Quantification, Lived Experience, and AI Bioethics: A Phenomenological Critique with a Documentation Case

YANAN LONG, University of Chicago, United States

This paper develops a general phenomenological critique of quantification overreach in AI bioethics. Its central claim is that metrics and model outputs can move from representational aids to authoritative stand-ins for the phenomena they are meant to describe. We call this mechanism *ethical substitution*. At a conceptual level, the argument draws on Husserl's critique of mathematization and Canguilhem's anomaly/abnormality distinction to show how statistical description can be converted into prescriptive normality in institutional practice. The paper then uses LLM-based EHR summarization for clinical handoffs as one exemplary use case, not as the conceptual center of the argument. This case makes visible three coupled infrastructure risks—compression, selective legibility, and authority shift—that can generate relational harms even when local factual quality appears acceptable. Constructively, the manuscript derives six design and governance requirements and introduces a relationship risk map for evaluating documentation systems beyond accuracy. The contribution is not methodological novelty in "phenomenology + AI" or novelty of each requirement in isolation, but a traceable chain from mechanism diagnosis to implementable governance implications.

1 Introduction

Across AI ethics, one recurring pattern has become difficult to ignore: as systems become easier to evaluate through benchmarks, optimization targets, and deployment-facing metrics, broader normative questions are repeatedly narrowed to whatever those targets can represent [23, 27, 29]. This narrowing does not usually result from bad faith. It emerges from institutional pressure to make difficult ethical problems tractable under time, budget, and coordination constraints, so that governance can proceed through measurable proxies rather than open-ended deliberation. Yet tractability and adequacy are not the same. In bioethical contexts, the gap is especially consequential because decisions concern vulnerable lives, asymmetries of testimony, distribution of uncertainty, and obligations generated through mediated relations of care. What becomes measurable can therefore become what is institutionally justifiable, even when the most consequential dimensions of harm remain relational, interpretive, and historically situated [14, 36]. When that happens, quantified artifacts do more than inform decisions: they structure which claims are admissible, which absences remain invisible, and which actors bear the burden of proving that institutional representations are incomplete.

This manuscript argues that one mechanism is central to this transition: *ethical substitution*. Ethical substitution names the process by which a quantified representation—a score, class label, risk profile, or generated summary—moves from being a partial aid for judgment to being treated as if it were the phenomenon itself in practical and normative decision-making. The mechanism is not reducible to factual error, fairness disparity, or opacity, although it can coexist with all three. A system can meet many local quality criteria and still produce substitution if workflow arrangements position representational artifacts as authoritative stand-ins for persons, situations, and relationships. The conceptual stakes are clarified through Husserl's critique of mathematization and Canguilhem's anomaly/abnormality distinction: descriptive regularities can be operationally converted into prescriptive normality when institutional practices forget the representational conditions under which those regularities were produced. The practical stakes are that once substitution is stabilized, correction becomes harder not because evidence is absent, but because authority has already migrated toward portable artifacts that are easier to circulate than to contest.

Author's Contact Information: Yanan Long, University of Chicago, Chicago, IL, United States, ylong@uchicago.edu, yanan.long439@gmail.com.

The paper develops this argument in a quantification-first sequence and treats LLM-supported documentation as an illustrative stress test rather than conceptual center. Documentation infrastructures are selected because they make authority dynamics unusually visible: generated summaries are produced in routine settings, reused across functions, and consumed by actors who did not witness the originating encounter. Under those conditions, representational convenience can be mistaken for epistemic sufficiency through ordinary workflow steps rather than exceptional breakdown. The case analysis identifies three coupled mechanisms that drive this movement: compression, which converts interpretive density into fluent closure; selective legibility, which privileges consistently encoded signals while marginalizing weakly represented context; and authority shift, which occurs when reused artifacts become default institutional truth through copy-forward circulation. The manuscript then derives governance implications directly from these mechanisms, proposing a coupled requirement bundle and a relationship risk map designed to evaluate documentation systems beyond first-pass accuracy.

The contribution is deliberately bounded. The manuscript does not claim novelty from "phenomenology + AI" as such, nor does it claim that each proposed requirement is individually unprecedented. It also does not claim to replace empirical deployment research with conceptual analysis. Its narrower claim is that current debates often lack a mechanism-specific bridge between philosophical critique and operational governance, and that this bridge can be built in a way that is explicit about intellectual debts, scope limits, and evaluative consequences. This positioning addresses multiple audiences without collapsing their concerns: conceptual-track readers gain clearer novelty boundaries and mechanism vocabulary; design and HCI researchers gain traceable links from critique to workflow commitments; and clinical-informatics and governance stakeholders gain an evaluative artifact that can structure procurement, deployment review, and post-incident analysis without requiring prior allegiance to any single ethical framework.

