ORIGINAL RESEARCH Open access

Cascading Risk in Autonomous Materials Design: Governance Failure Propagation

Sanjay Kulkarni[1], Meenal Joshi[1*], Rohan Patil[2]

Abstract

The integration of artificial intelligence, robotics, and high-throughput computation has transformed materials engineering into a domain of autonomous discovery, where self-driving laboratories execute closed-loop experimentation at scales previously unattainable. These systems ingest vast datasets, train predictive models, and steer experimental campaigns toward novel materials with minimal human intervention, promising to compress discovery timelines from decades to months. Yet this autonomy introduces a distinct class of systemic vulnerabilities. Governance failures—misalignments in data integrity protocols, model validation regimes, or decision orchestration logics—do not remain isolated; they propagate through the computational pipeline, amplifying epistemic uncertainties and eroding the reliability of downstream materials outcomes. Existing literature has catalogued the technical foundations of these platforms, from Bayesian optimization in active learning to graph neural networks for property prediction and multi-fidelity workflows. However, a conceptual gap persists: the infrastructure-level dynamics of governance failure propagation remain largely unarticulated within the data-driven materials ecosystem. This manuscript introduces the Cascading Governance Failure Propagation (CGFP) Framework, an original systems architecture that reframes autonomous materials design as a layered computational process governed by interconnected control nodes. The framework elucidates how local misalignments in data curation, inference alignment, and steering logics cascade across pipelines, generating interpretive insights into workflow resilience and infrastructure trade-offs. By positioning governance as an intrinsic computational layer rather than an external overlay, the CGFP Framework offers a conceptual scaffold for designing more robust autonomous discovery ecosystems. Its implications extend to the sustainable scaling of data-driven materials engineering, where failure propagation must be anticipated as a core design constraint.

Keywords Self-driving laboratories, Data-driven pipelines, Computational steering, Autonomous materials design, Governance failure, Cascading risk

*Correspondence:
Meenal Joshi
meenal.joshi@gmail.com
[1] Department of Materials Data Science, Faculty of Engineering, Savitribai Phule Pune University, Pune, India
[2] Department of Computational Materials Systems, Faculty of Engineering, IIT Bombay, Mumbai, India

Introduction

The rise of autonomous paradigms in materials engineering

Materials discovery has historically relied on iterative, human-led experimentation constrained by the combinatorial explosion of chemical and structural space. The emergence of autonomous systems has fundamentally altered this landscape. Self-driving laboratories now orchestrate robotic platforms, machine learning models, and high-performance computing to execute closed-loop campaigns that span synthesis, characterization, and optimization [1-3]. These platforms leverage transfer learning across material families [4], Bayesian active learning for efficient search [5, 6], and large language models for hypothesis generation [7, 8], achieving acceleration factors that redefine the tempo of innovation.

In thin-film deposition, for instance, autonomous workflows have demonstrated the capacity to explore compositional spaces orders of magnitude larger than traditional methods [3]. Similarly, in solid-state synthesis, integrated platforms couple ab initio predictions with robotic execution to realize novel compounds at rates unattainable by manual workflows [9]. Such capabilities rest on the convergence of three computational pillars: (i) data-driven surrogate models that map composition to properties [10-12], (ii) automated experimentation hardware that executes protocols with high fidelity [1, 13], and (iii) decision engines that select experiments under uncertainty [6, 14]. Collectively, these elements form the backbone of a new materials engineering paradigm grounded in computational autonomy.

The data-driven ecosystem and its computational foundations

The data-driven ecosystem in materials science has matured through systematic efforts to curate, represent, and learn from large-scale repositories. Foundational databases and high-throughput density functional theory calculations have supplied the training corpora for graph-based neural networks that predict electronic, mechanical, and thermodynamic properties with increasing accuracy [12, 15, 16]. Unsupervised embeddings extracted from the literature have further enriched this ecosystem, capturing latent chemical knowledge that informs generative design [17].

Parallel advances in multi-fidelity modeling have enabled efficient navigation of vast design spaces by balancing computational cost with predictive fidelity [18]. These technical achievements have been complemented by infrastructure-level developments, including standardized data schemas and benchmarking suites that facilitate reproducible workflows [19]. The result is a mature computational substrate upon which autonomous platforms operate, transforming materials design from a craft into a scalable, data-intensive discipline [20-22].

