Abductive Discretization and Residual Politics: From Kantian Schematism to “Open Schema” AI Governance

Se Hoon Son *

Posted Date: 23 January 2026

doi: 10.20944/preprints202601.1719.v1

Keywords: AI governance; algorithmic auditing; abductive reasoning; residuals; German Idealism; categorization; schematism; phenomenology

Article

Abductive Discretization and Residual Politics: From Kantian Schematism to "Open Schema" AI Governance

Se Hoon Son

Man-o College of General Education, Busan University of Foreign Studies, Republic of Korea; undiner@bufs.ac.kr

Abstract

Fairness and minority exclusion have emerged as the central concerns of contemporary AI ethics. However, standard auditing practices often fail to capture harms affecting edge cases and marginalized groups. This article argues that this failure is structural: the act of "discretization"—converting continuous reality into discrete governance categories—inevitably produces a "residual." Drawing on German Idealism (Kant, Fichte, Schelling) and continental philosophy (Dilthey, Gadamer, Merleau-Ponty), we reconceptualize residuals not as mere noise but as "surprising facts" that should trigger abductive hypothesis revision. We critique current audit-by-checklist approaches as "rituals of verification" that obscure these residuals. This article makes three key contributions: (i) a structural diagnosis of residual production using systems theory and topology; (ii) a philosophical reconstruction of abductive revision as a hermeneutic necessity; and (iii) an institutional design proposal—specifically, the Residual Ledger and Category Revision Protocols—to operationalize "Open Schema" governance.

Keywords: AI governance; algorithmic auditing; abductive reasoning; residuals; German Idealism; categorization; schematism; phenomenology

1. Introduction

Fairness and minority exclusion have become the signature concerns of contemporary AI ethics and governance. The dominant institutional response has been to expand documentation and evaluation practices—ranging from model cards and audit checklists to post hoc explanations—within a discourse dominated by the languages of fairness, transparency, and accountability [1,2]. Yet, as governance infrastructures proliferate, a disturbing paradox becomes clearer: governance improves its capacity to certify what is already legible to predefined categories, while repeatedly failing precisely where political stakes are highest—in edge cases, ambiguous subjects, culturally non-standard expressions, and minority forms of life. In short, AI governance often grows more verifiable without becoming more responsive.

A key reason for this failure is that many audits function less as epistemic inquiry than as organizational ritual. They tend to convert uncertainty into a manageable administrative form—"evidence" that a process was followed—while allowing the most troubling remainder to disappear from view. Classic critiques of auditing describe this phenomenon as a "ritual of verification," producing trust through procedure rather than through substantive understanding [5,6]. When AI auditing becomes a matter of checklist compliance, what is not easily captured by the checklist is treated as peripheral. Yet those very leftovers are not peripheral to politics: they are the specific situations in which harms materialize and contestation becomes unavoidable.

This article names that remainder the residual and argues that it is structurally produced by discretization—the transformation of continuous, heterogeneous social reality into discrete categories that can be measured, predicted, and governed. In classification studies, it has long been

 doi:10.20944/preprints202601.1719.v1

observed that every classificatory system generates "residual categories" (miscellaneous, borderline, "other"), and that such residues are not accidental but constitutive of the system's operation [7]. In the context of AI governance, residuals appear as misfits between lived experience and imposed labels, as underrepresented groups whose patterns are treated as noise, and as ambiguous cases that resist stable categorization. Importantly, residuals are not simply "data problems" to be solved by collecting more samples; they are epistemic and political artifacts of how systems decide what counts as a category in the first place.

Box 1. Residuals: Definition, Boundary Conditions, and Governance Implications.

Definition (governance diagnosis). In this paper, a residual is not merely a leftover prediction error. It is a recurrent mismatch between (i) the lived situation in which harms and claims arise and (ii) the institutionally authorized categories, thresholds, and evidentiary rules used to render those situations governable. Residuals therefore indicate a problem not only of accuracy but of category-formation, admissible evidence, and responsibility boundaries.

Boundary conditions (what residuals are NOT). Residuals may co-occur with outliers, distribution shift, or out-of-distribution (OOD) inputs, but they are conceptually distinct:

(1) Outliers and drift are primarily statistical diagnoses relative to a dataset or model behavior.
(2) OOD flags epistemic distance from training support.
(3) Residuals are governance failures: the schema cannot stably recognize a case as a case, cannot justify the boundary that excludes it, or cannot translate the harm into admissible evidence without erasing its meaning.

Typical residual modes (illustrative). Residuals tend to appear as:

- Category mismatch: the taxonomy lacks a distinction that the situation demands.
- Threshold ambiguity: small contextual shifts flip the outcome, relocating responsibility.
- Multi-label conflict: legitimate aspects of the case map to incompatible labels or procedures.
- Cultural/context mismatch: the schema presupposes a "standard" form of life that does not hold.
- Rights/harms not captured: the harm is real but cannot be expressed in the approved fields.

Governance implication. Treat residuals as first-order governance inputs. The question is not "How do we eliminate residuals?" but "How do we make residual patterns visible, actionable, and revision-triggering without collapsing them into exceptions?"

Current governance discourse tends to address residual harms by extending auditing coverage. Recent work on internal algorithmic auditing and governance artifacts emphasizes end-to-end documentation across the AI lifecycle and the production of inspectable audit materials [3,4]. While such work is valuable, it also reveals a fundamental limit: procedural completeness can coexist with conceptual rigidity. A system can be "fully audited" relative to its existing categories while remaining blind to what those categories systematically exclude. The missing philosophical layer concerns how categories are formed, stabilized, and—crucially—revised when confronted with what they cannot absorb.

To develop that layer, we turn to a philosophical tradition that treats conceptual formation not as mere labeling but as a dynamic relation between schema and resistance. Kant's schematism frames concepts as requiring mediating rules [9]. Fichte radicalizes the problem by foregrounding *Anstoß*—a "check" that interrupts self-positing activity [10]. Schelling interprets the remainder as an internal motor of development [11]. Dilthey treats meaning as historically formed within life-relations [12]. These trajectories anticipate Peirce's logic of discovery, where abduction arises from the "surprising fact" [15]. Across these figures, resistance is not merely an error signal; it is the condition under which conceptual life evolves. The remainder of the paper proceeds as follows. Section 2 clarifies residuals as a political artifact of classification, employing René Thom's catastrophe theory as a heuristic.

 doi:10.20944/preprints202601.1719.v1

Section 3 reconstructs abductive discretization through a philosophical lineage from Kant to phenomenology. Section 4 critiques contemporary XAI through Dilthey's distinction between explanation and understanding. Section 5 proposes open schema governance as a design orientation. Section 6 concludes with implications for future empirical validation.

