

THE EPISTEMIC SUBSTITUTION ERROR: WHY ACCURATE ALGORITHMIC DECISIONS CAN STILL BE ILLEGITIMATE

Prof. Dr. Kwan Hong Tan¹

¹ Singapore University of Social Sciences, Singapore, ORCID: 0009-0003-9276-2829, khtan055@suss.edu.sg

Corresponding author: Prof. Dr. Kwan Hong Tan, khtan055@suss.edu.sg

ABSTRACT

Institutions increasingly rely on algorithmic systems to assess people, allocate opportunities, and justify consequential decisions. Existing debates often ask whether these systems are accurate, fair, transparent, or explainable. This article identifies a distinct philosophical problem: an institution may reach a factually correct outcome while illegitimately replacing a person's standing to offer relevant reasons with an institutional prediction of what those reasons must be. Through analytical reconstruction, contrastive thought experiments, and normative stress tests, the article develops the concept of epistemic substitution, distinguishes predictive evidence from participatory justification, and proposes a non-substitution principle for decisions affecting persons as accountable agents. It demonstrates that prediction, explanation, and even accurate individualized classification do not by themselves establish that affected persons have been appropriately heard. The resulting framework specifies five conditions for legitimate institutional uptake: access, admissibility, responsiveness, traceability, and revisability. It also distinguishes bounded circumstances in which substitution is permissible from those in which it is not, and derives researchable expectations for public administration, employment, education, and automated peer review. The central conclusion is neither that algorithms should never decide nor that personal testimony should automatically prevail. Rather, decision legitimacy requires institutions to preserve the possibility that an affected person can change the reasons on which a decision is made.

Keywords: Epistemic Substitution, Algorithmic Governance, Epistemic Justice, Procedural Legitimacy, Contestability, Philosophy of Technology

Received: 12 October 2026

Accepted: 16 October 2026

Published: 16 October 2026

Cite: Tan, K. H. (2026). The Epistemic Substitution Error: Why Accurate Algorithmic Decisions Can Still Be Illegitimate. *AI-Driven Journal of Interdisciplinary Social Research*, 1(1), 105-123. <https://aijjsr.org/articles/AI-JISR-2026-0010>

CC BY 4.0 — This article is published under the Creative Commons Attribution 4.0 International License.

1. Introduction

Consider a scholarship committee that predicts with exceptional accuracy which applicants will remain enrolled. One student explains that a past interruption resulted from an extraordinary caregiving obligation that has now ended. A statistical profile nevertheless treats previous interruption as a stable indicator of future attrition, and the committee rejects the application without receiving or assessing the explanation. Suppose, further, that the predicted outcome happens to be correct: the student later leaves another programme. Has the institution necessarily treated the student justly? The inference from correct prediction to legitimate treatment is precisely what this article contests.

A decision may be epistemically successful in one respect and normatively deficient in another. Predicting what an individual will do, knowing why a prediction performs well, and giving that individual appropriate standing to contribute reasons to a decision are different achievements. The third cannot be reduced to the first two. This distinction matters wherever institutions convert data into judgments about eligibility, deservingness, competence, credibility, or risk. Such judgments not only estimate outcomes; they also decide which kinds of explanation count as relevant and whose voice can alter the institutional account of a case.

Studies of algorithmic governance identify disparate impacts and the limitations of technical abstraction (Barocas & Selbst, 2016; Selbst et al., 2019). Social epistemology examines exclusion from testimony and knowledge production (Fricker, 2007; Dotson, 2011), while procedural accounts of algorithmic voice show that an accurate output may coexist with deficient opportunities to be heard (Kinchin, 2024). The narrower question here is when a computational account of a person is treated not merely as evidence but as a substitute for that person's legitimate standing to contribute reasons.

This article defines the epistemic substitution error (ESE) as an institutional justificatory failure arising when (a) a proxy-based judgment is treated as dispositive, (b) the institution exercises material authority over an affected person on a matter admitting potentially relevant person-supplied reasons, (c) the person lacks a meaningful opportunity for those reasons to affect the justificatory record, and (d) no proportionate exception justifies that closure. These conditions diagnose a procedural wrong without presupposing that the prediction, outcome, or institution is otherwise biased or mistaken.

The argument proceeds by separating forms of epistemic success, examining foundations in social epistemology and political philosophy, articulating a non-substitution principle, testing that principle against hard cases, and translating it into operational obligations. Three claims organize the analysis. First, accuracy alone does not entail procedural legitimacy. Second, a right to receive reasons is not equivalent to a standing to supply reasons. Third, procedural standing creates genuine institutional obligations only when contribution can make a reason-sensitive difference. These claims are defended as normative propositions rather than reported as experimental findings.

2. Literature Review and Philosophical Foundations

Algorithmic fairness frameworks have exposed how predictive classifications may perpetuate social disadvantage (Barocas & Selbst, 2016) and how computational boundaries can conceal institutional context (Selbst et al., 2019). Fairness criteria are not politically neutral design choices (Binns, 2018). Yet removing measurable disparities does not settle whether affected persons can challenge the grounds on which an institution acts.

Technical transparency is equally incomplete. Burrell (2016) distinguishes several sources of opacity; Ananny and Crawford (2018) show why system visibility does not itself amount to accountability. Creel (2020) provides a more differentiated account of computational transparency. The present concern arises even if an intelligible explanation is supplied: explanation travels from the institution to the person, whereas epistemic uptake requires an accountable return channel.

Fricker (2007) theorizes harms connected to identity-prejudicial credibility deficits and interpretive disadvantage. Dotson (2011) examines silencing in testimonial exchange; Medina (2013) relates epistemic resistance to structural ignorance. These accounts help specify why persons are not simply objects of reliable inquiry. However, the error examined here need not rest on identity prejudice, an intention to silence, or an incorrect credibility judgment.