Emerging challenges in governance for autonomous systems

Despite these advances, the governance of autonomous materials pipelines remains an underexplored dimension. As systems transition from supervised optimization to higher degrees of autonomy, the loci of control shift from human operators to algorithmic agents [7, 8, 23]. This shift introduces novel failure modes. A subtle bias in curated training data can propagate through model inference, leading to systematically skewed experimental selections [24, 25]. An undetected drift in robotic calibration can cascade into invalid characterization data, contaminating subsequent learning cycles [3, 13]. More critically, misalignments between the objectives encoded in steering logics and the broader requirements of materials deployment—such as synthesizability, stability, or environmental impact—can generate discovery pathways that appear locally optimal yet yield globally compromised outcomes [26, 27].

Literature on autonomous platforms has primarily emphasized performance metrics and acceleration benchmarks [1, 3, 6]. Governance considerations, when addressed, tend to appear as ancillary discussions of data management or ethical guidelines rather than as intrinsic computational structures [20, 22, 23]. This creates a conceptual blind spot: the propagation dynamics of governance failures across the layered architecture of autonomous discovery remain unmodeled. Failures do not terminate at their point of origin; they interact with feedback loops, amplifying through the pipeline and manifesting as systemic risk in the final materials candidates.

Positioning the cascading governance failure propagation framework

The present work addresses this gap by introducing the Cascading Governance Failure Propagation (CGFP) Framework. Unlike prior taxonomies of risk in materials informatics, the CGFP Framework treats governance not as a supervisory layer but as a distributed computational substrate embedded within the discovery pipeline. It conceptualizes autonomous materials design as a multi-layer system in which data, models, orchestration, and validation interact through explicit governance nodes. These nodes monitor alignment, detect divergence, and enact steering corrections.

By focusing on propagation dynamics rather than isolated failure events, the framework provides interpretive insights into how local misalignments escalate into pipeline-wide instabilities. It offers a conceptual lens for analyzing infrastructure trade-offs, representation–inference interactions, and the steering logics required to maintain epistemic integrity. The following sections synthesize the theoretical background from the literature and elaborate the

CGFP Framework as a novel contribution to computational materials engineering.

Theoretical Background & Literature Synthesis

Foundations of self-driving laboratories and closed-loop discovery

Self-driving laboratories represent the operational embodiment of autonomous materials design. Early demonstrations integrated robotic platforms with Bayesian optimization to accelerate reaction screening and thin-film optimization [3, 14]. Subsequent systems extended this paradigm to solid-state synthesis, coupling literature-derived recipes with active learning to realize dozens of novel compounds in continuous operation [8, 9]. These platforms operate on closed-loop principles: experimental outcomes inform model updates, which in turn guide the next iteration of synthesis and characterization [1, 2, 5].

The computational architecture underlying these laboratories typically comprises three interdependent modules: (i) a planning agent that proposes experiments under uncertainty, (ii) an execution layer that translates plans into robotic actions, and (iii) an analysis layer that extracts features from characterization data [6, 13]. Transfer learning has further enhanced scalability, allowing models trained on one material family to initialize exploration in related domains [4]. Such capabilities have been documented across chemistry, inorganic materials, and polymer spaces, demonstrating consistent acceleration relative to traditional workflows [1-3].

Advances in machine learning for materials property prediction and inverse design

Machine learning has supplied the predictive engine for autonomous discovery. Graph neural networks and equivariant architectures now achieve state-of-the-art performance in forecasting formation energies, band gaps, and mechanical properties directly from atomic structure [12, 15, 16, 28]. These models operate within universal frameworks that span the periodic table, enabling zero-shot generalization across chemistries [16]. Complementing forward prediction, generative approaches and inverse design methods have emerged to propose novel compositions that satisfy target property constraints [10, 27].

Unsupervised techniques, such as word embeddings derived from the scientific literature, have uncovered latent structure–property relationships that inform hypothesis generation [17]. Benchmarking efforts have standardized evaluation protocols, revealing that while predictive accuracy continues to improve, generalization to out-of-distribution materials remains a persistent challenge [19]. These advances collectively define the model layer of autonomous pipelines, where representation learning and inference intersect to steer experimental selection.

Integration of high-throughput computing and autonomous experimentation

High-throughput computation has provided the foundational datasets that fuel data-driven autonomy. Materials databases generated through systematic density functional theory calculations have enabled the training of surrogate models at unprecedented scale [29-31]. Multi-fidelity workflows further optimize this integration by allocating expensive simulations only where low-fidelity approximations prove insufficient [18]. When coupled with robotic experimentation, these computational resources create hybrid pipelines in which virtual screening precedes physical validation, dramatically expanding the explored design space [5, 6, 9].