2. Residuals and Audit Rituals: The Politics of Classification and Verification

2.1. Residuals Are Not Accidents: The Constitutive Remainder of Discretization

AI governance often treats failures as deviations from an otherwise stable classificatory order. Yet residuals are not merely leftover errors; they are constitutive remainders generated whenever continuous, heterogeneous reality is discretized into governance-ready categories. The very act of drawing boundaries—between safe and unsafe, compliant and non-compliant—creates zones of indistinction. These zones are where power appears as the authority to decide what counts as a case. In this sense, residuals are not "exceptions to be fixed" after the fact; they are the predictable byproduct of abstraction and boundary-making [28].

As a formal heuristic (rather than a strict mathematical model), René Thom's catastrophe theory helps clarify why discretization yields stable forms alongside predictable breakdown zones. Thom defines "form" in terms of structural stability—patterns that remain qualitatively invariant under perturbation [39] (pp. 6–7)—and analyzes how continuous variation can pass through singular points where stability collapses [39] (pp. 38–40). Applied to governance, the analogy is straightforward: discretization aims to stabilize social variability into administrable "forms" (categories), yet certain cases concentrate precisely where the stabilization breaks. We do not claim an isomorphism between social life and dynamical systems; the point is that any attempt to enforce stable categorical forms upon heterogeneous continuities predictably generates rupture points. Topological heuristic (governance translation). Here "topology" is not invoked as a strict mathematical apparatus but as a vocabulary for reasoning about continuity, boundary-making, and singular points in governance categorization. Discretization seeks stable "forms" (administrable categories), yet it necessarily generates boundary regions where cases remain underdetermined. These regions function like singularities in the sense that small perturbations in context or wording can flip the classification outcome and thereby relocate responsibility. Open schema governance treats such boundary regions as sites where institutions must install explicit revision triggers, rather than as anomalies to be absorbed by discretionary exception-handling.

This rupture is not just mathematical but philosophical. In Schelling's terms, the residual is the "indivisible remainder" that the system's productivity cannot fully exhaust [11]. In Luhmann's terms, it is the inevitable "blind spot" generated by the system's need to reduce environmental complexity [8]. Residuals thus mark not only epistemic limits (what the schema cannot yet capture) but also political limits (what institutions cannot easily acknowledge without revising their accountability structure).

2.2. Audit as Ritual: When Verification Replaces Responsiveness

If discretization produces residuals, auditing often stabilizes them. Audits are built to produce legibility, comparability, and organizational coordination [5]. The problem is not that auditing exists, but that auditing can become ritualized: procedure substitutes for responsiveness. Strathern's "audit cultures" analysis highlights how accountability regimes can reorganize practices around what is measurable, producing performative compliance while displacing substantive judgment [6].

We therefore propose a diagnostic distinction between two layers of governance evidence. The first is a narrative layer: testimonies, complaints, and the phenomenology of harm. The second is a document layer: standardized vocabularies that convert events into reportable units. The problem is that institutions increasingly treat the second as sufficient, allowing the first to be compressed away. When the document layer becomes the only admissible reality, residuals are administratively "handled" while remaining conceptually unaddressed.

 doi:10.20944/preprints202601.1719.v1

2.3. The Residual Ledger: Making the Remainder Governable without Erasing It

If residuals are constitutive, governance must treat them as first-class objects rather than as embarrassing leftovers. We therefore introduce the **Residual Ledger**: a mandatory documentation module in which the system explicitly records what its categories cannot absorb. Unlike "miscellaneous" fields that silently flatten anomalies, a residual ledger is designed to expose patterns of exclusion and mismatch.

Concretely, the ledger should (i) taxonomize residual types (e.g., semantic ambiguity, borderline cases, cultural/context mismatch), (ii) quantify residual scale, and (iii) map residual concentration [4,7]. These elements transform residuals from "unstructured complaints" into governance-relevant signals that can trigger institutional learning and schema revision.

A second design implication follows: governance documents should not mix narrative and compliance evidence in a single column. They should present the two layers in parallel—one preserving the continuity of lived harm, the other representing the standardized abstraction. Such dual presentation prevents the document layer from silently overwriting the narrative layer and makes the act of schematization itself auditable.

2.4. Thresholds, Accountability, and Abductive Revision

Residual politics becomes most visible at thresholds. In technical terms, a threshold parameter separates acceptance from rejection. In governance terms, this same threshold functions as a responsibility boundary: it allocates who will be recognized as a subject of concern and who will be routed into residual categories. This is why thresholds must be justified not only mathematically but also publicly—through documented rationales that treat parameter choices as normative commitments rather than as inevitable facts [25–28].

Finally, if residuals are treated as governance signals, then governance must include a formal rule for abductive revision: when residual scale or concentration crosses a declared threshold, the category system itself must be revised—by introducing new categories, splitting existing ones, redefining criteria of evidence, or altering decision boundaries. This is the core of the open schema orientation. A schema is "open" not because it is vague, but because it contains explicit protocols for self-modification in response to the resistance of the real. Section 3 will reconstruct why such residual-driven revision is not merely a managerial preference but a philosophical necessity. Concretely, this requires specifying (a) who is authorized to initiate revision (a governance body with defined membership), (b) what evidence counts as revision-relevant (ledger trends, appeal overturn rates, narrative corroboration), and (c) how revisions are implemented (versioned taxonomies with changelogs and publicly stated rationales). Without these due-process constraints, "revision" remains an aspiration and residuals revert to discretionary handling.

