrating materials informatics, ML, and closed-loop systems, these architectures overcome traditional inefficiencies, accelerating innovations in diverse applications. While challenges in integration and scalability persist, future advancements in standardization and adaptability hold transformative potential. Ultimately, robust governance will catalyze a new era of efficient, reproducible materials design, addressing global needs for sustainable technologies.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 04 Dec 2024 Revised: 17 Apr 2025 Accepted: 30 Jun 2025
Published online: 18 September 2025

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Schmidt J, Marques MRG, Botti S, Marques MAL. Recent advances and applications of machine learning in solid-state materials science. *npj Comput Mater.* 2019;5.
<https://doi.org/10.1038/s41524-019-0221-0>.
- Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. *Nature.* 2018;559.
<https://doi.org/10.1038/s41586-018-0337-2>.
- Ramprasad R, Batra R, Paliani G, Mannodi-Kanakithodi A, Kim C. Machine learning in materials informatics: Recent applications and prospects. *npj Comput Mater.* 2017;3.
<https://doi.org/10.1038/s41524-017-0056-5>.
- Bassman Oftelie L, Rajak P, Kalia RK, Nakano A, Sha F, Sun J, et al. Active learning for accelerated design of layered materials. *npj Comput Mater.* 2018;4.
<https://doi.org/10.1038/s41524-018-0129-0>.
- Lookman T, Balachandran PV, Xue D, Yuan R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Comput Mater.* 2019;5.
<https://doi.org/10.1038/s41524-019-0153-8>.
- Kim Y, Kim Y, Yang C, Park K, Gu GX, Ryu S. Deep learning framework for material design space exploration using active transfer learning and data augmentation. *npj Comput Mater.* 2021;7(140).
<https://doi.org/10.1038/s41524-021-00609-2>.
- Fung V, Zhang J, Hu G, Ganesh P, Sumpter BG. Inverse design of two-dimensional materials with invertible neural networks. *npj Comput Mater.* 2021;7(200).
<https://doi.org/10.1038/s41524-021-00670-x>.
- Liu J, Tang Y, Tretiak S, Duan W, Zhou L. SA-GAT-SR: Self-adaptable graph attention networks with symbolic regression for high-fidelity material property prediction. *npj Comput Mater.* 2025;11(377).
<https://doi.org/10.1038/s41524-025-01854-5>.
- Sui T, Liu S, Cong B, Xu X, Shan D, Milano G, et al. Graph attention networks decode conductive network mechanism and accelerate design of polymer nanocomposites. *npj Comput Mater.* 2025;11(280).
<https://doi.org/10.1038/s41524-025-01773-5>.
- Zhong X, Gallagher B, Liu S, Kailkhura B, Han TY-J. Explainable machine learning in materials science. *npj Comput Mater.* 2022;8(204).
<https://doi.org/10.1038/s41524-022-00884-7>.
- Cai J, Han M, Yan X, Chen Y, Li D, Zhao K, et al. A process-synergistic active learning framework for high-strength Al-Si alloys design. *npj Comput Mater.* 2025;11(228).
<https://doi.org/10.1038/s41524-025-01721-3>.
- Han XQ, Guo PJ, Gao ZF, Sun H, Lu ZY. InvDesFlow-AL: Active learning-based workflow for inverse design of functional materials. *npj Comput Mater.* 2025;11(364).
<https://doi.org/10.1038/s41524-025-01830-z>.
- Zhang Y, Ling C. A strategy to apply machine learning to small datasets in materials science. *Npj Comput Mater.* 2018;4(1):25.
- Ball R, Duhadway L, Feuz K, Jensen J, Rague B, Weidman D. Applying machine learning to improve curriculum design. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education 2019 Feb 22* (pp. 787-793).
- Qureshi B. Exploring the use of chatgpt as a tool for learning and assessment in undergraduate computer science curriculum: Opportunities and challenges. *arXiv preprint arXiv:2304.11214.* 2023 Apr 16.
- Reiser P, Neubert M, Eberhard A, Torresi L, Zhou C, Shao C, et al. Graph neural networks for materials science and chemistry. *Commun Mater.* 2022;3(93).
- Tom G, Schmid SP, Baird SG, Cao Y, Darvish K, Hao H, et al. Self-Driving laboratories for chemistry and materials science. *Chem Rev.* 2024;124(16).
<https://doi.org/10.1021/acs.chemrev.4c00055>.
- Baird SG, Sparks TD. What is a minimal working example for a self-driving laboratory? *Matter.* 2022;5(12):4170-8.
- Bayley O, Savino E, Slattery A, Noël T. Autonomous chemistry: Navigating self-driving labs in chemical and material sciences. *Matter.* 2024;7(7).

Hung L, Yager JA, Monteverde D, Baiocchi D, Kwon HK, Sun S, et al. Autonomous laboratories for accelerated materials discovery: A community survey and practical insights. *Digit Discov.* 2024;3.

<https://doi.org/10.1039/d4dd00059e>.

Kusne AG, Yu H, Wu C, Zhang H, Hatrick-Simpers J, DeCost B, et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat Commun.* 2020;11.

<https://doi.org/10.1038/s41467-020-19597-w>.

Kavalsky L, Hegde VI, Muckley E, Johnson MS, Meredig B, Viswanathan V. By how much can closed-loop frameworks accelerate computational materials discovery? *Digit Discov.* 2023;2.

<https://doi.org/10.1039/d2dd00133k>.

Stach E, DeCost B, Kusne AG, Hatrick-Simpers J, Brown KA, Reyes KG, et al. Autonomous experimentation systems for materials development: A community perspective. *Matter.* 2021;4(9).

<https://doi.org/10.1016/j.matt.2021.06.036>.

Sim M, Vakili MG, Strieth-Kalthoff F, Hao H, Hickman RJ, Miret S, et al. ChemOS 2.0: An orchestration architecture for chemical self-driving laboratories. *Matter.* 2024;7(9):2959-77.

Stier SP, Kreisbeck C, Ihssen H, Popp MA, Hauch J, Malek K, et al. Materials acceleration platforms (MAPs): Accelerating materials research and development to meet urgent societal challenges. *Adv Mater.* 2024;36(45).

<https://doi.org/10.1002/adma.202407791>.

Molkeri A, Khatamsaz D, Couperthwaite R, James J, Arróyave R, Allaire D, et al. On the importance of microstructure information in materials design: PSP vs PP. *Acta Mater.* 2022;223(117471).

<https://doi.org/10.1016/j.actamat.2021.117471>.

Nigon N, Simionescu DC, Ekstedt TW, Tucker JD, Koretsky MD. Towards an adaptive learning module for materials science: comparing expert predictions to student performance. InASEE Annual Conference proceedings 2022.

Afolabi SO, Akinsooto O. Theoretical framework for dynamic mechanical analysis in material selection for high-performance engineering applications. *Noûs.* 2021;3(4):45-62.