Infrastructure developments, including standardized ontologies and cloud-based orchestration, have begun to support federated operation across laboratories [1, 20]. However, the literature reveals a consistent pattern: technical integration outpaces governance integration. Discussions of data provenance, model versioning, and validation protocols appear sporadically but lack systematic linkage to the propagation of failures across pipeline stages [20, 22, 23, 26].

Synthesizing the governance lacuna in data-driven pipelines

A synthesis of the literature exposes a structural lacuna. While the computational and experimental components of autonomous systems are richly described [1-6, 8, 9, 12-19, 26, 28], governance mechanisms are treated as external constraints rather than intrinsic architectural elements. Data

curation is acknowledged as critical [20, 21, 25], yet the downstream consequences of curation failures—such as biased active learning trajectories—are rarely traced through the full pipeline. Model alignment receives attention in the context of uncertainty quantification [6, 14, 18], but the propagation of misalignment into orchestration decisions remains unmodeled.

Steering logics, the algorithmic agents that select experiments, operate under objective functions that prioritize local metrics (e.g., predicted performance) over global governance criteria (e.g., synthesizability under realistic constraints) [7, 8, 27]. Feedback loops, while central to closed-loop operation [1-3], lack explicit governance nodes capable of detecting and correcting cascading divergence. The result is an ecosystem in which autonomy amplifies the impact of governance failures. A minor data inconsistency can cascade into model drift, misguided experimental selection, and ultimately materials candidates that fail downstream validation or deployment criteria.

This synthesis reveals that the computational materials community possesses mature tools for acceleration but lacks a unified conceptual framework for managing the governance dimension of autonomy. The CGFP Framework addresses this gap by embedding governance as a computational substrate and modeling its failure propagation explicitly.

Proposed conceptual framework

The Cascading Governance Failure Propagation (CGFP) framework

The Cascading Governance Failure Propagation (CGFP) Framework conceptualizes autonomous materials design ecosystems as stratified computational infrastructures in which governance is structurally embedded rather than externally imposed. Within contemporary self-driving discovery environments, oversight does not operate as a terminal checkpoint but as a distributed regulatory fabric interwoven across data ingestion, algorithmic inference, experimental orchestration, and translational validation. The framework therefore reframes governance from an administrative function into a computational systems property—one that co-evolves with discovery acceleration, epistemic uncertainty, and infrastructural autonomy. In this architecture, any disruption in governance integrity is not locally contained but propagates longitudinally across the discovery pipeline, amplifying distortions as they traverse computational and experimental strata.

At its foundational level, the Data Governance Layer constitutes the epistemic substrate upon which all downstream reasoning processes are constructed. Autonomous materials discovery systems ingest heterogeneous datasets spanning simulation repositories, robotic experimentation outputs, literature-derived extractions, and industrial process archives. These datasets differ widely in fidelity, descriptor completeness, thermodynamic coverage, and measurement uncertainty. Governance at this stage is therefore tasked with maintaining representational fidelity across compositional and structural design spaces. Without such regulation, uneven sampling densities produce algorithmic attention biases that skew predictive reasoning toward historically overrepresented chemistries. Provenance tracking further anchors each datapoint within a lineage architecture documenting simulation parameters, synthesis conditions, preprocessing transformations, and uncertainty annotations. This lineage infrastructure ensures that predictive outputs remain reproducible, auditable, and epistemically traceable. Integrity diagnostics operate concurrently, identifying distortions introduced by computational approximations, experimental noise, or publication bias. When governance weakens within this layer, structurally incomplete or biased datasets seed latent epistemic instabilities that propagate forward into model reasoning architectures.