3. German Idealism and Abduction: Schema, Resistance, and Conceptual Development

3.1. Kantian Schematism and the Structural Underdetermination of Categories

The problem of the residual finds its philosophical root in the gap between general rules and specific cases. Immanuel Kant's schematism clarifies that categories do not apply themselves; concepts require mediating rules (schemata) to connect with intuition [9]. Yet, as Niklas Luhmann later radicalized, this mediation is never exhaustive. Systems must reduce environmental complexity to remain operable [8,14]. Consequently, **Edmund Husserl** warns against the "mathematization of nature," arguing that formal systems risk forgetting the *lifeworld* (*Lebenswelt*)—the pre-theoretical horizon of meaning that sustains the system itself [13]. The residual is thus the trace of the lifeworld that the system's reduction was forced to exclude.

This diagnosis is deepened by a constellation of thinkers who expose the fragility of rigid categorization. Ludwig Wittgenstein's concept of "family resemblances" challenges the assumption

 doi:10.20944/preprints202601.1719.v1

that concepts have sharp boundaries [40]. Elena Esposito adds a temporal dimension, highlighting the "asynchrony of translation" between algorithmic and social systems [41]. Furthermore, Hans-Georg Gadamer and Maurice Merleau-Ponty ground the residual in the conditions of understanding and perception: Gadamer argues that "misunderstanding" is not a bug but a condition of hermeneutic engagement [42], while Merleau-Ponty emphasizes that every visible figure relies on an invisible background ("ground") that cannot be fully objectified [43].

From these converged perspectives, residuals appear as systematic gaps between (i) the richness of lived meaning (Lifeworld/Ground) and (ii) the operative simplifications (Schema/Form) required for institutional decision-making.

3.2. Fichte's Anstoß: Residuals as the "Check" that Forces Reconfiguration

Fichte radicalizes schematism by making resistance not a mere empirical inconvenience but a transcendental condition for agency and cognition. The *Anstoß*—often glossed as a "check" or "impulse"—interrupts self-positing activity and compels a reconfiguration of the I's practical-cognitive orientation [10]. What matters here is not the metaphysical status of the *Anstoß* but its functional role: it marks the point where the system encounters something it cannot assimilate through its current rule-set.

Within AI governance, residuals can be interpreted as institutional *Anstöße*. They are the moments when the system's categories fail to do their job—when an audit produces a formally complete record that nonetheless cannot settle a dispute, repair harm, or justify a threshold decision. Residuals therefore do not merely indicate uncertainty; they indicate a normative interruption. The apparatus is forced to ask whether its categories remain legitimate. This shift matters because it relocates the "problem" from the data to the schema.

This Fichtean interruption is structurally identical to what Peirce called the "surprising fact" that breaks habitual expectation and triggers inquiry [15]. Abduction, for Peirce, is the logic by which thought moves when deduction and induction cannot absorb what appears [15]. Contemporary work on abductive cognition reinforces this point: abductive reasoning is the capacity to generate and revise hypotheses under constraint, especially when locked schemas fail to accommodate novelty. In governance terms, residual concentration should be treated as precisely such a surprising fact—an institutional signal that category revision is required rather than deferred.

3.3. Schelling: The Dynamic Remainder as a Motor of Formation

Schelling's contribution is to interpret remainder and opacity not as external obstacles but as internal conditions of formation. In his philosophy of nature, productivity is not exhausted by its products; it exceeds them [11]. The "negative" is not merely a lack but an active dimension of becoming. Transposed into our context, the residual is not a defect to be eliminated but a dynamic surplus generated by the living encounter between schema and world.

This is why an audit-only approach is philosophically impoverished: it presupposes that the right category system is already given, and that the task is merely to enforce compliance. A Schellingian view treats governance as an ongoing formation process: categories are historically situated and must remain responsive to what exceeds them. Residuals become the sites where governance learns—provided that institutions do not treat them as reputational threats to be suppressed.

3.4. Dilthey: The Epistemological Rift Between Explanation and Understanding

Wilhelm Dilthey provides the decisive epistemological ground for distinguishing governance from mere engineering. In *Introduction to the Human Sciences*, he famously demarcates the "explanation" (*Erklären*) of the natural sciences from the "understanding" (*Verstehen*) of the human sciences [12]. Interpretive understanding cannot be replaced by formal explanation alone, because what counts as relevant context is itself historically variable. In AI governance, this implies that

 doi:10.20944/preprints202601.1719.v1

attempts to fix categories once and for all—especially across cultures—will inevitably generate residuals that are not noise but traces of unacknowledged context. Taken together, Kant, Fichte, Schelling, and Dilthey allow us to define abductive discretization not merely as a technique but as a normative stance. Because discretization necessarily produces residuals (Kant), and because these residuals act as checks that force reconfiguration (Fichte), drive formation (Schelling), and remain historically situated (Dilthey), governance cannot rely solely on deduction (applying rules). It must institutionalize abduction—the capacity to infer new rules from the surprise of the residual.

This distinction is critical for criticizing contemporary AI "explainability." Most XAI (Explainable AI) tools offer only *Erklären*: they trace causal pathways or feature importances within a frozen model. However, they fail at *Verstehen*: they cannot grasp the "nexus of life" (*Lebenszusammenhang*) from which the data was extracted.

For Dilthey, meaning arises only within this historical nexus. When an AI system discretizes reality, it breaks this nexus, treating lived moments as atomic data points. The "residual," then, is the trace of the life-nexus that resists being reduced to a causal variable. Therefore, "abductive redescription" in governance is not merely technical error correction; it is a hermeneutic act. It restores the context that natural-scientific explanation necessarily stripped away, re-integrating the residual into a meaningful whole.

4. From Audit to Abduction: Why Explanation and Interpretability Do Not Eliminate Residuals

4.1. Why "More Explanation" Does Not Eliminate Residuals

Contemporary AI governance often responds to residual harms by demanding "more explanation." Explanations are expected to make decisions transparent, contestable, and accountable [23,24]. Yet the demand for explanation can inherit the same limitation as audit-by-checklist: it treats the existing category system as given. In that case, an explanation clarifies why the system classified a case under a category, but it does not ask whether the category system itself is adequate for the lived situation that generated harm.