An explicit boundary condition follows from this framing. The argument is not anti-measurement. Quantification and standardization remain indispensable in healthcare, including for equity and accountability when coupled to reflexive governance [10, 13]. The critique is directed at totalization, the point at which quantified artifacts are granted authority as if they exhausted ethically and clinically relevant reality. If the paper succeeds, it should make that threshold more explicit and actionable: not by rejecting formal methods, but by showing where representation ceases to function as revisable aid and begins to function as unacknowledged substitute.

2 Quantification and Its Limits in Bioethics

Quantification in medicine has non-negotiable value, and any serious critique has to begin by acknowledging that value rather than treating it as a rhetorical concession. Standardized indicators enable comparability across institutions, support epidemiological coordination, and make harmful variation visible at scales where purely narrative approaches cannot reliably intervene [10, 13]. In many domains, counting is not the opposite of care but one of the conditions under which care can be audited, resourced, and defended. The question pursued in this paper is therefore narrower and more demanding than a generic anti-metric posture: where, in AI-mediated bioethical practice, does quantification move from disciplined representation into epistemic overreach. Classical naturalistic accounts of health and disease already reveal why this transition can look deceptively neutral. Statistical regularity appears descriptive, yet normativity reenters through reference-class construction, threshold setting, and institutional interpretation [3, 4]. AI systems inherit this ambiguity and often intensify it because they operate on representational infrastructures assembled through billing logics, workflow constraints, and schema design decisions that are practical but never exhaustive of clinical meaning [30, 31]. What seems to be objective technical substrate is therefore already a historically layered mediation.

Husserl's critique of mathematization and Canguilhem's anomaly/abnormality distinction provide a precise language for diagnosing the resulting failure mode. Husserl does not reject formal method; he identifies the moment when method forgets its own condition of validity and begins to appear as if it directly presents the world [17]. In healthcare AI, this forgetting occurs when operational tractability is treated as epistemic sufficiency and generated or classified outputs are treated as if they exhausted the phenomenon they summarize. Canguilhem's distinction sharpens the ethical stake because it separates anomaly as descriptive deviation from abnormality as prescriptive disvalue [6, 7]. AI classification systems can make this conversion institutionally efficient: once a category has high predictive utility, downstream workflows may begin to treat it as normatively authoritative in eligibility rules, triage priorities, or credibility judgments, even when the category was never designed to bear those judgments. The key point is that this conversion rarely happens through a single explicit declaration. It is assembled through defaults, copy-forward practices, and role-based incentives that progressively reduce interpretive friction while increasing institutional confidence in portable artifacts. Ethical substitution names this process: representations begin as aids, then acquire stand-in authority, and eventually reorganize what institutions can see, contest, and repair.

The payoff of Canguilhem's distinction in machine-learning settings is sharper than a general reservation about statistical reasoning, because it targets how classification systems constitute, rather than merely approximate, the reference classes through which clinical populations are rendered legible to institutional action. Supervised learning procedures map individual cases onto class means, thresholds, or decision boundaries defined across a training distribution, while unsupervised clustering homogenizes populations according to optimization criteria sensitive to the same distributional regularities. In both regimes, deviation from the learned structure is formally treated as error relative to the target class rather than as variation intrinsic to the population under study. When the outputs of such systems enter clinical workflow through risk stratification, documentation templates, triage scoring, or eligibility filtering, descriptive distance from the modal case can be institutionally translated into a prescriptive judgment that the divergent case requires correction, exclusion, or additional scrutiny. This translation is not a correctable bias artifact that disappears with better data collection; it is structural to how prediction-as-classification turns distributional rarity into a default signal of deficit [2, 33]. The consequence for clinical populations with high within-group variation—patients with chronic illness, disability-related embodiments, socially structured access barriers, and caregiving trajectories that are difficult to encode—is that anomaly is routinely presented to downstream actors as if it already warranted abnormality treatment. Disability studies and crip theory have long argued that alienation of this kind arises from normative infrastructures rather than from intrinsic pathology [21, 24], and Canguilhem supplies the philosophical grammar for why the infrastructural point holds even when individual predictions are statistically justified relative to the training distribution. Boorse's biostatistical framework helps identify the reference-class structure but cannot name this conversion, because it presupposes that a correct reference class is already available; Canguilhem keeps the conversion visible as a normative event that technical justification cannot discharge [9, 22].