The Model Inference Layer transforms curated datasets into predictive and generative intelligence capable of navigating vast materials search spaces. Within the CGFP architecture, governance at this stage operates through constraint alignment, uncertainty calibration, and objective oversight. Representation learning systems encode materials into latent embeddings that must remain physically plausible. Governance mechanisms therefore enforce symmetry compliance, thermodynamic feasibility, and conservation constraints, preventing the emergence of chemically invalid generative outputs. Concurrently, uncertainty quantification infrastructures monitor epistemic confidence across prediction regimes. Bayesian ensembles, evidential learning architectures, and probabilistic surrogate models collectively assess whether predictive certainty aligns with training data density. Governance failures here often manifest as overconfident predictions within sparsely sampled regions, creating algorithmic hallucinations that masquerade as discovery

signals. Objective functions are likewise subject to regulatory scrutiny. Optimization targets driven purely by performance maxima risk generating materials that are computationally optimal yet environmentally unsustainable, economically infeasible, or experimentally unrealizable. Distortions originating within this inference layer propagate forward as flawed experimental directives, embedding representational bias into laboratory execution pathways.

The Orchestration Layer operationalizes algorithmic predictions into executable experimental trajectories. It governs robotic synthesis platforms, high-throughput characterization infrastructures, and adaptive experimental design engines that collectively constitute the physical armature of autonomous discovery. Governance within this layer regulates how computational intent is translated into laboratory action. Experiment selection logics are continuously evaluated to ensure epistemic balance between exploratory probing and exploitative optimization. Without such oversight, closed-loop systems may prematurely converge on narrow design corridors, suppressing high-uncertainty exploration zones where transformative discoveries often reside. Resource allocation governance further mediates how laboratory time, precursor materials, robotic bandwidth, and energy expenditures are distributed across candidate experiments. Autonomous systems operating without such constraints risk infrastructural inefficiency or unsustainable operational footprints. Safety and synthesizability screening mechanisms provide an additional governance membrane, filtering proposed experiments through feasibility classifiers and hazard prediction systems prior to execution. When governance fractures within this orchestration stratum, laboratories may pursue experimentally infeasible candidates, allocate disproportionate resources to low-value search regions, or execute synthesis pathways that violate safety or regulatory thresholds.

The Discovery Output Layer represents the epistemic translation interface through which computational predictions are transformed into validated materials knowledge. Governance at this terminal yet feedback-connected stage ensures that discovery claims withstand empirical scrutiny and translational evaluation. Validation infrastructures benchmark predicted properties against experimental measurements, recalibrated simulations, and independent datasets, establishing reproducibility baselines. Interpretability governance interrogates whether observed performance arises from mechanistic phenomena or dataset artifacts, thereby safeguarding against spurious correlations being misinterpreted as scientific insight. Translational oversight extends beyond laboratory confirmation to assess manufacturability, lifecycle sustainability, regulatory compliance, and industrial scalability. Failures within this output layer produce epistemic false positives—materials that appear viable within computational environments but collapse under experimental replication or deployment conditions. Such breakdowns initiate upstream corrective cascades, triggering dataset re-curation, model recalibration, and orchestration redesign.

Across all four layers, the CGFP Framework is unified by a central discovery pipeline that flows from raw data ingestion to validated materials deployment. Governance nodes are embedded at each interlayer interface, functioning as regulatory checkpoints that evaluate alignment between system outputs and predefined integrity criteria. The transition from data governance to model inference, for instance, activates representativeness audits and bias diagnostics, while the transition from inference to orchestration evaluates synthesizability, safety, and feasibility constraints. These checkpoints operate not as static gates but as adaptive evaluative systems capable of learning from prior governance breaches.

Bidirectional feedback loops further reinforce the architecture’s resilience. Divergences detected at the discovery output stage—such as discrepancies between predicted and experimentally measured properties—propagate upstream to initiate corrective interventions. Training datasets may be re-curated, representation embeddings recalibrated, or optimization objectives redefined in response to validation failures. Computational steering logics implement these adjustments through differentiable optimization pathways or rule-encoded governance policies, enabling real-time infrastructural adaptation. In this sense, governance operates as a cybernetic control system embedded within autonomous discovery, continuously balancing exploratory ambition against epistemic accountability.

Through this layered architecture, the CGFP Framework demonstrates that governance failures are neither isolated nor terminal. Instead, they propagate as cascading distortions across interconnected computational and experimental strata, reshaping discovery trajectories, resource distributions, and scientific legitimacy. By rendering these propagation pathways visible, the framework provides a systems-level foundation for

designing oversight infrastructures capable of scaling alongside autonomous materials innovation.

The dynamics of failure propagation within the CGFP Framework can be conceptualized as:

$$Fl + 1 = \alpha l Fl + \beta l (1 - \Gamma l) \quad (1)$$

where F_l denotes the governance failure intensity at layer l, αl is the intrinsic propagation coefficient determined by layer interconnectivity , β_l represents external misalignment injected from upstream processes, and Γ_l ($0 \leq \Gamma_l \leq 1$) quantifies the governance alignment maintained by control nodes at layer l. This expression captures how even modest local failures F_l can escalate when alignment Γ_l degrades.

