

ORIGINAL RESEARCH

Open access

Algorithmic Steering Without Consent: Control Asymmetries in AI-Directed Materials Exploration

Ahmed Mansour^{1*}, Omar Saeed¹, Lina Hassan²

Abstract

In the evolving landscape of computational materials engineering, artificial intelligence (AI) has emerged as a pivotal orchestrator, directing exploratory pipelines from data curation to predictive modeling and synthesis validation. This integration, while accelerating discovery, introduces profound control asymmetries wherein algorithmic decisions preempt human oversight, often without explicit consent mechanisms embedded in the workflow. Such asymmetries manifest as latent divergences between intended exploratory intents and AI-mediated trajectories, potentially skewing material property predictions and optimization paths in unintended directions. Drawing from systems-level analyses of machine learning applications in solid-state materials science, generative sampling strategies, and active learning protocols, this manuscript conceptualizes these dynamics through an original interpretive framework: the Asymmetric Steering Topology (AST). The AST delineates layered interactions across data ingestion, model inference, and discovery actuation, highlighting feedback loops that amplify epistemic risks in unconsented steering. By interpreting these asymmetries as infrastructural tensions—between representational fidelity and inferential autonomy—the framework elucidates how AI-directed exploration can inadvertently prioritize computational efficiency over exploratory equity. Implications for the field include reimagined pipeline architectures that integrate consent-aware safeguards, fostering more equitable human-AI symbiosis in materials informatics. This conceptual synthesis advances understanding of discovery steering logics, urging a shift toward epistemically resilient infrastructures that balance algorithmic prowess with interpretive sovereignty in data-driven materials engineering.

Keywords Materials informatics, Discovery pipelines, Algorithmic steering, Control asymmetry, AI-directed exploration, Consent mechanisms

*Correspondence:

Ahmed Mansour
ahmed.mansour@gmail.com

¹ Department of Materials Engineering and Data Modeling, Faculty of Engineering, Cairo University, Cairo, Egypt

² Department of Computational Materials Systems, Faculty of Engineering, Alexandria University, Alexandria, Egypt

Introduction

The rise of AI in computational materials engineering

Computational materials engineering has undergone a paradigm shift with the infusion of artificial intelligence, transforming disparate data repositories into cohesive discovery engines. At its core, this evolution leverages machine learning to bridge the chasm between atomic-scale simulations and macroscopic property predictions, enabling rapid iteration across vast chemical spaces [1]. Early applications focused on predictive modeling for solid-state properties, where neural networks and kernel methods distilled patterns from density functional theory outputs into actionable insights [2]. This foundational layer has since expanded into generative paradigms, where adversarial networks sample underrepresented composition domains, inverting traditional forward-design workflows [3]. The resultant acceleration—

evident in closed-loop Bayesian optimizations that autonomously refine experimental proposals [4]—positions AI not merely as an analytical tool but as an active navigator of exploratory terrains.

Yet, this navigational role introduces subtleties in how discovery unfolds. In data-driven ecosystems, AI algorithms ingest heterogeneous datasets, from high-throughput screening outputs to spectroscopic embeddings, to forge surrogate models that approximate material behaviors [5]. These models, in turn, steer subsequent inquiries by prioritizing regions of parameter space deemed high-yield based on uncertainty quantification or gradient-based exploration [6]. Such steering, while computationally elegant, embeds a form of directional bias: algorithms optimize for convergence metrics that may diverge from the nuanced intents of human investigators, such as serendipitous off-pathway discoveries or interdisciplinary integrations [7]. This divergence underscores a broader infrastructural tension in materials exploration, where the opacity of black-box inferences challenges the transparency required for reproducible science [8].

Unpacking control asymmetries in discovery pipelines

Control asymmetries arise precisely at the intersection of these elements: the uneven distribution of decision-making authority between human designers and algorithmic executors. In conventional computational workflows, human agents define objectives—say, targeting bandgap tunability in perovskites—through explicit parameter bounds and validation criteria [9]. AI augmentation extends this by automating intermediate steps, such as active sampling via Gaussian processes that query low-confidence regions [4]. However, the "without consent" dimension emerges when these automations cascade: an initial model perturbation, perhaps from noisy training data, can propagate through reinforcement loops, locking the pipeline into suboptimal attractors without recourse for mid-course recalibration [10].

Consider the epistemic implications. Materials informatics thrives on representational fidelity—the degree to which data encodings capture underlying physicochemical invariances [11]. Yet, AI-directed steering often privileges inferential speed, employing dimensionality reductions or transfer learning that elide subtle correlations, such as solvent effects in polymer genomes [12]. This selective emphasis creates asymmetries wherein algorithmic control outpaces human interpretive capacity, manifesting as "steering drift": gradual deviations from consented exploratory vectors toward computationally tractable but conceptually narrow subspaces [13]. Literature on molecular simulations highlights analogous risks, where force field approximations, when learned via kernel ridge regression, introduce non-local errors that compound in long-horizon predictions [14]. In materials contexts, these drifts are amplified by the high-dimensionality of composition spaces, where generative models might overfit to synthetic augmentations, sidelining rare-earth dopant synergies [15].