This is a structural point about discretization. Explanation typically functions deductively: it offers a justification within the system's vocabulary, thereby converting external disagreement into internal coherence. Here, the epistemological limit identified by Dilthey becomes a governance critique: contemporary XAI tools provide *Erklären* (causal tracing of weights and features) but structurally lack *Verstehen* (the hermeneutic grasp of the life-context). Because the system has already discretized the "nexus of life" (*Lebenszusammenhang*) into variables, its explanation can only trace the path of variables, not the meaning of the nexus [12]. Residuals remain—especially where the mismatch is not between input and output but between life and schema.

4.2. Internal Auditing and Governance Artifacts: When Documentation Becomes the End

Governance artifacts are built for legibility. They summarize complex socio-technical events into reportable units: risk levels, mitigation steps, accountability assignments, compliance status. This is not a flaw; it is the condition of organizational coordination. The flaw arises when legibility is mistaken for completeness. When a template dictates what can be recorded, what cannot be recorded becomes residual.

Recent work on internal algorithmic auditing emphasizes end-to-end documentation and the production of auditable artifacts [4]. Model cards and related documentation aim to communicate intended use, limitations, and performance characteristics to relevant stakeholders [3]. These initiatives are crucial, but residual politics introduces a further requirement: governance artifacts must contain an abductive pathway. Without a formal procedure that converts residual concentration into category revision, auditing risks becoming a stabilizer of the status quo—verifying correct application of existing categories while routing misfits into exceptions, unknowns, or discretionary handling.

 doi:10.20944/preprints202601.1719.v1

4.3. Interpretability vs. Explanation: Two Governance Promises, One Residual Limit

Interpretability and explanation are often conflated. Yet interpretability aims at making model reasoning understandable (sometimes by design), whereas explanation often provides post hoc rationales for decisions [36]. The literature reflects this variety: calls for a more rigorous science of interpretability [20], local post hoc explainers such as LIME [21], and attribution methods such as SHAP [22]. Governance can benefit from these tools, but residuals reveal a shared limit: these methods typically operate within the existing schema.

This suggests that governance should treat explanation and interpretability as necessary but insufficient conditions. They can help contest decisions within a schema, but they do not guarantee that the schema itself will evolve. In our terms, explanation and interpretability operate largely in the document layer, whereas residual politics originates in the tension between document layer and narrative layer. What is needed is a bridge: an institutional mechanism that converts narrative residuals into document revisions.

4.4. Abductive Governance as a Protocol: From Residual Evidence to Category Change

If residuals are governance signals, governance needs a protocol that specifies how signals become change. We propose three minimal steps. First, residual evidence must be made durable: recorded in a residual ledger with types, frequency, trend, and concentration [7]. Second, residual evidence must be made contestable: affected parties must have institutional access to challenge categorizations and propose alternative hypotheses, in line with procedural accountability commitments [35]. Third, residual evidence must be made actionable: when declared thresholds are crossed, the governance body must initiate a category revision procedure, including versioned updates and public rationales [25–27].

This protocol makes explicit what audit and explanation often leave implicit: categories are provisional and must be governable as such. In other words, abductive governance is not a mood of openness; it is a structural constraint that prevents residuals from being treated as reputational hazards. Checklists can be valuable safeguards, but they are structurally biased toward closure: they reward what can be ticked off [5,6]. Residual politics begins where ticking off fails. Abductive governance therefore requires that governance institutions treat failure-to-tick as a first-order input into category revision—an opportunity for abduction—rather than as an embarrassment to be managed away.

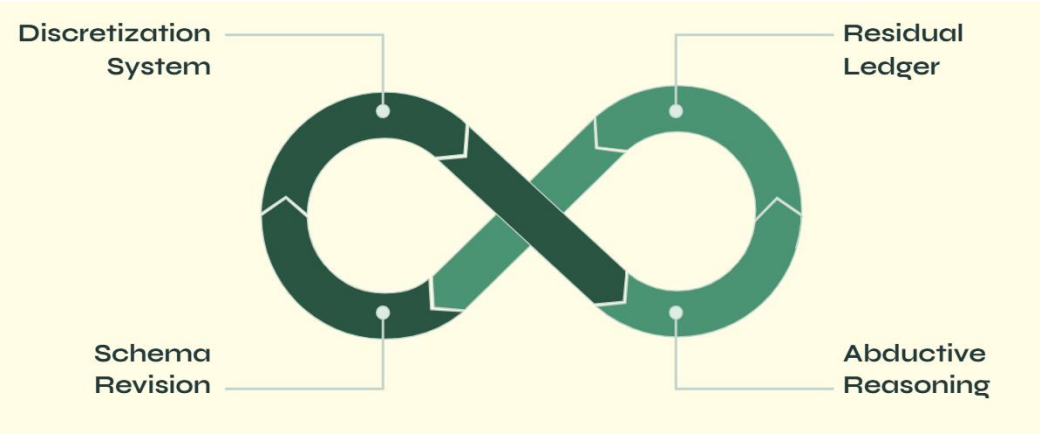

Figure 1. The Open Schema Governance Loop. This continuous feedback mechanism illustrates how the governance system evolves. It begins with the Discretization System filtering reality. Unresolved cases are logged in the Residual Ledger, triggering Abductive Reasoning when thresholds are crossed. This leads to Schema Revision, updating the system to better accommodate complex realities.

5. Open Schema Governance: Institutional Design for Residual Visibility and Revision

 doi:10.20944/preprints202601.1719.v1

To clarify what is new in an "open schema" orientation, Table 1 contrasts checklist-centered governance with open schema governance at the level of epistemic aim, admissible evidence, and institutional change. The core distinction is that open schema governance not only verifies compliance within categories but also authorizes revision of the categories themselves under explicit conditions.

Table 1. Checklist governance vs. Open schema governance (conceptual contrast).