The corrective action enacted by steering logics may be expressed as:

$$\Delta Gl = \kappa l \cdot \partial E \partial Dl \quad (2)$$

where ΔG_l is the governance adjustment at layer l, κ_l is a layer-specific steering gain, E represents epistemic error accumulated across the pipeline, and D_l is the decision divergence at the current layer. This formulation formalizes steering as a gradient-driven process that minimizes misalignment.

Finally, the accumulation of systemic risk through repeated pipeline cycles captures the interaction between autonomy and governance investment:

$$R_s = \int_0^T P_{p(t)} \cdot \left(\frac{A(t)}{G(t)}\right) dt \quad (3)$$

where R_s is cumulative systemic risk, $P_{p(t)}$ is the instantaneous propagation probability, A(t) denotes the level of autonomy, and G(t) represents governance resource allocation. Higher autonomy amplifies risk unless counterbalanced by proportional governance investment. As conceptualized in **Figure 1**, governance operates as a vertically embedded computational substrate across the autonomous discovery pipeline, where local oversight failures propagate longitudinally through data, inference, orchestration, and validation strata.

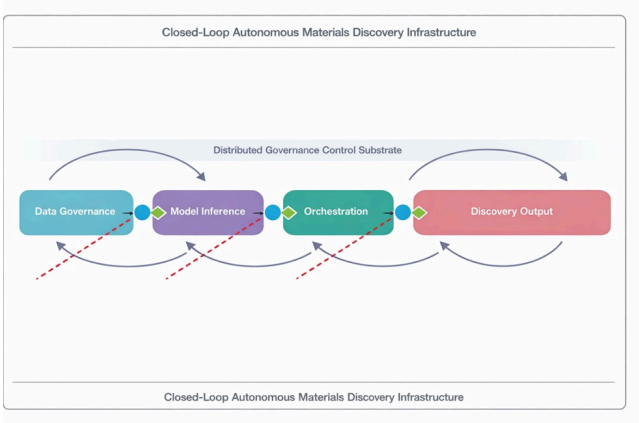

Figure 1. Cascading Governance Failure Propagation (CGFP) Framework architecture.

The schematic represents autonomous materials discovery as a horizontal closed-loop pipeline spanning data governance, model inference, experimental orchestration, and discovery validation. Vertically embedded governance layers operate as distributed control infrastructures, each containing monitoring nodes and intervention actuators that regulate epistemic alignment across system transitions. Dashed propagation vectors illustrate how governance failures originating in upstream strata cascade through inference and orchestration processes, ultimately shaping downstream discovery outcomes. Bidirectional feedback loops enable computational steering corrections, demonstrating the cybernetic coupling between oversight mechanisms and autonomous decision infrastructures. **Table 1** synthesizes the layer-specific governance functions, failure triggers, and propagation consequences mapped within the CGFP architecture.

Table 1. Governance Failure Propagation Dynamics Across Autonomous Discovery Layers

Discovery Layer	Primary Governance Function	Typical Failure Origin	Propagati Mechanis
Data Governance	Dataset curation, provenance tracking, bias diagnostics	Sampling asymmetry, descriptor sparsity, labeling noise	Encoded into trainir distributior and laten embeddin

Model Inference	Representation learning alignment, uncertainty calibration, constraint enforcement	Overfitting, miscalibrated uncertainty, objective misalignment	Translate into erroneous predictive confidenc and generativ outputs
Orchestration	Experiment selection, resource allocation, feasibility screening	Exploitative convergence, unsafe synthesis proposals, planning bias	Embedde in closed loop steeri policies
Discovery Output	Validation benchmarking, interpretability, translational assessment	Measurement drift, interpretive bias, scalability oversight	Feeds erroneous confirmatic signals upstream

The CGFP Framework thus provides a coherent interpretive structure for analyzing the governance dimension of autonomous materials systems. It shifts the focus from isolated technical components to the systemic dynamics of failure propagation, offering a foundation for designing pipelines that are both autonomous and governable.