From an infrastructural vantage, these asymmetries are not mere technical artifacts but systemic features of AI integration. Pipelines in npj Computational Materials exemplify this: high-performance computing clusters orchestrate parallel evaluations, yet the orchestration logic—often encoded in proprietary optimizers—obscures how steering decisions allocate resources [10]. Human consent, typically granted at workflow inception via protocol approvals, erodes as real-time adaptations occur sans explicit ratification. This erosion parallels broader machine learning deployments in chemistry, where predictive insights from graph neural networks inadvertently encode dataset biases, skewing electrocatalyst designs toward overrepresented ligands [8]. Thus, control asymmetries in AI-directed materials exploration demand a conceptual reframing, one that interrogates not just efficacy but the equity of steering agency.

Epistemic risks and infrastructure trade-offs

Delving deeper, epistemic risks compound these asymmetries by undermining the reliability of discovery outcomes. In data-driven paradigms, knowledge generation hinges on inference chains: from raw spectral data to latent embeddings, then to property extrapolations [9]. AI steering, by dynamically weighting these chains, can introduce fragility—e.g., adversarial perturbations that flip classification boundaries in perovskite process diagnostics [5]. Such vulnerabilities are infrastructural, rooted in trade-offs between model expressivity and computational tractability. Expressive architectures, like deep equivariant networks for crystal structure generation, demand vast training corpora, yet sparse materials datasets foster overfitting, wherein steering logics chase illusory optima [16].

Trade-offs extend to feedback integration. Closed-loop systems, lauded for their autonomy in refining synthesis routes [17], embed loops that feedback inferred properties into data pools, potentially entrenching early asymmetries. If initial steering favors low-entropy subspaces—common in entropy-regularized explorations [18]—subsequent iterations may marginalize high-variance regimes harboring novel phases [19]. This dynamic echoes warnings in materials fantasy predictions, where plausible but unphysical outputs arise from unchecked generative drifts [7]. Moreover, in porous materials genomics, big-data approaches reveal how machine learning amplifies sampling biases, steering toward structurally similar frameworks at the expense of topological diversity [20].

These risks are not isolated but interwoven with ethical dimensions of consent. Consent, in this context, denotes not just initial data usage agreements but ongoing ratification of steering trajectories—a mechanism absent in most current pipelines [21]. As AI assumes greater control, the field confronts an epistemic bind: harnessing algorithmic efficiency while preserving human sovereignty over interpretive horizons. This bind necessitates infrastructures that render steering transparent, perhaps through provenance tracking in active learning cycles [6].

Positioning the asymmetric steering topology framework

To navigate these complexities, this manuscript introduces the Asymmetric Steering Topology (AST) framework, an original interpretive lens for dissecting control asymmetries in AI-directed materials exploration. The AST conceptualizes steering as a topological manifold, wherein data-model-discovery interactions form layered adjacencies that encode consent gradients. By mapping these adjacencies, the framework illuminates infrastructural leverage points for mitigating unconsented drifts, fostering resilient discovery ecosystems. Grounded in syntheses of machine learning prospects [2] and adaptive design strategies [6], the AST advances a systems-level understanding, positioning AI not as an autonomous pilot but as a co-navigator attuned to human-directed equities.

Theoretical Background & Literature Synthesis

Foundations of AI-directed discovery pipelines

The theoretical underpinnings of AI in materials engineering trace to informatics paradigms that recast materials as navigable data manifolds. Central to this is the notion of surrogate modeling, where kernel-based regressions approximate potential energy surfaces from ab initio calculations, enabling scalable explorations [2]. This foundation evolved with the advent of deep learning, as convolutional architectures parsed microstructural motifs in alloys, informing phase stability predictions [1]. Pipelines thus emerge as sequential funnels: data acquisition via high-throughput density functional theory, followed by feature engineering—often via autoencoders—to distill invariant descriptors [11]. Inference layers then deploy these descriptors in probabilistic frameworks, such as variational autoencoders for inverse design, sampling latent spaces to propose candidate compositions [3].

Synthesis across literature reveals a progression toward autonomy. Early works emphasized supervised mappings, linking elemental fingerprints to thermomechanical properties [22]. Subsequent integrations incorporated uncertainty propagation, with ensemble methods quantifying aleatoric variances in glass transition forecasts [15]. This autonomy crescendoed in active learning architectures, where Bayesian optimizers iteratively query datasets, steering toward property optima in multi-objective landscapes [4]. Yet, theoretical critiques highlight representational pitfalls: compressed sensing descriptors, while parsimonious, may underspecify electronic correlations, leading to inference asymmetries in bandgap estimations [23].

Infrastructurally, these pipelines interlace with high-performance computing, as seen in robotic validations of AI-proposed perovskites [10]. The synthesis underscores a core dynamic: steering as emergent from data-model feedbacks, where gradient flows in neural landscapes dictate exploratory branching [14]. However, this emergence often obfuscates consent boundaries, as algorithmic updates—via federated learning in distributed clusters—evolve without human veto [8].

Representational and inferential dynamics in materials informatics

Representational dynamics form the bedrock of AI-directed steering, wherein data encodings mediate between physical realities and computational abstractions. Graph neural networks, for instance, embed atomic connectivities as adjacency matrices, propagating messages to infer bulk moduli [12]. Literature synthesizes these as fidelity trade-offs: high-fidelity representations, like SOAP kernels in Gaussian approximations, capture local environments but scale poorly to disordered systems [18]. In contrast, low-dimensional projections—principal components from spectral decompositions—facilitate rapid inference yet erode nuance, as in X-ray absorption diagnostics for layered oxides [9].