Hidden normativity and the anomaly-to-abnormality collapse share a deeper presupposition that is worth naming in its own right: the identification of objectivity with the elimination of perspective. From a phenomenological standpoint, this identification is a category error. Genuine objectivity is intersubjectively constituted, which is to say that a claim counts as objective when it is available to anyone positioned to encounter and contest it, not because it has been purified of all perspectival grounding [25]. That formulation is not mystical; it is a recognition that meaningful claims about clinical situations depend on the situated interpretations through which those situations become recognizable in the first place. Removing interpretation does not deliver a more reliable reality; it removes the layer on which judgment, testimony, and contestation operate. For AI bioethics, the consequence is that a "view from nowhere" posture in

Table 1. Three modes of quantification overreach in bioethics

Mode of overreach	How it appears technically legitimate	What is displaced	Downstream authority effect
Category closure	Stable labels and thresholds improve reproducibility and coordination.	Local context, evolving interpretation, and contested narratives.	"Class is case": categories become hard to contest once reused.
False objectivity	Aggregate metrics appear perspective-free and institutionally neutral.	Voice asymmetries and interpretive labor required for judgment.	"Score is truth": outputs treated as complete evidence.
Ethical substitution	Representation is operationally efficient and portable across workflows.	Phenomenon-level meaning in lived and relational contexts.	"Artifact is phenomenon": summaries stand in for persons/cases.

documentation or decision support is both unattainable and counterproductive: it strips the interpretive context without which the represented claim cannot be assessed, while simultaneously presenting the result as more authoritative than any situated account. This objectivistic posture is precisely what the substitution mechanism requires in order to function smoothly, because treating quantified outputs as if they were phenomenon-level truth presupposes that phenomenon-level truth is what a purified aggregate already delivers. Naming the error is not a concession to relativism; it is what allows accountability to be reattached to representation, because responsibility can only be assigned to claims whose interpretive grounding is still in view.

Positioning this argument requires explicit continuity with adjacent literatures, because the contribution is diagnostic refinement rather than framework replacement. The analysis aligns with established critiques of quantification-centric AI ethics and epistemic narrowing in sociotechnical systems [14, 36], and it remains in productive dialogue with phenomenology-informed concerns about decontextualization, testimonial erosion, and first-person displacement in medicine [8]. The nearest neighboring concept is epistemic calcification [15], and the overlap is substantial enough that a forced distinction would be artificial. The distinction retained here is interventional rather than ontological: calcification emphasizes a rigidified epistemic state, while ethical substitution tracks the workflow mechanism by which representational artifacts become practical stand-ins over time. These are not rival diagnoses. Substitution can be one pathway through which calcification stabilizes, and calcified infrastructures can in turn make substitution harder to interrupt. Keeping both terms available helps identify where intervention should be directed, whether at deep representational closure, at concrete authority transfer points, or at the coupling between the two.

Because phenomenology is already a mature resource in AI scholarship, this paper does not claim novelty from simply combining phenomenological vocabulary with technology ethics. Its register is transcendental in a limited, operational sense: it asks what conditions make institutional representations appear complete and authoritative before downstream normative frameworks are applied [1, 16, 28, 38]. That register matters because substitution can occur under deontological, consequentialist, or justice-oriented governance alike if representational mediation has already been forgotten at the level of everyday workflow. Post-phenomenological mediation theory is therefore a neighbor, not an opponent, especially for analyzing artifact-level human-technology relations in medicine [19, 20, 32, 35, 37]. The present manuscript works at a more infrastructural scale where representational compression, institutional portability, and normative force are braided. This is also why documentation systems become the decisive site in the next section:

they are where abstraction, authority, and accountability are continuously co-produced under practical pressure, making substitution dynamics both ordinary and empirically inspectable.

3 An Illustrative Case: LLM Documentation as a Hard Test

The manuscript's conceptual claim is general, but it becomes operationally inspectable through one concrete infrastructure case: LLM-based summarization of electronic health records used in handoff and adjacent documentation workflows [26, 34]. This case is selected not because it exhausts the argument, but because it condenses the key authority dynamics in a tractable form. Documentation artifacts are produced in routine practice, often under severe temporal pressure, by actors managing competing obligations of care, coordination, and institutional compliance. Yet these same artifacts later function as durable reference points for clinicians, administrators, and patients who were absent from the original encounter and must still act on what the record renders salient. In other words, documentation sits at the exact intersection where representational efficiency becomes governance-relevant authority. If substitution can be shown clearly here, it provides a strong basis for identifying similar mechanisms in other AI-mediated bioethical domains.

The case is not a speculative projection. Ambient-scribing pilots, LLM-assisted note generation, and structured-summary tools are already deployed or under active evaluation across integrated health systems, typically framed as relief from clinician documentation burden and as a way to standardize communication across shifts and care transitions [26, 34]. These framings are coherent on their own terms and respond to real operational pressures, including documentation overload, inconsistent handoff quality, and time asymmetries between inpatient and ambulatory contexts. The analytical point is that the same features that deliver the operational benefit—speed, fluency, and reuse across contexts—are the features that enable substitution when institutional assumptions about human authorship are not explicitly updated alongside adoption. In other words, the risk analyzed here is not a forecast about future technology; it concerns how existing deployment patterns interact with documentation's role as durable institutional memory under workflows that were originally designed on the assumption that record entries indexed human interpretive work.