Dimension	Checklist-centered Governance	Open schema governance
Primary aim	Verifiability of procedure; compliance demonstration	Responsiveness under inevitable discretization; revision capacity
Unit of governance	Fixed categories + predefined checklist items	Versioned taxonomy + explicit revision triggers
What counts as evidence	Standardized, document-layer fields; audit artifacts	Dual-layer evidence (narrative + document) linked but not collapsed
Default treatment of edge cases	Exception-handling; "miscellaneous/other"; discretionary overrides	Residual logging as first-order input; patterns become revision-relevant
Failure signal	Non-compliance (a missed item)	Residual concentration/scale + appeal overturn + threshold instability
Accountability style	Accountability-by-proof (show the checklist was followed)	Accountability-by-revisability (show how/when schema changes)
Change mechanism	Rare, informal, reactive Updates	Formal protocol: docket → review → decision → changelog → monitoring
Typical blind spot	Harms that do not fit approved fields; minority lifeworld contexts	Strategic flooding / over-revision (addressed by safeguards)
Outputs	Model cards, checklists, post hoc explanations	Residual ledger, trigger dashboard, revision decisions, versioned standards

5.1. The Residual Ledger as a Governance Primitive

An open schema orientation begins by treating residuals as governance primitives. We therefore propose the Residual Ledger as a mandatory module within AI governance documentation. Its function is not to catalogue anomalies as curiosities, but to record the systematic remainder that the current taxonomy and thresholds cannot absorb. In this sense, the ledger is not merely "extra transparency"; it is a structural correction to the bias of checklist governance toward closure.

At minimum, the ledger should include four fields: (i) Residual Type (unknown; borderline; multi-label conflict; cultural/context mismatch; threshold ambiguity; rights/harms not captured), (ii) Residual Scale (counts, rates, trends), (iii) Residual Concentration (distribution across groups, contexts, environments, with privacy-preserving aggregation where required), and (iv) Disposition (manual override; appeal; deferral; reclassification; policy exception), with rationales and responsible roles.

These fields are not merely administrative details; they operationalize the philosophical core of this paper. They treat the residual precisely as a Fichtean *Anstoß*—a check that interrupts the system's complacency and forces self-reflection [10]—and as a Peircean "surprising fact" that demands the formation of a new hypothesis [15]. Instead of treating anomalies as noise to be discarded or hidden, the ledger institutionalizes them as essential signals for abductive reasoning, ensuring that the system remains open to the resistance of reality.

Minimal measurement spec. Residual Scale can be operationalized as a rolling-window rate (e.g., residuals per 1,000 decisions) and a trend slope over time. Residual Concentration can be reported as privacy-preserving aggregates across protected groups and contexts, using simple inequality indicators (e.g., max-to-median ratio, entropy of distribution, or a Gini-like index over group-context

 doi:10.20944/preprints202601.1719.v1

cells). Disposition should record not only the outcome (override/appeal/deferral) but also the responsible role and justification category, enabling an "overturn-rate" signal that directly feeds revision triggers.

5.2. Dual-Layer Evidence: Preserving Narrative without Sacrificing Standardization

Because residual politics arises at the tension between lived harm and standardized documentation, governance must preserve both layers. We propose a dual-layer evidence format in which each report contains (A) a narrative layer describing the phenomenology of harm and context and (B) a document layer containing standardized fields required for auditability. The two layers should be linked but not collapsed.

This design counters a predictable failure mode: narrative becomes "unstructured" and is discarded as non-evidence, while the document layer becomes the only admissible reality. Dual-layer formats make schematization itself visible. They allow a reviewer to see not only what category was assigned but what was lost in the assignment. This design is structurally an attempt to preserve what Husserl called the *lifeworld* (*Lebenswelt*)—the pre-theoretical horizon of meaning that exists before and beneath formalization [13]. While metrics capture the discretized form, the narrative layer retains the trace of the lifeworld. Ethnographic accounts repeatedly show that harms become visible first as stories rather than as metrics [33,34]; thus, preserving this layer is a necessary condition for just governance.

5.3. Threshold Justification as Public Reason: Parameters as Normative Commitments

Open schema governance requires that threshold choices be treated as normative commitments. Technical systems often present thresholds as calibration details—confidence cut-offs, decision boundaries, acceptable error rates. But thresholds are also political boundaries: they define who is recognized as a valid case and who is relegated to residual status. We propose that every high-stakes threshold be accompanied by a public justification record that states intended trade-offs, affected populations, and the rationale for choosing one boundary over another [25–27].

Linking threshold justification to the residual ledger closes the loop: when residual concentration rises beyond a declared level, the threshold and/or taxonomy must be revisited rather than defended as technically "optimal." This is especially important because fairness criteria and error trade-offs are plural and contested [28–31,37,38].

5.4. Category Revision Protocol: Versioned Taxonomies and Explicit Triggers

A governance framework is "open" only if it contains explicit rules for self-modification. We propose a Category Revision Protocol with three components: (i) a versioned taxonomy (categories treated like evolving standards with changelogs), (ii) revision triggers (pre-declared conditions requiring revision, e.g., residual concentration for a protected group exceeds X; appeal overturn rates exceed Y), and (iii) review procedures (a documented process for proposing, evaluating, and approving changes).

The protocol's guiding premise is abductive: it treats residuals not as noise but as hypothesis prompts. A concentration of residuals suggests that the existing schema is failing to capture a stable pattern of life. Revision thus involves hypothesis generation ("What is the missing distinction?"), empirical probing, and normative deliberation. This is precisely what checklist governance lacks: it verifies application but does not authorize schema change [5–7]

5.4.1. Governance Body and Due Process.

Revision authority should be assigned to a defined governance body (e.g., an internal review board with external and/or affected-party representation where feasible). The board's remit is not to relitigate individual cases but to determine whether residual patterns indicate a schema mismatch that warrants revision. Due process minimally includes: (i) a docketed proposal that cites ledger

 doi:10.20944/preprints202601.1719.v1

trends and representative narratives; (ii) an impact assessment on affected groups, operational workload, and downstream accountability boundaries; (iii) a decision record with stated reasons and recorded dissent where applicable; and (iv) publication of the change as a versioned update with a short changelog. Without these due-process constraints, "revision" collapses into ad hoc discretion.

5.4.2. Safeguards Against Gaming and Over-Revision.

Open schemas are vulnerable to strategic "residual flooding," performative complaint cycles, or oscillatory over-correction. To mitigate this, protocols should include triage rules (severity × recurrence), sampling and corroboration requirements for narrative reports, and stability constraints (e.g., cooling-off periods, staged pilots, or sunset clauses for newly introduced categories). In addition, revision triggers should be paired with a burden-of-revision principle: changes must specify what evidence will count as success or failure in the next review window. The aim is to preserve openness to revision without turning revision into an unbounded channel for reputational management or bureaucratic churn.