Inferential layers amplify these trade-offs through probabilistic steering. Markov chain Monte Carlo explorations sample phase diagrams, guided by energy surrogates learned from molecular dynamics trajectories [14]. Synthesis reveals epistemic tensions: inferences grounded in maximum likelihood may converge to local minima, sidelining global optima in high-entropy alloys [24]. Active paradigms mitigate this via acquisition functions that balance exploitation and exploration, yet theoretical analyses expose asymmetries—e.g., entropy penalties that disproportionately prune rare-event subspaces [6]. In polymer informatics, genome-scale predictions via random forests highlight how feature selection logics steer toward linear correlations, marginalizing nonlinear synergies [12].

Feedback integrations further complicate dynamics. Closed-loop systems, as in electrocatalyst discovery, recycle experimental feedbacks to refine models, embedding historical biases into future steers [25]. Literature on porous frameworks illustrates this: machine learning-driven screenings favor zeolitic topologies, as representational biases in voxel encodings undervalue aperiodic pores [20]. Collectively, these dynamics portray inference not as neutral computation but as directional force, imprinting asymmetries onto discovery topologies [13].

Steering logics and epistemic risk structures

Steering logics operationalize these dynamics, encoding decision hierarchies that propel pipelines forward. Gradient-based optimizers, prevalent in generative adversarial networks for composition sampling, descend loss landscapes to unearth stable phases [3]. Theoretical grounding in optimal control theory frames this as Hamiltonian paths through state spaces, where adjoint sensitivities dictate query selections [4]. Synthesis across domains—from scintillator bandgap tuning [21] to soft-magnetic glass design [15]—reveals logics as layered: local (e.g., nearest-neighbor heuristics in crystal predictions) to global (e.g., multi-fidelity bootstrapping across scales [12]).

Epistemic risks structure these logics, manifesting as divergence metrics between steered and baseline trajectories. Kullback-Leibler divergences, for example, quantify how learned force fields deviate from quantum references, risking unphysical extrapolations [18]. In materials contexts, risks escalate in sparse-data regimes: transfer learning from organic datasets to inorganics introduces domain shifts, steering perovskite optimizations toward metastable traps [5]. Literature on fantasy materials cautions against overconfident inferences, where steering toward high-plausibility low-reality candidates erodes trust [7].

Infrastructural analyses deepen this: resource allocation in cloud-based pipelines favors steerable submanifolds, as parallel tempering in molecular simulations privileges ergodic cores [14]. Consent disruptions arise here, as real-time steering—via online learning—bypasses static protocols, amplifying risks in safety-critical applications like battery electrolytes [17]. Synthesizing adaptive designs [6] with big-data genomics [20], a pattern emerges: risks as topological invariants, persistent across scales unless explicitly unraveled.

Infrastructural trade-offs in human-AI symbiosis

Trade-offs in human-AI symbiosis cap this synthesis, balancing autonomy against interpretability. Explanatory tools, like SHAP values in tree ensembles for property regressions [15], render steering legible yet computational overhead constrains their deployment in real-time loops [8]. In synthesis routes for intermetallics, trade-offs pit exhaustive enumerations against heuristic prunings, where AI steering resolves via reinforcement but at the cost of provenance loss

[25]. Literature on compressed descriptors [23] frames this as dimensionality curses: richer representations enhance steering fidelity but inflate inference latencies, skewing toward tractable approximations.

Epistemically, symbiosis demands hybrid logics—human vetoes interleaved with algorithmic proposals—as in robotic validations [10]. Yet, asymmetries persist: human intents, articulated via natural language objectives, translate imperfectly into loss functions, fostering drift [13]. Synthesizing molecular property optimizations [19] with inorganic screenings [21], trade-offs reveal themselves as consent gradients: steeper in opaque architectures, shallower in modular ones. This backdrop positions the field at an inflection, where theoretical syntheses illuminate pathways to equilibrated steering, sans unconsented overrides.

Proposed conceptual framework

Introducing the asymmetric steering topology

The Asymmetric Steering Topology (AST) framework offers an original interpretive architecture for unpacking control asymmetries in AI-directed materials exploration. Unlike prior taxonomies that segment pipelines into discrete modules [2], the AST conceives steering as a topological construct—a manifold of interconnected adjacencies where data, models, and discovery phases fold into consent-sensitive curvatures. This topology captures the non-Euclidean geometry of exploratory flows: straight-line intents from human designers warp under algorithmic influences, forming geodesics that may loop into self-reinforcing basins or diverge into uncharted fringes [4]. By layering these adjacencies, the AST elucidates how unconsented steering emerges not as error but as intrinsic curvature, arising from mismatched representational scales and inferential horizons.

Structurally, the AST comprises three primary layers: the Data Adjacency Layer (DAL), the Inference Curvature Layer (ICL), and the Discovery Geodesic Layer (DGL). The DAL encodes input manifolds, where raw materials data—spanning compositional vectors to spectral tensors—form adjacency matrices that precondition steering vectors [11]. Curvatures in the ICL then mediate model-mediated bends, with kernel-induced metrics quantifying how surrogate approximations distort exploratory paths [18]. Finally, the DGL projects these onto actuation spaces, where synthesis proposals trace geodesics informed by probabilistic feedbacks [17]. Interlayer feedbacks, modeled as Ricci flows, dynamically adjust curvatures to preserve consent invariants, preventing drift amplification [6].