A workflow-level stress test makes the mechanism visible without relying on dramatic edge cases. A clinician receives a generated summary that appears coherent and serviceable, accepts it with limited edits because time and cognitive load are real constraints, and then reuses portions of the text in downstream artifacts oriented to different purposes. Later readers encounter these artifacts under conditions where the originating interaction is inaccessible, and practical decisions are made on the basis of what has become portable institutional memory. None of these steps is implausible, and none individually looks like ethical failure. The risk emerges cumulatively: each handoff prioritizes legibility and continuity, each reuse increases durability, and each additional distance from originating context makes contestation more costly. This cumulative pathway is exactly why substitution cannot be diagnosed by one-off output inspection. It is a temporal and organizational process, not a single defective generation event.

Within this process, the three mechanism terms used throughout the paper—compression, selective legibility, and authority shift—operate as coupled dynamics rather than separable defects. Compression is unavoidable because documentation must reduce interpretive density, but it becomes risky when reduction is presented as closure and unresolved uncertainty is linguistically smoothed into declarative certainty. Selective legibility is produced when model and schema preferences privilege consistently encoded signals while rendering weakly represented but clinically meaningful context informationally peripheral, including caregiver observations, social barriers, stigmatized experiences, and contested narrative timelines. Authority shift follows when repeated reuse makes generated text the default institutional case object for actors who did not witness the original encounter and therefore have fewer resources for

reconstructing omitted context. As these dynamics reinforce one another, institutions can become highly coordinated around representations that are locally coherent yet ethically incomplete.

The case also clarifies why factual adequacy and relational adequacy can diverge in consequential ways. A summary can be accurate on coded diagnoses, medications, and chronology while still undermining recognition, trust, and accountability if it flattens voice distinctions, hides disagreement, or makes correction pathways cumbersome and non-propagating [8, 12]. In patient-facing contexts, this can appear as misrecognition: individuals encounter institutionally polished narratives that do not reflect how their situation was experienced or contested. In clinician-facing contexts, this can appear as false certainty: downstream teams inherit prose that reads complete even when key ambiguities were unresolved at the point of care. The novelty claim is therefore infrastructural rather than sensational. Documentation has always abstracted, and earlier regimes also excluded; LLM-mediated documentation intensifies speed, fluency, and portability in ways that increase the likelihood that abstraction will be mistaken for sufficiency.

It is worth spelling out what is specifically new relative to pre-ML documentation risks, because the reviewer-facing version of this question cannot be answered simply by appeal to scale. Pre-LLM records were also reductive, yet reduction was typically performed by humans who were present in or proximate to the encounter and who could exercise clinical judgment at the moment of writing, including judgments about what to leave tentative, contestable, or deliberately underspecified. LLM-mediated summarization separates compression from encounter presence: the compressor has no situated grasp of the case, the output is produced at near-zero marginal cost, and the resulting prose is fluent enough to invite limited editing even when interpretive labor has not actually been performed. These properties together make compressed narratives unusually portable and unusually authoritative, because the surface signals that previously tracked careful human authorship—readability, coherence, stylistic restraint—no longer track the underlying interpretive labor they used to index [11]. The infrastructural novelty therefore lies in this mismatch between the surface form of the artifact and its actual epistemic warrant, which is also why mitigation has to operate through workflow, authorship policy, and reuse governance rather than only through model-level performance improvements. On this reading, the relevant question is not whether LLM-generated summaries are more reductive than their predecessors, but whether institutional practices have updated to reflect that the usual cues for trusting a well-written record no longer apply under automated generation.

A compact phenomenological corrective follows directly from this diagnosis. Lifeworld orientation requires that abstractions remain answerable to situated practice rather than treated as autonomous reality [17]. Intersubjective accountability requires representational forms that preserve perspective differences and make disagreement visible rather than absorbed into one neutral-seeming institutional voice. Record-making must be treated as active mediation that configures relevance and credibility, not passive capture of pre-given facts [5, 30, 31]. Renewal names the institutional obligation to revisit representational defaults as harms, incentives, and workflows change [18]. These commitments do not replace empirical study; they specify what empirical evaluation must track when the ethical problem is not only whether outputs are correct, but whether representations remain contestable and corrigible as they circulate.

4 Design and Evaluation Implications

If ethical substitution is the primary mechanism-level risk, governance cannot remain model-centric. The practical design task is to govern how representational artifacts acquire, retain, and propagate authority as they move through organizational workflows where actors, purposes, and accountability demands differ. This reframing shifts the question from “how do we improve summary quality” to “how do we prevent summaries from becoming unchallengeable

stand-ins for persons and situations." The answer requires workflow commitments rather than isolated model capabilities, because substitution is produced through socio-technical coupling: interface defaults, authorship conventions, copy-forward habits, performance incentives, and correction infrastructure. It is also why the manuscript emphasizes derivation and interaction instead of checklist novelty. Many proposed controls sound familiar in isolation, but their ethical force depends on whether they jointly interrupt authority transfer under realistic institutional constraints.