Analytical implications

The Cascading Governance Failure Propagation (CGFP) Framework generates interpretive insights that reframe key trade-offs in autonomous materials design. At the systems level, it reveals how governance functions as a computational resource whose under-allocation relative to autonomy creates structural vulnerabilities. As pipelines scale toward full self-driving operation [1-3, 8, 9], the framework shows that local governance nodes—responsible for data integrity, inference alignment, and steering fidelity—must scale nonlinearly to counteract propagation effects. This is not a matter of adding oversight but of embedding governance as an active computational layer whose capacity directly modulates pipeline resilience.

The interaction between autonomy and governance investment can be conceptualized as:

$$\Lambda = G \cdot (1+\phi) A2 \quad (4)$$

Where Λ represents the latent propagation multiplier, A is the normalized autonomy index (0–1), G is the governance capacity index, and ϕ is a layer-specific interconnectivity factor. The quadratic dependence on autonomy captures how even modest increases in closed-loop freedom amplify failure intensity when governance lags, providing a quantitative lens for infrastructure design decisions without invoking empirical thresholds.

A second implication concerns representation–inference interactions within the model layer. Graph-based architectures [12, 15, 16, 28] and equivariant networks [28] encode chemical knowledge through learned embeddings, yet these representations remain susceptible to upstream data governance failures. The CGFP Framework interprets such failures as inducing a representational drift that propagates into inference outputs, ultimately skewing orchestration decisions toward chemically plausible but governance-misaligned candidates. This dynamic underscores an epistemic trade-off: richer representations enhance predictive power [10, 11, 17] but simultaneously increase the surface area for failure propagation unless explicit alignment nodes are present.

Finally, the framework illuminates discovery steering logics as sites of cumulative risk. Steering agents [6-8, 14] optimize under uncertainty using Bayesian or language-model-driven mechanisms, yet without embedded governance, they can converge on locally optimal but globally unstable pathways. The accumulation of systemic risk across repeated cycles is captured by:

$$R_{c(t)} = R_0 + \int \left(\sum l = 1 L \alpha l F l(\tau)\right) e^{-\gamma (t-\tau)} d\tau \quad (5)$$

Where $R_{c(t)}$ is cumulative risk at cycle t, R_0 is baseline risk, αl are layer-specific propagation rates, F_l denotes instantaneous failure intensity, and γ is a decay constant representing corrective feedback efficacy. This integral formulation interprets risk as a memory-dependent process, where earlier governance lapses continue to influence later stages unless actively damped by steering interventions.

These analytical implications position the CGFP Framework as a tool for infrastructure-level reasoning. It

shifts design conversations from component optimization to the orchestration of governance capacity across layers, offering computational materials engineers a structured way to anticipate and mitigate the hidden costs of autonomy.

Results and Discussion

The CGFP Framework integrates with and extends the existing literature on autonomous materials systems by foregrounding governance as an intrinsic computational substrate rather than an external ethical or data-management concern. Where prior work has catalogued technical accelerations [1-3, 5, 9] and identified perils of unchecked machine learning in chemical spaces [23], the framework provides a unified interpretive architecture that traces how isolated misalignments become systemic through propagation. It complements studies on multi-fidelity modeling [18] and active learning [5, 6, 24] by showing that uncertainty quantification alone is insufficient when governance nodes are absent from the pipeline architecture.

Infrastructure-level analysis reveals a consistent pattern across the synthesized literature: the most advanced platforms [3, 8, 9, 13] achieve remarkable experimental throughput yet operate with governance mechanisms that remain largely implicit. The CGFP Framework makes these mechanisms explicit, demonstrating how feedback loops—central to closed-loop discovery [1, 2]—can either dampen or amplify divergence depending on the strength of interlayer governance. This insight has direct bearing on the design of next-generation ecosystems, where federated laboratories [20] and cross-platform transfer learning [4] will further increase interconnectivity and, by extension, propagation potential. Operationalizing governance as an infrastructural variable requires identifying the intervention levers capable of dampening cascading distortions. These mitigation capacities and their systemic effects are structured in **Table 2**.