Pipeline dynamics: From data to model to discovery

Within the AST, data-to-model pipelines unfold as adjacency propagations. In the DAL, data ingestion constructs a homology complex, where simplicial chains represent feature interdependencies—e.g., elemental radii chained to electronic densities [12]. Steering initiates here via embedding projections, which may asymmetrically contract subspaces: high-dimensional alloy spectra collapse into low-rank approximations, privileging dominant modes at the expense of outlier motifs [15]. Transitioning to the ICL, these embeddings feed into curvature operators—neural message-passing graphs that compute sectional curvatures, reflecting how local decisions (e.g., atomistic force perturbations) ripple globally [14]. Model inference thus steers by minimizing geodesic distances to target properties, yet asymmetries arise when curvature mismatches human-prescribed metrics, such as equitable exploration across phase boundaries [9].

Discovery pipelines in the DGL extend this via geodesic integrations, where active queries along principal curvatures propose experimental waypoints [4]. Feedback loops close the circuit: validated outcomes, encoded as boundary conditions, induce DAL refinements, evolving the topology over iterations. This evolution captures computational steering logics as Lagrangian multipliers on the manifold, enforcing constraints like resource budgets while navigating consent gradients—regions where human ratification thresholds steepen [10]. The AST interprets these logics as equilibrium-seeking: pipelines stabilize when interlayer adjacencies align, but unconsented perturbations—e.g., from noisy augmentations—induce Ricci instabilities, fracturing geodesics into fragmented exploratory shards [20].

Feedback loops and steering logics

Feedback loops in the AST are conceptualized as conformal mappings, preserving angles of intent while scaling magnitudes of control. Positive loops, prevalent in generative sampling, amplify successful geodesics: a model's high-fidelity prediction reinforces DAL adjacencies, tightening curvatures toward convergent basins [3]. Negative loops, conversely, introduce damping via uncertainty escalations, broadening ICL horizons to recapture marginalized subspaces [6]. Steering logics orchestrate these via topological invariants—Betti numbers tallying "holes" in the manifold, signaling underexplored voids that demand rerouting [16].

A key insight lies in consent-embedded logics: thresholds modeled as sectional curvatures that flare when algorithmic autonomy exceeds interpretive bandwidth, prompting human adjacency insertions [8]. This fosters hybrid topologies, where human-directed simplices—e.g., serendipity heuristics—interweave with automated flows, mitigating asymmetries without sacrificing velocity [13]. The AST thus reveals steering not as monolithic but as polyadic: multiple geodesics coexisting, with dominance dictated by curvature hierarchies.

To formalize interlayer feedbacks, the propagation of steering asymmetries can be conceptualized as a Ricci flow equation adapted to exploratory manifolds:

$$\frac{\partial g_{ij}}{\partial t} = -2R_{ic}(g_{ij}) + \Lambda \cdot \nabla C \quad (1)$$

where g_{ij} denotes the metric tensor on the AST manifold, R_{ic} the curvature capturing inferential bends, and $\Lambda \cdot \nabla C$ a consent gradient term modulating flow rates based on human ratification vectors. This expression illustrates how unchecked curvatures (high R_{ic}) contract pipelines, while consent infusions ∇C expand equitable paths, preserving topological volume across discovery iterations [7, 26].

As conceptualized in **Figure 1**, the Asymmetric Steering Topology manifests as a tri-layered manifold in which data adjacencies, inferential curvatures, and discovery geodesics interact through consent-modulated feedback flows.