The first control cluster concerns how generated text enters the record. Draftness-by-default should make uptake explicit and revision expected, so that model prose does not silently inherit finished status through convenience. Draftness is not a cosmetic banner; it is an authorship policy encoded in interaction design, review workflow, and documentation norms. Provenance should similarly exceed source traceability and include omission signaling, so readers can distinguish "not encoded" from "not clinically relevant" and avoid treating representational absence as evidentiary absence. Multi-voice representational structure should preserve distinctions among patient account, caregiver context, and clinician interpretation, because voice fusion routinely creates false objectivity by hiding where testimony ends and institutional interpretation begins. Together, these controls keep representational partiality visible at the point of entry, where substitution pressure is often highest due to workload and time compression.

The second control cluster concerns what happens after entry, when artifacts travel. Uncertainty and disagreement must be first-class representational states that persist across handoffs without being penalized as documentation failure, since epistemic closure can be dangerous in exactly those cases where safe care depends on keeping multiple hypotheses alive. Repair loops must be visible, routine, and role-appropriate, allowing both clinician amendments and structured channels for patient or caregiver contestation where feasible. Most importantly, repair has to propagate across reused artifacts; local correction that remains trapped in one node leaves downstream authority effects intact elsewhere in the system. Purpose separation should govern migration across functions so language generated for one task does not silently acquire authority in another with different interpretive stakes, such as when handoff prose is repurposed for billing, quality surveillance, or patient-facing communication. Without explicit recontextualization rules, reuse becomes the institutional engine by which convenience is converted into normative force.

These controls are most effective as a coupled bundle because substitution dynamics exploit gaps between isolated safeguards. Draft indicators without uncertainty support still produce fluent closure. Provenance without omission signaling still allows weakly encoded context to disappear as if irrelevant. Repair pathways without propagation still fail after authority has shifted through copy-forward reuse. Purpose tagging without voice differentiation can still transport flattened narratives across organizational boundaries. For this reason, governance assessment should prioritize interaction effects, burden distribution, and failure propagation pathways rather than binary feature presence. The traceability map below summarizes how each mechanism corresponds to recurrent workflow failure and to derived control logic.

Evaluation criteria should mirror this mechanism-level view. If risk is infrastructural and relational, output-level metrics alone will under-detect the most consequential failures, especially those that emerge after reuse rather than at generation time. A robust assessment bundle therefore combines quantitative signals such as omission frequency, uncertainty suppression rates, override trajectories, and correction propagation depth; workflow analysis of adoption under realistic time pressure, revision burden by role, and correction latency; qualitative inquiry into recognition, trust, and contestability for affected stakeholders; and governance audit of whether repair remains accessible and effective once artifacts migrate across contexts. Importantly, these layers should not be treated as parallel dashboards with equal interpretive status: they should be read together to identify where local technical adequacy coexists with relational degradation.

Table 2. Traceability from mechanism diagnosis to derived requirement

Risk mechanism	Typical workflow failure	Derived requirement(s)	Primary governance objective	Status in standard frameworks
Compression	Uncertainty and disagreement collapse into fluent closure during handoff and reuse.	Draftness by default; uncertainty/disagreement as first-class.	Keep interpretive openness visible and institutionally legitimate.	Partly addressed; explicit uncertainty carry-forward is uncommon.
Selective legibility	Weakly encoded context is omitted and interpreted as clinically irrelevant.	Provenance with omission signaling; multi-voice structure.	Prevent absence from becoming non-existence in downstream reasoning.	Partly addressed; omission signaling and voice structure are uncommon.
Authority shift	Reused summaries become default institutional truth beyond original context and purpose.	Repair loops with propagation; purpose separation and reuse governance.	Keep contestability and correction effective after artifact migration.	Rarely explicit as coupled workflow governance.

Temporal design also matters because substitution harms have mixed latency. Some are immediate, as when uncertainty is erased in a handoff that requires cautious coordination. Others are delayed, as when partial narratives harden through repeated copy-forward and become institutional defaults before contestation is possible. Evaluation windows that stop at first-pass accuracy can therefore produce a false sense of safety by missing precisely those failures that depend on organizational time and artifact travel. Longitudinal monitoring of correction propagation, reuse pathways, and burden asymmetries is necessary to determine whether governance mechanisms are interrupting substitution or merely documenting it after the fact.

To operationalize this stance, the relationship risk map in Table 3 links relationship-artifact pairs to mechanism type, likely relational harm, repair pathway, and concrete evaluative question. Its value is practical: design teams, procurement reviewers, and governance committees must specify where authority is exercised, who can challenge it, and what institutional process turns challenge into downstream change. The map is intended for deployment readiness, post-incident retrospectives, and periodic governance review, and it can be locally adapted as long as the mechanism-harm-repair-observation chain remains explicit.

Under this view, benchmark strength is necessary but not sufficient for institutional trustworthiness. A system cannot count as governance-ready if it systematically suppresses uncertainty, erases perspective distinctions, or makes correction practically non-propagating after reuse. In documentation ecologies, those properties are not optional ethical enhancements layered on top of performance; they are validity conditions for deciding whether quantified artifacts should be granted durable institutional authority.