Table 2. Governance Intervention Levers and Mitigation Capacity in Autonomous Discovery Pipelines

Governance Intervention Lever	Targeted Failure Domain	Operational Mechanism	Mitigati Capaci
Adaptive dataset rebalancing	Representational bias	Dynamic sampling redistribution	Reduce embedd skew
Provenance graph enforcement	Data lineage discontinuity	Blockchain / graph-based traceability	Restore auditabi
Uncertainty recalibration engines	Predictive overconfidence	Bayesian ensemble correction	Aligns confiden with da density
Multi-objective optimization governance	Objective misalignment	Sustainability + feasibility constraints	Broade discove criteria
Autonomous feasibility classifiers	Unsafe synthesis pathways	Pre-execution screening models	Filters hazardo experime
Cross-platform validation consortia	Benchmarking fragility	Federated replication testing	Strengthe confirma fidelity
Interpretability audit modules	Mechanistic opacity	Feature attribution + causal inference	Detect spuriou correlati
Governance resource scaling protocols	Autonomy–oversight imbalance	Dynamic allocation of oversight compute	Dampe propagat coefficie

Epistemic risk structures emerge as a central theme. The framework interprets data-driven materials engineering as an environment in which epistemic integrity is not a static property of models [10-12, 16] but a dynamic outcome of governance interactions. Failures in literature-derived embeddings [17] or benchmark datasets [19] do not remain confined; they cascade into discovery outputs that may satisfy surrogate objectives while violating broader materials engineering constraints such as long-term stability or manufacturability. By modeling these interactions, the CGFP Framework encourages a shift from performance-centric to resilience-centric pipeline design.

Limitations of the present conceptual contribution are inherent to its interpretive nature. The framework does not

prescribe specific implementations for governance nodes nor quantify exact propagation coefficients; these remain open challenges for future computational realization. Nevertheless, it supplies a coherent language and structural logic for addressing them. As autonomous systems continue to scale [7-9], the ability to reason about governance failure propagation will become a core competency in computational materials engineering, determining not only the speed but the reliability of discovery.

Conclusion

The Cascading Governance Failure Propagation (CGFP) Framework introduces a novel systems-level perspective on autonomous materials design. By treating governance as a distributed computational layer embedded within data–model–orchestration–output pipelines, it elucidates how local misalignments escalate into pipeline-wide instabilities through feedback-amplified propagation. The framework's structural layers, steering logics, and conceptual formalizations provide interpretive tools for analyzing infrastructure trade-offs, representation–inference dynamics, and epistemic risk accumulation in data-driven discovery ecosystems.

Positioned at the intersection of computational autonomy and materials engineering, the CGFP Framework offers a conceptual scaffold for designing pipelines that are simultaneously high-throughput and governable. Its implications extend beyond individual platforms to the sustainable evolution of the field, where the scaling of self-driving laboratories must be accompanied by commensurate governance architectures. As the materials community advances toward increasingly autonomous workflows, attention to failure propagation dynamics will be essential to maintaining the integrity of discovered knowledge.

The framework thus contributes an original lens through which to view the future of computational materials engineering—one in which governance is not an afterthought but a foundational element of autonomous discovery.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 06 Aug 2024 Revised: 25 Sep 2024 Accepted: 31 Oct 2024
Published online: 18 March 2025

References

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):9633-732.

Pyzer-Knapp EO, Pitera JW, Staar PWJ, Takeda S, Laino T, Sanders DP, et al. Accelerating materials discovery using artificial intelligence, high performance computing and robotics. npj Comput Mater. 2022;8:84. https://doi.org/10.1038/s41524-022-00765-z.

MacLeod BP, Parlane FGL, Morrissey TD, Häse F, Roch LM, Dettelbach KE, et al. Self-driving laboratory for accelerated discovery of thin-film materials. Sci Adv. 2020;6(20):eaaz8867. https://doi.org/10.1126/sciadv.aaz8867.

Yoshida N, Iwabuchi Y, Igarashi Y, Iwasaki Y. Networking autonomous material exploration systems through transfer learning. npj Comput Mater. 2025;11:362. https://doi.org/10.1038/s41524-025-01851-8.

Kusne AG, Yu H, Wu C, Zhang H, Hattrick-Simpers J, DeCost B, et al. On-the-fly closed-loop materials discovery via Bayesian active learning. Nat Commun. 2020;11(1):5966.

Noack MM, Doerk GS, Li R, Streit JK, Vaia RA, Yager KG, et al. Autonomous materials discovery driven by Gaussian process regression with inhomogeneous measurement noise and anisotropic kernels. Sci Rep. 2020;10(1):17663. https://doi.org/10.1038/s41598-020-74394-1.

Miret S, Krishnan NMA. Enabling large language models for real-world materials discovery. Nat Mach Intell. 2025;7:991-8. https://doi.org/10.1038/s42256-025-01058-y.