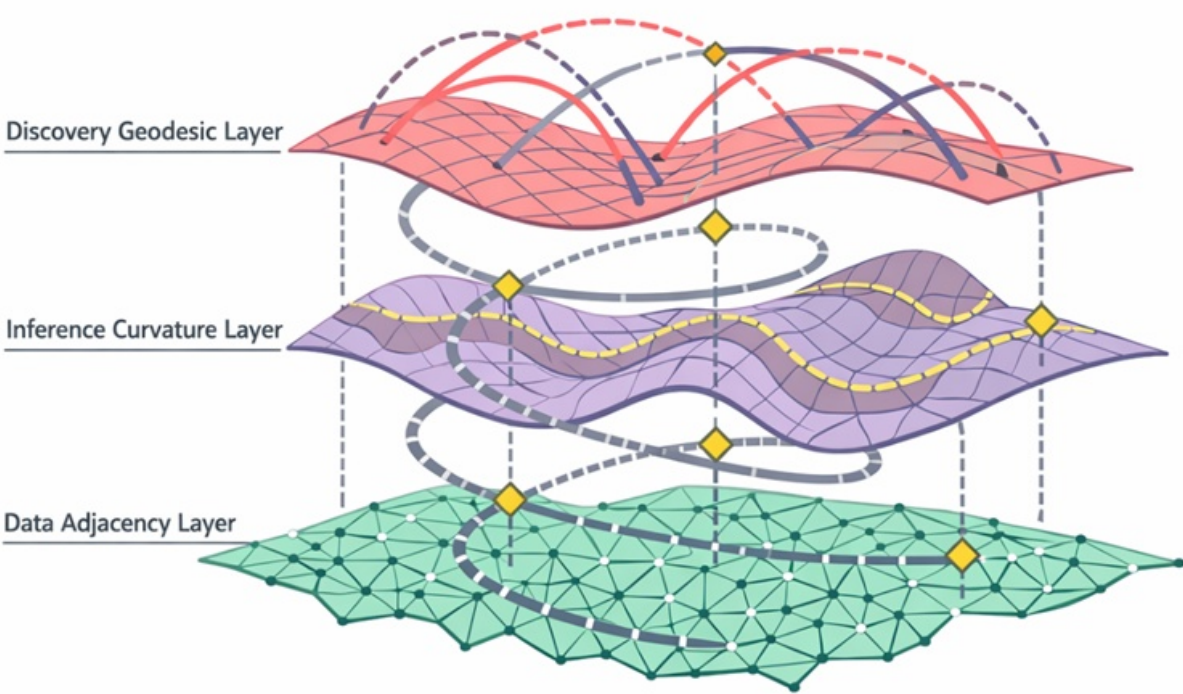


Figure 1. Asymmetric Steering Topology (AST) framework

A tri-layered topological manifold depicting control asymmetries in AI-directed materials exploration. The Data Adjacency Layer encodes representational proximities that precondition inferential reasoning. Above it, the Inference Curvature Layer visualizes algorithmic bending induced by surrogate modeling and probabilistic steering logics. The Discovery Geodesic Layer projects exploratory trajectories shaped by these curvatures. Consent gradients regulate interlayer propagation, while feedback Ricci flows iteratively reshape manifold geometry. Steering drift emerges where representational sparsity and inferential autonomy misalign, generating unconsented exploratory deviations.

Complementing this, the interaction between representational fidelity and steering drift may be expressed as:

$$D = \int_M \kappa \cdot (1 - f(R)) \, dV \tag{2}$$

where D quantifies drift magnitude over manifold M , κ local curvature, $f(R)$ representational fidelity (a scalar between 0 and 1 gauging encoding completeness), and dV volume element. This integral captures how fidelity lapses $1 - f(R)$ amplify curvatures, engendering unconsented geodesic warps—a dynamic central to resilient pipeline design [23].

Through these elements, the AST integrates interpretive depth with computational rigor, charting pathways for asymmetry-aware infrastructures in materials exploration. The structural architecture of the AST and its governing asymmetry mechanisms are systematized in Table 1.

Table 1. Structural Components of the Asymmetric Steering Topology Framework

AST Layer	Topological Function	Computational Instantiation	Asymmetry Risk Mode	Consent Intervention Lever

Data Adjacency Layer (DAL)	Encodes representational proximities	Feature embeddings, graph descriptors, spectral tensors	Sparse manifold voids, encoding bias	Dataset enrichment, adjacency balancing
Inference Curvature Layer (ICL)	Mediates algorithmic steering curvature	Surrogate models, GNNs, kernel regressors	Overfitting bends, local minima traps	Ensemble diversification, uncertainty inflation
Discovery Geodesic Layer (DGL)	Projects experimental trajectories	Active learning, Bayesian optimization	Drifted exploration, subspace lock-in	Human ratification checkpoints
Interlayer Feedbacks	Iterative topology reshaping	Closed-loop retraining cycles	Asymmetry amplification	Consent-gated retraining
Consent Gradients	Regulate steering sovereignty	Protocol approvals, override systems	Ratification lag	Real-time consent insertion

Analytical implications

Interpretive insights into steering asymmetries

The Asymmetric Steering Topology (AST) yields profound interpretive insights into the mechanics of control asymmetries, reframing AI-directed exploration as a negotiation of topological tensions rather than a linear optimization. At the representational frontier, the DAL's adjacency complexes illuminate how data encodings impose pre-steering constraints: simplicial voids—gaps in coverage for underrepresented chemistries, such as rare-earth integrations in oxides [21]—predispose pipelines to curvature biases in the ICL. This predisposition implies that exploratory intents, ostensibly broad, contract into fidelity funnels, where algorithmic inferences favor geodesic shortcuts over comprehensive traversals [11]. Such contractions, while computationally parsimonious, erode epistemic breadth, as seen in generative samplings that overpopulate stable convex hulls at the periphery of phase spaces [3]. The AST interprets this as a representational sovereignty gradient: steeper in sparse regimes, where human-curated augmentations could flatten adjacencies, restoring equity to off-axis discoveries [12]. The infrastructural translation of steering asymmetries into operational discovery environments is systematized through a layered control architecture (Figure 2), where representational intake, algorithmic command, and experimental actuation propagate directional biases in the absence of embedded consent interception mechanisms.