5 Tradeoffs, Boundaries, and Related Positions

The argument in this manuscript is intentionally strong about substitution risk, but it is not a maximalist rejection of quantification. Quantified infrastructures remain indispensable for continuity of care, public-health coordination, quality auditing, and equity monitoring, and in many settings those functions are preconditions for accountability

Table 3. Relationship risk map (worked examples for documentation AI)

Relationship	Artifact	Primary mechanism	Likely relational harm	Repair pathway	Evaluation question
Clinician–clinician	Handoff summary	Compression	Unresolved differentials collapse into one storyline, weakening coordination under ambiguity.	Uncertainty addendum and contested-items field.	How often are unresolved items preserved versus collapsed at handoff?
Patient–clinician	Portal or discharge reuse	Selective legibility	Lived context is omitted, producing misrecognition and trust erosion in patient-facing communication.	Patient flag-and-amend with visible clinician response.	What share of patient-reported omissions is acknowledged and integrated?
Patient–institution	Secondary reuse across billing, triage, and quality	Authority shift	Partial narratives become difficult-to-contest institutional defaults across administrative decisions.	Purpose-gated reuse with propagated correction logs.	When a summary is corrected, where else is the correction reflected and how fast?

rather than obstacles to it [10, 13]. Documentation AI can likewise provide real operational benefits when clinical teams are under severe temporal and cognitive pressure, including faster handoff preparation and reduced omission of routine details. These gains are not peripheral caveats; they are part of why institutions adopt such systems and why governance has to be precise rather than oppositional. The normative question is therefore not whether medicine should simplify complex encounters, because some simplification is unavoidable, but how simplification can be kept answerable to lived and relational context once artifacts begin to circulate with institutional authority.

Overreach begins when representational convenience is treated as epistemic sufficiency. In practice this transition is incremental: generated prose is accepted because it is coherent, reused because it is available, and trusted because repeated visibility is mistaken for evidentiary completeness. No single step requires blatant falsehood, which is why substitution should not be reduced to factual error. A summary can remain highly accurate on codable details while still becoming harmful as an authority object if uncertainty is erased, perspective distinctions are flattened, and correction pathways are burdensome or non-propagating. Under these conditions, institutions can become more coordinated around partial representations at the same time that they become less responsive to contestation from people most affected by those representations. Governance regimes that focus only on first-pass factuality will miss this dynamic because the core problem is not isolated inaccuracy but authority drift through ordinary reuse.

A predictable objection is that many requirements proposed in this paper resemble established responsible-AI language and therefore may appear conceptually redundant. The manuscript accepts this lexical overlap and does not claim novelty for individual controls taken one by one. Its narrower claim is that mechanism-specific derivation and coupling change how those controls should be interpreted and evaluated: draftness, provenance, voice differentiation, uncertainty representation, repair propagation, and purpose separation are treated as interacting safeguards against identifiable substitution pathways rather than as a general ethics menu. This positioning is also why the paper is framed as complementary to adjacent critique traditions instead of framework-competitive. It extends quantification-centered AI ethics by specifying where institutional authority transfer is operationalized in documentation infrastructures [14, 36], and it remains in productive proximity to epistemic calcification accounts by distinguishing a workflow process of stand-in formation from a structural state of epistemic rigidity [15]. The practical value of this distinction is interventional: it helps identify whether governance should target representational closure, authority transfer points, or both.

Scope limits are substantive and should be explicit. The argument is scenario-based and mechanism-driven, not prevalence-estimating; it proposes testable governance hypotheses rather than validated deployment prescriptions across health systems. The applied analysis is centered on documentation infrastructures, so transfer to triage, imaging, or decision-support domains should be argued empirically rather than assumed by analogy. The paper also concentrates on workflow and institutional governance and therefore does not model political-economy constraints in full detail, including procurement asymmetries, vendor dependence, or reimbursement incentives that shape implementation behavior. These limits produce a direct empirical agenda: trace where compression, selective legibility, and authority shift arise in live settings; compare requirement bundles under realistic staffing and time conditions to identify interaction effects; and measure whether corrections propagate across downstream artifacts quickly enough to alter decisions in practice rather than only in audit trails.

The boundary condition is therefore bounded pluralism. Responsible documentation AI requires quantitative discipline and interpretive accountability together, with neither register allowed to substitute for the other. Quantified artifacts remain legitimate when they are explicitly partial, contestable in use, and corrigible across reuse chains; institutional trust should become conditional when systems suppress uncertainty, erase perspective differences, or render repair effectively non-propagating. Framed this way, the governing question is not whether a system quantifies, but how quantified representations are positioned, challenged, and revised within the relations of care they mediate over time.