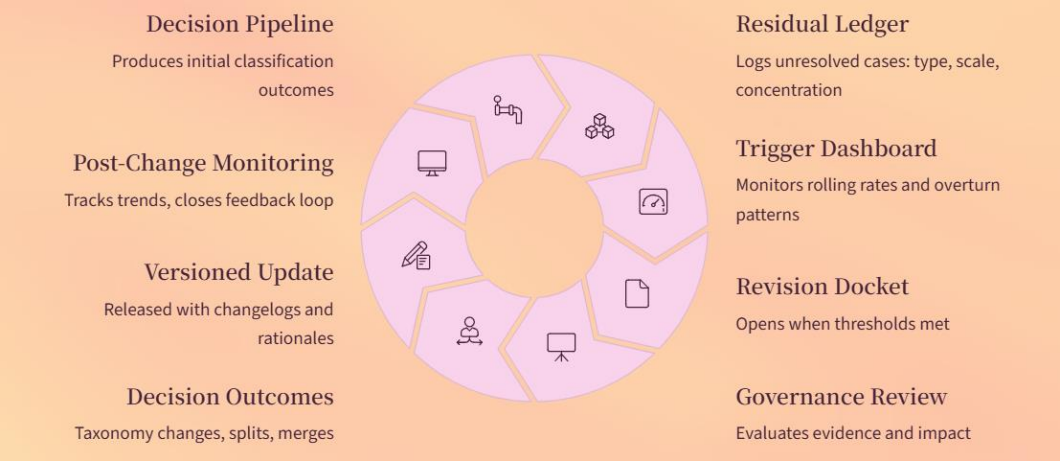

Figure 2. Category Revision Protocol (operational flow). The protocol translates residual visibility into authorized schema change under due process: **(1)** Decision pipeline produces outcomes; **(2)** unresolved/contested cases are logged into the Residual Ledger with type/scale/concentration/disposition; **(3)** aggregation generates a trigger dashboard (rolling rates, concentration, overturn); **(4)** when pre-declared triggers are met, a revision docket is opened; **(5)** the governance board reviews ledger evidence + representative narratives + impact assessment; **(6)** decision outcomes include taxonomy changes (add/split/merge categories), threshold updates, or evidentiary rule updates; **(7)** changes are released as a versioned update with changelog and rationale; **(8)** monitoring evaluates post-change residual trends and unintended shifts, closing the loop.

5.5. Minimal Implementation: How to Adopt Open Schema Governance

A common objection is feasibility. Institutions are busy, and governance capacity is limited. Open schema governance should therefore start with pragmatic integration: (i) add a residual ledger to existing audit templates; (ii) require dual-layer evidence for a subset of high-stakes decisions; (iii) pilot one revision trigger tied to residual concentration; (iv) publish short changelogs for taxonomy updates. These steps keep systems auditable while creating a formal pathway for abductive revision.

Importantly, this approach avoids a false alternative between "pure philosophy" and "full technical redesign." It is a governance design proposal grounded in a philosophical diagnosis of discretization, aligned with emerging risk governance frameworks and standards [25–27]. The institutional claim is simple: residual visibility and revision pathways are prerequisites for just governance under inevitable discretization. These pilots also generalize beyond purely informational systems, including physical AI settings where residuals manifest as action-level ambiguity rather than label mismatch.

 doi:10.20944/preprints202601.1719.v1

5.6. Worked Example: From Residual Ledger to Schema Revision (Illustrative)

Philosophical status of the example. The following case is not offered as an empirical report about a particular institution. It functions as a conceptual demonstration (in the spirit of a thought experiment) meant to make visible what the argument has claimed in abstract terms: that discretization generates a normative remainder, and that this remainder is epistemically and politically significant. The point is not the administrative details of welfare provision, but the structure of misrecognition produced when a schema of admissible evidence fails to capture a lived situation. The worked example therefore illustrates how "residuals" mark breakdowns in justification and responsibility-boundaries, and why an abductive orientation requires institutionalized revision rather than discretionary exception-handling.

All numerical thresholds and rates below are purely illustrative, used only to show how a revision trigger could be operationalized under due process. To demonstrate how an open schema orientation differs from checklist governance in practice, this section provides an illustrative end-to-end example: (i) a residual pattern emerges, (ii) the pattern is logged in a residual ledger, (iii) a pre-declared trigger is met, (iv) a governance body opens a revision docket and issues a versioned taxonomy update, and (v) post-change monitoring evaluates whether the residual pattern declines without shifting harm elsewhere.

Scenario. Consider a municipal service chatbot used for welfare eligibility guidance. The operational schema classifies user requests into administrative categories such as "Eligibility—Income Threshold," "Documentation—Missing Forms," and "Appeal Procedure." The system is compliant with a standard checklist: it provides disclosure, logs interactions, and uses approved templates. However, a recurring class of users—recently separated migrants with irregular employment and mixed household arrangements—report that the chatbot's responses repeatedly direct them to "missing documentation" rather than recognizing that the relevant issue is the schema's presupposition of stable household categories and standard wage records. The harm is not merely an incorrect answer; it is a patterned exclusion of a lived situation that does not map cleanly to the authorized categories.

Residual Ledger entry (example). The organization logs these cases using the residual ledger fields described in Section 5.1. A simplified ledger excerpt is shown below:

Residual Type: Category mismatch + threshold ambiguity

Short description: Mixed-household/irregular-income cases repeatedly mapped to "Missing Forms," obscuring eligibility-relevant evidence.

Residual Scale (rolling window): 27 residuals / 1,000 decisions (30-day window), up from 9 / 1,000 the previous month.

Residual Concentration (privacy-preserving aggregate): Concentrated in "migrant status = recent" × "employment = irregular" contexts; max-to-median ratio across group–context cells = 3.6.

Disposition: 61% escalated to human agent; 24% user appeal; 15% drop-off.

Overturn rate (appeal outcomes): 42% of escalated cases result in eligibility-relevant guidance that contradicts the system's initial categorization.

Narrative layer link: Representative narratives indicate that "missing forms" framing is experienced as misrecognition and triggers disengagement.