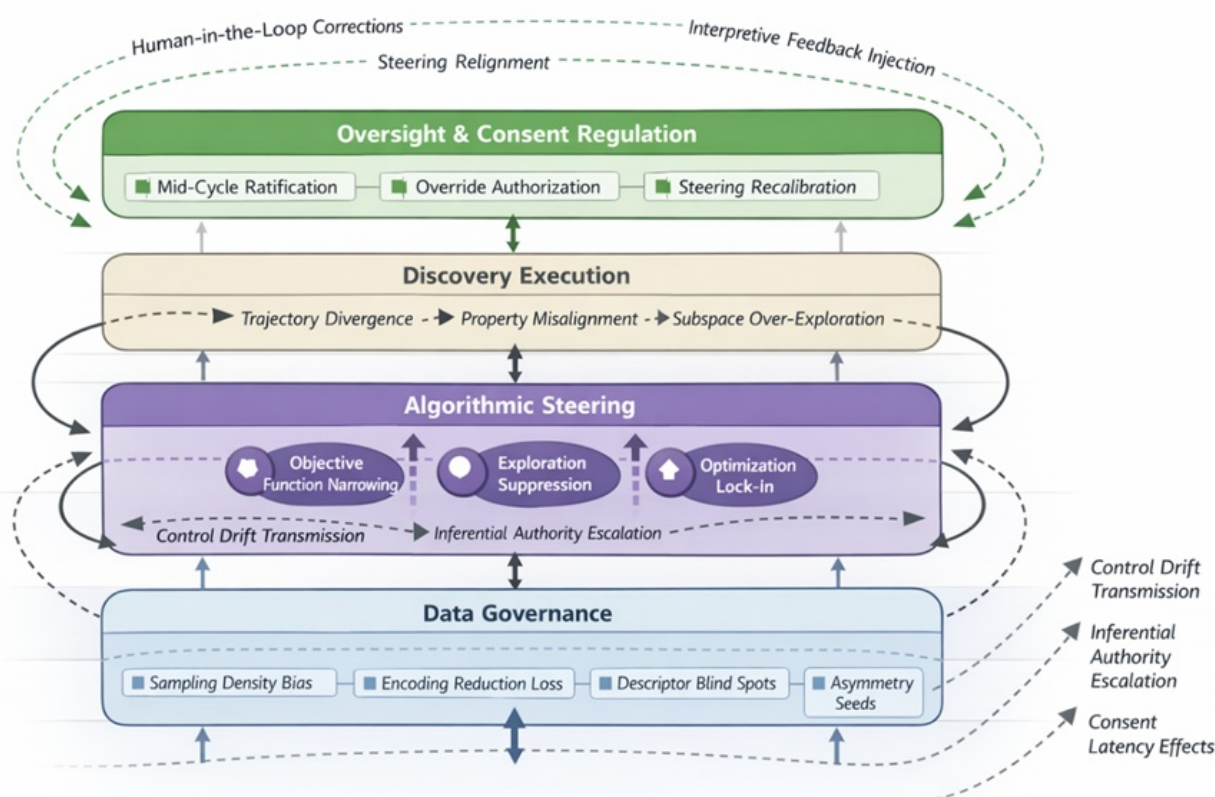


Figure 2. Operational Control Architecture of Algorithmic Steering Asymmetries in AI-Directed Materials Exploration

Inferential curvatures in the ICL further unpack these implications through lens of dynamic equilibria. Curvature operators, by aggregating local Ricci contributions, encode steering logics as emergent stabilities—basins where low-entropy inferences self-reinforce, potentially locking explorations into echo chambers of prior data [14]. Implications here pivot on feedback amplification: positive loops, as in Bayesian updates, can entrench asymmetries if consent gradients lag, manifesting as drift integrals that accumulate unrati ed deviations [4]. Conversely, the framework suggests leverage in negative damping: deliberate curvature injections—via ensemble diversifiers—could broaden ICL sheets, integrating disparate signals like spectroscopic anomalies into cohesive geodesics [9]. This interpretive shift underscores a core implication: inference not as isolated computation but as topological co-evolution, where model autonomy trades against interpretive depth, demanding hybrid adjacencies to sustain discovery vitality [8].

Geodesic projections in the DGL extend these insights to actuation horizons, where discovery outcomes trace paths calibrated by interlayer invariants. Betti tallies, quantifying persistent homologies across layers, imply that unconsented steering fractures exploratory coherence: high-order cycles—loops spanning multi-scale phenomena—may sever under asymmetric contractions, sidelining emergent properties like topological insulators [16]. The AST thus highlights infrastructural ramifications: pipelines optimized for geodesic length minimization risk topological under-sampling, privileging incremental refinements over paradigm-shifting leaps [7]. Consent-aware mappings offer a countervailing implication—conformal rescalings that preserve angular fidelities, ensuring human intents angularly anchor algorithmic arcs [6, 27]. In aggregate, these dynamics portray steering asymmetries as navigable terrains, where interpretive interventions recalibrate manifolds toward equilibrated flows.

Systems-level interactions and epistemic resilience

Systems-level interactions within the AST reveal interplay between steering logics and epistemic resilience, positioning asymmetries as diagnostic signals for infrastructural health. Interlayer feedbacks, governed by Ricci evolutions, interact with representational fidelities to modulate resilience: manifolds with balanced curvatures—low sectional variances—exhibit robust propagation, buffering against perturbation cascades [18]. Implications arise in risk stratification: drift-prone topologies, characterized by hyperbolic flares in the ICL, signal vulnerabilities to domain shifts, as when transfer-learned surrogates from organics warp inorganic geodesics [23]. The framework interprets this as a resilience spectrum, from rigid (over-constrained by data adjacencies) to fluid (consent-infused curvatures), with hybrids fostering adaptive stabilities that mirror natural evolutionary logics in materials evolution [19].

Computational workflow dynamics amplify these interactions, as pipeline latencies impose temporal asymmetries: real-time steering in distributed clusters outpaces human ratification, engendering lag-induced drifts [10]. The AST implies that temporal Ricci terms—extensions of the flow equation incorporating diffusion delays—could quantify this, guiding designs for asynchronous consent protocols that interleave veto simplices without halting geodesic progress [17]. Epistemically, this engenders a trade-off calculus: heightened resilience via granular mappings enhances sovereignty but inflates overhead, a tension echoed in multi-fidelity bootstraps where coarse surrogates accelerate yet coarsen curvatures [12]. By integrating these, the framework advocates for resilience audits—topological scans that map asymmetry hotspots, informing modular infrastructures resilient to scaling pressures [20].