6 Conclusion

This manuscript advanced a general phenomenological critique of quantification overreach in AI bioethics and named one mechanism at its center: ethical substitution, the process by which representational artifacts begin to function as stand-ins for the phenomena they were meant to support. The argument did not depend on any single application domain, but it was made concrete through one illustrative use case—LLM-supported clinical documentation—because documentation infrastructures reveal substitution dynamics with unusual clarity.

The paper's core contribution is the explicit chain connecting levels that are often separated in practice: conceptual diagnosis, workflow mechanism analysis, design and governance requirements, and evaluation artifacts. At the conceptual level, Husserlian critique of mathematization and Canguilhem's anomaly/abnormality distinction clarified how descriptive regularities can be institutionally converted into prescriptive authority. At the applied level, the documentation case showed how compression, selective legibility, and authority shift can produce relational harm even when local factual quality appears acceptable. At the practical level, the manuscript derived a coupled requirement bundle and proposed a relationship risk map to support evaluation beyond accuracy.

Equally important are the limits the paper set for itself. It did not claim novelty from applying phenomenology to AI in general; it did not claim each governance control is individually unprecedented; and it did not claim scenario-based analysis can replace empirical deployment research. Its narrower claim was that mechanism-specific derivation improves governance precision by making intervention assumptions explicit and auditable.

The normative stance remained bounded throughout. Quantification is indispensable in healthcare and often necessary for accountability. The critique was directed at totalization, not measurement as such. On this view, phenomenological bioethics contributes a practical discipline of limits: keep representation answerable to lived context, preserve uncertainty where it is epistemically warranted, maintain contestability across workflow transitions, and ensure that correction remains effective after artifacts travel. If those conditions are not met, model-centered success metrics are insufficient grounds for institutional trust.

References

[1] Andreea Smaranda Aldea. 2016. Phenomenology as Critique: Teleological–Historical Reflection and Husserl' s Transcendental Eidetics. *Husserl Studies* 32, 1 (April 2016), 21–46. doi:10.1007/s10743-016-9186-8

[2] Giuseppe Bianco, Charles T. Wolfe, and Gertrudis Van De Vijver (Eds.). 2023. *Canguilhem and Continental Philosophy of Biology*. History, Philosophy and Theory of the Life Sciences, Vol. 31. Springer International Publishing, Cham. doi:10.1007/978-3-031-20529-3

[3] Christopher Boorse. 1977. Health as a Theoretical Concept. *Philosophy of Science* 44, 4 (1977), 542–573. https://www.jstor.org/stable/186939 Publisher: [Cambridge University Press, The University of Chicago Press, Philosophy of Science Association].

[4] Christopher Boorse. 2014. A Second Rebuttal On Health. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 39, 6 (Dec. 2014), 683–724. doi:10.1093/jmp/jhu035

[5] Geoffrey C. Bowker and Susan Leigh Star. 2000. *Sorting things out: classification and its consequences*. MIT Press, Cambridge (Mass.).

[6] Georges Canguilhem. 2018. *Histoire des sciences, épistémologie, commémorations: 1966-1995*. Oeuvres complètes, Vol. 5. Librairie philosophique J. Vrin, Paris.

[7] Georges Canguilhem. 2021. *Écrits de médecine et de philosophie*. Œuvres complètes, Vol. 2. Vrin, Paris.

[8] Benjamin Chin-Yee and Ross Upshur. 2019. Three Problems with Big Data and Artificial Intelligence in Medicine. *Perspectives in Biology and Medicine* 62, 2 (2019), 237–256. https://muse.jhu.edu/pub/1/article/728485

[9] Bas de Boer and Ciano Aydin. 2023. Empowerment: Freud, Canguilhem and Lacan on the ideal of health promotion. *Medicine, Health Care and Philosophy* 26, 3 (Sept. 2023), 301–311. doi:10.1007/s11019-023-10145-z

[10] Catherine D'Ignazio. 2024. *Counting Feminicide: Data Feminism in Action*. The MIT Press. doi:10.7551/mitpress/14671.001.0001

[11] Henry Farrell, Alison Gopnik, Cosma Shalizi, and James Evans. 2025. Large AI models are cultural and social technologies. *Science* 387, 6739 (March 2025), 1153–1156. doi:10.1126/science.adt9819

[12] Adam Frank, Marcelo Gleiser, and Evan Thompson. 2024. *The blind spot: why science cannot ignore human experience*. The MIT Press, Cambridge, Massachusetts.