Trigger calculation and docket opening. Open schema governance pre-declares revision triggers to prevent ad hoc discretion. In this example, the organization uses three triggers: (T1) residual rate exceeds 20 / 1,000 for two consecutive rolling windows; (T2) concentration index exceeds 3.0 for a protected-group × context cell; (T3) appeal overturn rate exceeds 30% for a residual-linked pathway. Here, T1–T3 are satisfied. A revision docket is opened with: (i) ledger statistics and trends, (ii) a small set of representative narratives (anonymized and corroborated), and (iii) an impact assessment that estimates operational burden (agent escalations, user drop-off) and accountability risk (systematic misrecognition of a vulnerable group).

Governance review and schema revision. The governance board (Section 5.4.1) evaluates whether the residual is a "data problem" solvable by more training or a "schema problem" requiring

 doi:10.20944/preprints202601.1719.v1

taxonomy revision. The board concludes that the primary issue is schema mismatch: the taxonomy lacks a category that admits alternative evidence forms for irregular household and income situations, and the threshold logic for "documentation completeness" wrongly functions as a gatekeeping proxy for legitimacy. The board issues a versioned update:

Taxonomy change (Version 1.2 → 1.3):

(1) Add new category: "Eligibility—Nonstandard Household/Income Evidence" (with admissible evidence rules and escalation guidance).
(2) Split "Documentation—Missing Forms" into (a) "Missing Forms—Standard Case" and (b) "Evidence Not Expressible in Standard Forms" (residual-sensitive pathway).
(3) Update threshold rule: if the system detects irregular-income indicators and mixed household references, it must not default to "Missing Forms"; it must route to the new category and present an evidence menu (acceptable substitutes and a human-assistance option).

Changelog and rationale. The change is published as a short changelog: "v1.3 introduces a nonstandard-evidence eligibility category to reduce systematic misrecognition of irregular household/income cases. Trigger basis: residual rate > 20/1,000 (two windows), concentration index > 3.0, overturn rate > 30%. Expected effect: reduce escalations and drop-offs while improving rights-relevant guidance."

Post-change monitoring. Open schema governance closes the loop by monitoring both intended improvements and potential harm-shifting. Over the next 60 days, the organization tracks: (i) residual rate for the relevant pathway, (ii) concentration index across group–context cells, (iii) appeal overturn rate, (iv) user drop-off after classification, and (v) workload shifts (agent escalations). Suppose the results show: residual rate declines from 27/1,000 to 11/1,000; concentration index declines from 3.6 to 1.9; and drop-off declines by 18%. The board nonetheless observes a small increase in a different residual type ("translation ambiguity" in multilingual prompts), prompting a new, separate docket. This illustrates the core principle: discretization cannot eliminate residuals, but it can institutionalize how residuals become intelligible signals for abductive revision rather than being suppressed as noise or handled as discretionary exceptions.

The lesson is philosophical: residuals are not merely technical noise but sites where the schema's conditions of intelligibility and legitimacy become contestable. An open schema regime makes such contestability institutionally visible and answerable.

6. Conclusions: From Audit Rituals to Open Schema Governance

This article began from a practical paradox in contemporary AI governance. As checklists, documentation requirements, and auditing practices proliferate, governance becomes more verifiable without becoming more responsive. The reason is structural: discretization produces residuals, and audit-as-ritual stabilizes them [5–7]. Residuals are not accidental noise but the political remainder of category-making—especially visible where harms concentrate in edge cases, ambiguous subjects, and minority contexts.

To move beyond this paradox, we reconstructed discretization as an abductive process by tracing a lineage from German Idealism to the logic of discovery. Kant's schematism clarifies that categories require mediating rules which inevitably underdetermine certain cases [9]. Ludwig Wittgenstein's concept of "family resemblances" exposes the inherent fluidity of concepts that discretization attempts to freeze [40]. Niklas Luhmann and Elena Esposito demonstrate the systemic necessity and temporal asynchrony of this reduction, revealing why "blind spots" are functional inevitabilities [8,14,41]. René Thom's catastrophe theory serves as a heuristic to model the residual as a singularity where stability breaks down [39]. Hans-Georg Gadamer and Maurice Merleau-Ponty ground the residual in the hermeneutic condition of understanding and the perceptual ground that sustains the figure [42,43]. Fichte's *Anstoß* interprets resistance as a check that forces reconfiguration [10], structurally anticipating Peirce's surprising fact that triggers abduction [15]. Schelling reframes remainder and opacity as motors of formation [11], and Wilhelm Dilthey emphasizes that technical

 doi:10.20944/preprints202601.1719.v1

"explanation" (*Erklären*) cannot replace the "understanding" (*Verstehen*) of the historical nexus of life [12]. Finally, Husserl deepens the diagnosis by warning that formal systems can erase the lifeworld conditions under which harms become intelligible [13].

Together these resources ground a normative claim: because residuals are inevitable and politically concentrated, governance must institutionalize abductive pathways for category revision rather than treating revision as an optional extra.

On this basis, we proposed an open schema orientation for AI governance. Its core instruments include: (i) a Residual Ledger that records the remainder the taxonomy cannot absorb; (ii) a dual-layer evidence format that preserves narrative harm alongside standardized documentation; (iii) public justification records for high-stakes thresholds as normative commitments; and (iv) category revision protocols with explicit triggers and versioned taxonomies [25–27]. These instruments do not aim to eliminate residuals—an impossible goal—but to prevent residual concentration from becoming invisible, and to protect the institutional conditions under which residuals can drive conceptual evolution.

The broader implication is a shift in what counts as accountability. Governance should not be evaluated solely by the completeness of audit artifacts or by the availability of post hoc explanations. It should also be evaluated by its capacity to revise its schemata under pressure—by whether it can own its discretization rather than hiding it behind ritual verification. In this sense, the politics of residuals is not a peripheral concern but a diagnostic of governance maturity: just governance is not governance without residuals, but governance that can be transformed by them.

Future work can complement this philosophical framework with empirical validation. For example, one may quantify residual concentration and its relationship to threshold choices, track appeal overturn rates as a signal of schema mismatch, or—in physical AI—measure how intermediate action entropy correlates with task success when a system faces ambiguous environments. These studies would not replace the conceptual claim; they would operationalize it across domains. The essential point remains: governance must institutionalize abductive revision, because residuals are where politics and cognition meet.