Boiko DA, MacKnight R, Kline B, Gomes G. Autonomous chemical research with large language models. Nature. 2023;624(7992):570-8. https://doi.org/10.1038/s41586-023-06792-0.

Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. Nature. 2023;624(7990):80-5.

Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018;559(7715):547-55. https://doi.org/10.1038/s41586-018-0337-2.

Schmidt J, Marques MRG, Botti S, Marques MAL. Recent advances and applications of machine learning in solid-state materials science. npj Comput Mater. 2019;5:83. https://doi.org/10.1038/s41524-019-0221-0.

Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. Chem Mater. 2019;31(9):3564-72. https://doi.org/10.1021/acs.chemmater.9b01294.

DeCost BL, Hattrick-Simpers JR, Trautt Z, Kusne AG, McClure E, Borg CK, et al. Scientific AI for materials discovery. Matter. 2023;6(9):2677-80.

Shields BJ, Stevens J, Li J, Parasram M, Damani F, Alvarado JIM, et al. Bayesian reaction optimization as a tool for chemical synthesis. Nature. 2021;590(7844):89-96. https://doi.org/10.1038/s41586-021-03213-y.

Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. Phys Rev Lett. 2018;120(14):145301. https://doi.org/10.1103/PhysRevLett.120.145301.

Chen C, Ong SP. A universal graph deep learning interatomic potential for the periodic table. Nat Comput Sci. 2022;2:718-28. https://doi.org/10.1038/s43588-022-00349-3.

Tshitoyan V, Dagdelen J, Weston L, Dunn A, Rong Z, Kononova O, et al. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature. 2019;571(7763):95-8. https://doi.org/10.1038/s41586-019-1335-8.

Pilania G, Gubernatis JE, Lookman T. Multi-fidelity machine learning models for accurate bandgap predictions of solids. Comput Mater Sci. 2017;129:156-63.

Dunn A, Wang Q, Ganose A, Dopp D, Jain A. Benchmarking materials property prediction methods: The Matbench test set and Automatminer reference algorithm. npj Comput Mater. 2020;6:138. https://doi.org/10.1038/s41524-020-00406-3.

Himanen L, Geurts A, Foster AS, Rinke P. Data-Driven materials science: Status, challenges and perspectives. Adv Sci. 2019;6(21):1900808. https://doi.org/10.1002/advs.201900808.

Wang AY-T, Murdock RJ, Kauwe SK, Oliynyk AO, Gurlo A, Brgoch J, et al. Machine learning for materials scientists: An introductory guide toward best practices. Chem Mater. 2020;32(12):4954-65.

Morgan D, Jacobs R. Opportunities and challenges for machine learning in materials science. Annu Rev Mater Res. 2020;50:71-103.

Shankar S, Zare RN. The perils of machine learning in designing new chemicals and materials. Nat Mach Intell. 2022;4:314-5.

Lookman T, Balachandran PV, Xue D, Yuan R. Active learning in materials science. npj Comput Mater. 2019;5:21.

Jha D, Ward L, Paul A, Liao W-k, Choudhary A, Wolverton C, et al. ElemNet: Deep learning the chemistry of materials from only elemental composition. Sci Rep. 2018;8:17593. https://doi.org/10.1038/s41598-018-35934-y.

Zhao Y, Cui Y, Xiong Z, Jin J, Liu Z, Dong R, et al. Machine learning-based prediction of crystal systems and space groups from inorganic materials compositions. ACS omega. 2020;5(7):3596-606.

Zunger A. Inverse design in search of materials with target functionalities. Nat Rev Chem. 2018;2:0121. https://doi.org/10.1038/s41570-018-0121.

Batatia I, Kovács DP, Simm GNC, Ortner C, Csányi G. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. Adv Neural Inf Process Syst. 2022;35:11423-36.

Yuan WL, He L, Tao GH, Shreeve JN. Materials-genome approach to energetic materials. Acc Mater Res. 2021;2(9):692-6.

Liu C, Zhang Y, Zhang T, Wu X, Gao L, Zhang Q. High throughput vehicle coordination strategies at road intersections. IEEE Trans Veh Technol. 2020;69(12):14341-54.

Luo S, Li T, Wang X, Faizan M, Zhang L. High-throughput computational materials screening and discovery of optoelectronic semiconductors. Wiley Interdiscip Rev Comput Mol Sci. 2021;11(1):e1489.