[13] Lawrence Goldman. 2022. *Victorians and Numbers: Statistics and Society in Nineteenth Century Britain* (1 ed.). Oxford University PressOxford. doi:10.1093/oso/9780192847744.001.0001

[14] Thilo Hagendorff. 2022. Blind spots in AI ethics. *AI and Ethics* 2, 4 (Nov. 2022), 851–867. doi:10.1007/s43681-021-00122-8

[15] Mahi Hardalupas. 2024. Contributory Injustice, Epistemic Calcification and the Use of AI Systems in Healthcare. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 7 (Oct. 2024), 573–583. doi:10.1609/aies.v7i1.31659

[16] Sara Heinämaa. 2021. On the transcendental undercurrents of phenomenology: the case of the living body. *Continental Philosophy Review* 54, 2 (June 2021), 237–257. doi:10.1007/s11007-021-09534-z

[17] Edmund Husserl. 1976. *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie: eine Einleitung in die phänomenologische Philosophie* (2. aufl., photomechan. nachdr ed.). Husserliana gesammelte Werke, Vol. VI. Nijhoff, Haag.

[18] Edmund Husserl. 1989. *Aufsätze und Vorträge (1922-1937)*. Husserliana gesammelte Werke, Vol. XXVII. Kluwer Academic Publishers, Dordrecht. OCLC: 17108390.

[19] Don Ihde. 2016. *Husserl's missing technologies* (first edition ed.). Fordham University Press, New York.

[20] Don Ihde. 2019. *Medical Technics*. University of Minnesota Press. doi:10.5749/j.ctvpmw56v

[21] Corinne Lajoie. 2023. Disability Studies and Phenomenology. In *Encyclopedia of Phenomenology*, Nicolas de Warren and Ted Toadvine (Eds.). Springer International Publishing, Cham, 1–13. doi:10.1007/978-3-030-47253-5_330-1

[22] Céline Lefève. 2014. De la philosophie de la médecine de Georges Canguilhem à la philosophie du soin médical. *Revue de métaphysique et de morale* 82, 2 (2014), 197–221. doi:10.3917/rmm.142.0197

[23] Zari McFadden and Lauren Alvarez. 2024. Performative Ethics From Within the Ivory Tower: How CS Practitioners Uphold Systems of Oppression. *Journal of Artificial Intelligence Research* 79 (March 2024), 777–799. doi:10.1613/jair.1.15423

[24] Robert McRuer. 2006. *Crip theory: cultural signs of queerness and disability*. New York University Press, New York. OCLC: 173511594.

[25] Maurice Merleau-Ponty. 2013. *Phénoménologie de la perception*. Gallimard, Paris. OCLC: 937881458.

[26] Bertalan Meskó and Eric J. Topol. 2023. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digital Medicine* 6, 1 (July 2023), 1–6. doi:10.1038/s41746-023-00873-0

[27] Brent Mittelstadt. 2019. Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* 1, 11 (Nov. 2019), 501–507. doi:10.1038/s42256-019-0114-4 Number: 11.

[28] Dermot Moran. 2008. Husserl' s transcendental philosophy and the critique of naturalism. *Continental Philosophy Review* 41, 4 (Dec. 2008), 401–425. doi:10.1007/s11007-008-9088-3

[29] Luke Munn. 2023. The uselessness of AI ethics. *AI and Ethics* 3, 3 (Aug. 2023), 869–877. doi:10.1007/s43681-022-00209-w Number: 3.

[30] Minna Ruckenstein and Natasha Dow Schüll. 2017. The Datafication of Health. *Annual Review of Anthropology* 46, 1 (Oct. 2017), 261–278. doi:10.1146/annurev-anthro-102116-041244 Number: 1.

[31] Clare Southerton. 2022. Datafication. 358–361 pages. doi:10.1007/978-3-319-32010-6_332

[32] Daniel Susser. 2016. Ihde' s Missing Sciences: Postphenomenology, Big Data, and the Human Sciences. *Techné: Research in Philosophy and Technology* 20, 2 (Aug. 2016), 137–152. doi:10.5840/techne201672754

[33] Samuel Talcott. 2019. *Georges Canguilhem and the Problem of Error.* Springer International Publishing, Cham. doi:10.1007/978-3-030-00779-9

[34] Lu Tang, Jinxu Li, and Sophia Fantus. 2023. Medical artificial intelligence ethics: A systematic review of empirical studies. *DIGITAL HEALTH* 9 (Jan. 2023), 20552076231186064. doi:10.1177/20552076231186064

[35] Shannon Vallor. 2016. Ihde, Technoscience, and the Resilience of Phenomenology. *Techné: Research in Philosophy and Technology* 20, 2 (Aug. 2016), 90–94. doi:10.5840/techne201672550

[36] David Gray Widder. 2024. Epistemic Power in AI Ethics Labor: Legitimizing Located Complaints *(FAccT '24)*. Association for Computing Machinery, New York, NY, USA, 1295–1304. doi:10.1145/3630106.3658973

[37] Harald A. Wiltsche. 2017. Mechanics Lost: Husserl ' s Galileo and Ihde ' s Telescope. *Husserl Studies* 33, 2 (July 2017), 149–173. doi:10.1007/s10743-016-9204-x

[38] Dan Zahavi. 2017. *Husserl's legacy: phenomenology, metaphysics, and transcendental philosophy.* Oxford university press, Oxford.