References

1. Jobin, A.; Ienca, M.; Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399.
2. Mittelstadt, B. Principles alone cannot guarantee ethical AI. *Nat. Mach. Intell.* **2019**, *1*, 501–507.
3. Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I.D.; Gebru, T. Model Cards for Model Reporting. In Proceedings of the FAT*; ACM: New York, NY, USA, **2019**; pp. 220–229.
4. Raji, I.D.; Smart, A.; White, R.N.; Mitchell, M.; Gebru, T.; Hutchinson, B.; Smith-Loud, J.; Theron, D.; Barnes, P. Closing the AI Accountability Gap. In Proceedings of FAccT; ACM: New York, NY, USA, **2020**; pp. 33–44.
5. Power, M. *The Audit Society: Rituals of Verification*; Oxford University Press: Oxford, UK, **1997**.
6. Strathern, M. (Ed.) *Audit Cultures*; Routledge: London, UK, **2000**.
7. Bowker, G.C.; Star, S.L. *Sorting Things Out*; MIT Press: Cambridge, MA, USA, **1999**.
8. Luhmann, N. *Social Systems*; Stanford University Press: Stanford, CA, USA, **1995**.
9. Kant, I. *Critique of Pure Reason*; Guyer, P., Wood, A., Trans.; Cambridge University Press: Cambridge, UK, **1998**.
10. Fichte, J.G. *Foundations of the Entire Science of Knowledge*; Heath, P., Lachs, J., Trans.; Cambridge University Press: Cambridge, UK, **1982**.
11. Schelling, F.W.J. *System of Transcendental Idealism*; Heath, P., Trans.; University of Virginia Press: Charlottesville, VA, USA, **1978**.
12. Dilthey, W. *Einleitung in die Geisteswissenschaften*; Duncker & Humblot: Leipzig, Germany, **1883**.
13. Husserl, E. *The Crisis of European Sciences and Transcendental Phenomenology*; Carr, D., Trans.; Northwestern University Press: Evanston, IL, USA, **1970**.

 doi:10.20944/preprints202601.1719.v1

14. Luhmann, N. *Risk: A Sociological Theory*; de Gruyter: Berlin, Germany, **1993**.
15. Peirce, C.S. *Collected Papers of Charles Sanders Peirce*; Hartshorne, C., Weiss, P., Burks, A.W., Eds.; Harvard University Press: Cambridge, MA, USA, **1931–1958**.
16. Magnani, L. Human abductive cognition vindicated. *Philosophies* **2022**, *7*, 15.
17. Rudin, C. Stop explaining black box machine learning models. *Nat. Mach. Intell.* **2019**, *1*, 206–215.
18. Miller, T. Explanation in artificial intelligence. *Artif. Intell.* **2019**, *267*, 1–38.
19. Wachter, S.; Mittelstadt, B.; Russell, C. Counterfactual explanations without opening the black box. *Harv. J. Law Technol.* **2018**, *31*, 841–887.
20. Doshi-Velez, F.; Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv* **2017**, arXiv:1702.08608.
21. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why should I trust you?". In Proceedings of SIGKDD; ACM: New York, NY, USA, **2016**; pp. 1135–1144.
22. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30; **2017**; pp. 4765–4774.
23. Floridi, L. Establishing the rules for building trustworthy AI. *Nat. Mach. Intell.* **2019**, *1*, 261–262.
24. Mittelstadt, B.D.; Allo, P.; Taddeo, M.; Wachter, S.; Floridi, L. The ethics of algorithms. *Big Data Soc.* **2016**, *3*, 1–21.
25. European Commission. *Ethics Guidelines for Trustworthy AI*; Brussels, Belgium, **2019**.
26. NIST. *AI Risk Management Framework (AI RMF 1.0)*; NIST: Gaithersburg, MD, USA, **2023**.
27. ISO/IEC. *ISO/IEC 23894:2023*; ISO: Geneva, Switzerland, **2023**.
28. Selbst, A.D.; Boyd, D.; Friedler, S.A.; Venkatasubramanian, S.; Vertesi, J. Fairness and abstraction in sociotechnical systems. In Proceedings of FAT*; **2019**; pp. 59–68.
29. Barocas, S.; Hardt, M.; Narayanan, A. *Fairness and Machine Learning: Limitations and Opportunities*; MIT Press: Cambridge, MA, USA, **2023**.
30. Dwork, C.; et al. Fairness through awareness. In Proceedings of ITCS; **2012**; pp. 214–226.
31. Hardt, M.; Price, E.; Srebro, N. Equality of opportunity in supervised learning. In Advances in NIPS; **2016**; pp. 3315–3323.
32. Gebru, T.; et al. Datasheets for datasets. *Commun. ACM* **2021**, *64*, 86–92.
33. Eubanks, V. *Automating Inequality*; St. Martin's Press: New York, NY, USA, **2018**.
34. Noble, S.U. *Algorithms of Oppression*; NYU Press: New York, NY, USA, **2018**.
35. Kroll, J.A.; et al. Accountable algorithms. *Univ. Pa. Law Rev.* **2017**, *165*, 633–705.
36. Gunning, D.; Aha, D.W. DARPA's explainable artificial intelligence (XAI) program. *AI Mag.* **2019**, *40*, 44–58.
37. Binns, R. Fairness in machine learning. In Proceedings of FAT*; **2018**; pp. 149–159.
38. Passi, S.; Barocas, S. Problem formulation and fairness. In Proceedings of FAT*; **2019**; pp. 39–48.
39. Thom, R. *Structural Stability and Morphogenesis*; W. A. Benjamin: Reading, MA, USA, **1975**.
40. Wittgenstein, L. *Philosophical Investigations*; Anscombe, G.E.M., Trans.; Basil Blackwell: Oxford, UK, **1953**.
41. Esposito, E. *Artificial Communication*; MIT Press: Cambridge, MA, USA, **2022**.
42. Gadamer, H.-G. *Truth and Method*; Weinsheimer, J., Marshall, D.G., Trans.; Continuum: London, UK, **2004**.
43. Merleau-Ponty, M. *L'Œil et l'Esprit*; Gallimard: Paris, France, **1964**.