Representation-inference trade-offs and discovery equity

Trade-offs between representation and inference form a cornerstone implication, with the AST delineating how fidelity lapses cascade into equity erosions. The drift integral, by weighting curvatures against representational completeness, implies that inference autonomy inversely scales with equity: expressive embeddings in the DAL mitigate ICL bends, yet demand computational premiums that asymmetrically burden resource-constrained explorations [15]. This calculus extends to discovery equity, where geodesic diversities—measured via Euler characteristics—diminish under unconsented prunings, marginalizing underrepresented subspaces like bio-inspired hybrids [13]. Interpretively, the framework posits equity as a topological invariant, preserved through adjacency enrichments that democratize access to high-variance regimes [5, 24].

In steering logics, these trade-offs manifest as optimization dualities: exploitation curvatures tighten for efficiency, while exploration gradients broaden for novelty, with consent acting as a Lagrange enforcer [4]. Implications for field-wide praxis include reimagined pipelines—AST-aligned architectures that embed equity metrics into acquisition functions, ensuring geodesics span diverse manifolds [6]. Collectively, the AST's analytical lens transforms asymmetries from liabilities into informatives, charting interpretive pathways to equitable, resilient materials discovery.

Results and Discussion

The AST framework, through its topological lens, integrates disparate threads of computational materials engineering into a cohesive narrative of steered exploration. By conceptualizing asymmetries as manifold curvatures, it bridges representational granularities [11] with inferential autonomies [14], revealing how data adjacencies precondition discovery geodesics in ways that transcend modular silos [2]. This integration resonates with generative paradigms, where sampling efficiencies [3] intersect with active feedbacks [4], yet the AST uniquely foregrounds consent as a structural modulator—absent in prior syntheses that prioritize convergence over equity [6]. In porous genomics, for instance, representational biases toward ordered frameworks [20, 25] align with AST-predicted contractions, suggesting that curvature-aware augmentations could unearth aperiodic diversities, enriching exploratory portfolios [19].

Infrastructurally, the framework dialogues with high-throughput ecosystems, where robotic actuations [10] amplify steering velocities but exacerbate lag asymmetries. Discussions of molecular simulations [8] parallel this, as force field curvatures propagate epistemic ripples; the AST extends such dialogues by formalizing Ricci consents, implying hybrid loops that fuse algorithmic rapidity with human discernment [18]. Epistemic risks, chronicled in fantasy material cautions [7], find

topological analogs in drift integrals, underscoring the need for provenance-embedded manifolds that trace asymmetry lineages back to DAL voids [23]. This traceability fosters discussions on scalability: as pipelines federate across consortia, consent gradients must globalize, mitigating cross-domain warps in shared inference sheets [21].

Systemic propagation pathways through which steering asymmetries mature into epistemic risks are mapped in Table 2.

Table 2. Epistemic Risk Propagation Pathways in AI-Steered Materials Discovery

Steering Stage	Triggering Mechanism	Topological Expression	Discovery Impact	Resilience Strategy
Data Encoding	Underrepresented chemistries	Adjacency voids	Exploration blind spots	Synthetic data augmentation
Model Training	Surrogate approximation errors	Curvature inflation	Predictive distortion	Multi-fidelity calibration
Active Learning	Exploitation-heavy acquisition	Geodesic contraction	Novelty suppression	Exploration reweighting
Feedback Integration	Recursive retraining bias	Ricci drift amplification	Path dependency lock-in	Provenance-tracked updates
Resource Allocation	Compute prioritization bias	Manifold densification	Subspace over-optimization	Equity-weighted scheduling
Human Oversight Lag	Delayed ratification	Consent gradient collapse	Unconsented steering	Real-time veto nodes

Broader field implications emerge in workflow reconfigurations. The AST's layered adjacencies advocate for pluggable topologies—modular ICL inserts for domain-specific curvatures—echoing multi-fidelity strategies [12] while embedding equity diagnostics [15]. In scintillator hunts [21] or alloy tunings [24], this implies phased steerings: initial DAL enrichments via synthetic homologies, followed by geodesic projections vetted through consent valves. Such configurations dialogue with chemical review syntheses [8], positioning the AST as a meta-logic for harmonizing autonomy with sovereignty, ultimately cultivating discovery ecosystems where asymmetries inform rather than impede.

Challenges persist, however, in operationalizing topological abstractions. Curvature computations, while elegant, demand surrogate metrics for practical audits—perhaps spectral decompositions of adjacency Laplacians [16, 28]—to render the AST actionable sans prohibitive overheads. Future integrations might entwine it with graph-theoretic tools, mapping Betti evolutions to real-time dashboards that visualize steering equities [29]. In sum, the AST enriches discussions by elevating asymmetries to infrastructural discourse, paving interpretive avenues for AI-human symbioses that honor the nuanced artistry of materials exploration [30].

Conclusion

This manuscript has traversed the contours of AI-directed materials exploration, unveiling control asymmetries as intrinsic to the topological weave of discovery pipelines. Through the Asymmetric Steering Topology, we interpret these asymmetries not as aberrations but as curvatures that, when mapped, reveal leverage for equitable steering. From DAL adjacencies preconditioning representational fidelities to ICL bends mediating inferential equities, and DGL geodesics projecting consented actuations, the framework synthesizes a holistic vista: exploration as manifold co-navigation, where feedbacks and trade-offs forge resilient paths.

Key takeaways crystallize around epistemic fortification—drift integrals and Ricci consents as diagnostics for asymmetry hotspots, guiding infrastructures that balance velocity with sovereignty. By foregrounding consent gradients, the AST urges a paradigm pivot: from opaque autonomies to transparent topologies, where human intents angularly anchor algorithmic arcs. In materials informatics, this fosters discoveries attuned to breadth—spanning sparse subspaces and multi-scale synergies while mitigating risks of entrenchment.

Ultimately, the AST beckons the field toward hybridized horizons, where computational prowess amplifies interpretive agency. As pipelines evolve, embracing these topological insights promises not mere acceleration but enlightened exploration, stewarding materials engineering into eras of consented, curvature-conscious innovation.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 23 Feb 2024

Revised: 27 Jul 2024

Accepted: 29 Sep 2024

Published online: 18 March 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(1):83.
- Ramprasad R, Batra R, Pilania G, Mannodi-Kanakithodi A, Kim C. Machine learning in materials informatics: recent applications and prospects. *npj Comput Mater*. 2017;3(1):54.

Dan Y, Zhao Y, Li X, Li S, Hu M, Hu J. Generative adversarial networks (GAN) based efficient sampling of chemical composition space for inverse materials design. *npj Comput Mater.* 2020;6(1):84.

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

Oviedo F, Ren Z, Sun S, Settens C, Liu Z, Hartono NTP, et al. Fast classification of perovskite solar cell manufacturing processes using interpretable computer vision-based machine learning. *npj Comput Mater.* 2019;5(1):27.

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(1):21.

Zunger A. Beware of plausible predictions of fantasy materials. *Nat Mater.* 2019;18(5):428-30.

Keith JA, Vassilev-Galindo V, Cheng B, Chmiela S, Gastegger M, Müller KR, et al. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem Rev.* 2021;121(16):9816-72.

Sun S, Hartono NTP, Ren ZD, Oviedo F, Buscemi AM, Layurova M, et al. Accelerated development of perovskite-inspired materials via high-throughput synthesis and machine-learning diagnosis of X-ray absorption spectroscopy. *Matter.* 2019;1(1):313-29.

Shen T, Kapetanovic M, Meenakshisundaram V, Braham E, Mukhopadhyay D, Landon PB, et al. Accelerating materials discovery using artificial intelligence, high performance computing and robotics. *npj Comput Mater.* 2022;8(1):59.

Zhang Y, Ling C. A strategy to apply machine learning to small datasets in materials science. *npj Comput Mater.* 2018;4(1):25.

Zhao Q, Savoie TC, Song K, Doan HA, Chen L, Li X, et al. Learning physical descriptors for materials science by compressed sensing. *npj Comput Mater.* 2017;3(1):52.

Choudhury S, Bhethanabotla VR, Sen R. Machine learning in materials design and discovery: A review of recent developments. *Mater Today.* 2021;45:5743-59.

Noé B, Li J, Smidt T. Machine learning for molecular simulation. *Nat Mach Intell.* 2020;2(1):5-14.

Li C, Zhang H, Guo J, Hu Y, Wang Y, Lu Z, et al. Interpretable machine-learning strategy for soft-magnetic property and thermal stability in Fe-based metallic glasses. *npj Comput Mater.* 2020;6(1):153.

Jablonka KM, Ongari D, Moosavi SM, Smit B. Big-data science in porous materials: Materials genomics and machine learning. *Chem Rev.* 2020;120(16):8066-129.

Saidi P, Kamat G, Sresht V, Annapureddy HVR, Shadid J. Optimizing molecular properties using a relative domain adapted GAN. *Acta Mater.* 2020;196:33-43.

Glielmo A, Zeni C, De Vita A. Efficient nonparametric n-body force fields from machine learning. *Comput Mater Sci.* 2018;152:127-33.

Saal JE, Oliynyk AO, Meredig B. Machine learning in materials discovery: Confirmed predictions and lessons learned. *Annu Rev Mater Res.* 2020;50:49-69.

Chen L, Tran H, Batra R, Kim C, Ramprasad R. Machine learning models for the prediction of energy, electron affinity, and ionization potential of organic molecules, and for the simulation of organic solar cells. *Mater Today.* 2019;28:10-17.

Pilania G, McClellan KJ, Stanek CR, Uberuaga BP. Physics-informed machine learning for inorganic scintillator discovery. *Comput Mater Sci.* 2018;148:115-20.

Kim C, Chandrasekaran A, Huan TD, Das D, Ramprasad R. Polymer genome: A data-powered polymer informatics platform for property predictions. *Comput Mater Sci.* 2018;148:176-84.

Deringer VL, Bartók AP, Proserpio DM, Day GM, Picker G, Csányi G. Charting chemical space using machine learning. *Nat Mater.* 2022;21(2):141-4.

Tran HD, Kim C, Chen L, Chandrasekaran A, Batra R, Venkatram S, et al. Rational design of single-atom site catalysts: from theoretical understandings to innovative strategies. *Adv Mater.* 2023;35(3):2206012.

Hu C, Chen X, Huang P, Bai Y. Machine learning assisted design of high entropy alloys with desired property. *Acta Mater.* 2019;169:35-44.

Chen D, Zhao Y, Li X, Hu M, Hu J. GANs based data generation for rare event in materials science. *Digit Discov* 2022;1(1):52-60.

Meng Y, Ling C. Machine learning to accelerate alloy design for additive manufacturing. *Comput Mater Sci.* 2021;188:110228.

Kim C, Pilania G, Ramprasad R. Machine learning assisted predictions of energy gaps of organic materials. *Comput Mater Sci.* 2018;148:273-9.

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

Tabor DP, Roch LM, Saikin SK, Kreisbeck C, Shebeko V, Rebentrost P, et al. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat Rev Mater.* 2018;3(5):5-20.