Anderson (2012) shifts attention from the virtues of individual knowers to the properties of social institutions. This matters when officials trust an accurate model but the workflow prevents evidence outside fixed feature categories from entering the decision. Institutional exclusion can occur without personal animus or technical malfunction.

Young (2000) argues for inclusive communication across social differences, and Tyler (2006) links perceived procedural fairness to legitimacy and compliance. More directly, Kinchin (2024) identifies a procedural gap in algorithmic justice centered on voice and the opportunity to be heard. Perdana (2026) maps gaps between technical fairness and procedural correctness. These analyses establish that algorithmic voice and contestability are not novel discoveries of this article. The additional claim defended here concerns the particular inference by which predictive evidence is allowed to exhaust an individual's standing to introduce reasons, together with a diagnostic test for that inference.

Tan (2026) develops a broader humanistic framework for assessing algorithmic governance, including contestability and dignity. The present article narrows the unit of analysis to an individual institutional decision episode, gives necessary diagnostic elements for the substitution error, and specifies how a reason-sensitive pathway differs from symbolic participation. This relation to earlier work is acknowledged as an extension rather than offered as proof of originality.

Further nearby contributions deserve attention. Ibitoye et al. (2026) examine how artificial intelligence reorganizes judgment through decision substitution; Wong (2026) asks whether legally consequential public judgments are attributable to an answerable office. Kinchin (2024) and Perdana (2026) emphasize procedural voice and correctness, respectively. The distinctive proposed object is the affected person's opportunity to change the justificatory

record even where the model is accurate and an empowered official owns the final verdict. An institution can meet a requirement of accountable decision ownership while still excluding relevant testimony.

The expression “epistemic substitution” is also used in another sense: Theoharakis and Mylonopoulos (2026) study consumers who substitute artificial intelligence for their own evaluative judgment. Their subject is self-delegation to a tool. The error considered here is an institution substituting its representation for someone else’s participation. No priority claim is made for the expression itself.

2.1 Conceptual Boundaries and Diagnostic Counterexamples

The epistemic substitution error is neither synonymous with testimonial injustice nor a label for every procedurally unfair decision. Testimonial injustice in the strict sense involves a credibility deficit arising from identity prejudice (Fricker, 2007). By contrast, substitution can occur under a uniformly applied, prejudice-free rule that refuses all individualized evidence because the prediction is deemed sufficient. Conversely, an official can commit testimonial injustice by discounting a speaker because of identity prejudice even in a process that otherwise allows corrections and reason-sensitive review. The two wrongs may coincide but are not coextensive.

Procedural fairness is the broader family of standards governing consistency, impartiality, notice, voice, and correction. Epistemic substitution is one proposed member of that family: the specific displacement of an affected person’s warranted role in producing the reasons for a decision by a proxy representation. An unfair allocation without any proxy does not automatically instantiate the error. Nor does every unappealable automated transaction instantiate it, because there may be no material, contestable matter requiring individualized uptake.

Explanation and contestability are related but distinct. A person can receive an exact account of why an automated score was assigned without being allowed to supply evidence that changes the grounds used. Equally, a functioning appeal may reject a person’s claim after properly assessing it, without committing substitution. The critical counterfactual is whether relevant new reasons could alter the institutional justification under publicly defensible standards, not whether the claimant always wins or whether the institution always gives an exhaustive technical explanation.

Accordingly, the diagnostic has four elements: material authority, a contestable evidential or normative issue, proxy displacement of person-supplied reasons, and absence of a justified proportional exception. These elements limit its reach. A defective model can be inaccurate without causing substitution if a legitimate review pathway remains open, while a perfectly calibrated model can coexist with substitution if the affected person is excluded from that pathway.

The scope of standing is institutional rather than automatically legal. A public agency deciding a basic entitlement faces stronger justificatory duties than a private supplier making a reversible recommendation. Yet a private employer, credentialing organization, or

educational gatekeeper can exercise comparable practical authority where access to livelihood, recognition, or basic opportunities depends on its decision and alternatives are limited. The proposed obligation tracks the justified power to settle a person-affecting matter, not the simple fact that a computer was used. Ordinary consumer suggestions, freely avoidable transactions, and nonmaterial decisions may therefore fall outside it. These distinctions prevent the principle from silently converting every discretionary private decision into a right to a personal hearing.

3. Method: Analytical Reconstruction and Normative Stress Testing

This is an original theoretical research article in philosophy, informed by political theory, social epistemology, human-computer interaction, and public administration. No interviews, participant recruitment, field observations, or statistical experiments were performed. Its method is analytical reconstruction: identifying an implicit inference in current governance practice, making the premises explicit, examining their normative validity, and proposing a more precise alternative. The article does not present an unreported systematic literature search, a validated scale, or estimates of real-world prevalence.

The reconstruction proceeds in four steps: separating prediction, explanation, substantive permissibility, and reason-giving standing; constructing contrast cases that fix predictive outcomes while varying opportunities for uptake; testing the resulting principle against objections about expertise, burden, manipulation, emergencies, and third-party rights; and specifying institutional duties capable of being evaluated in future research. The contrast cases are conceptual tests of entailment, not evidence about the frequency of real-world harms.

Thought experiments in this article are devices for testing conceptual entailments, not stylized evidence that actual institutions behave in a particular manner. The scholarship, employment, social assistance, and peer-review examples are hypothetical unless a published empirical finding is explicitly cited. The resulting propositions are defended by reasons and counterexamples. Any claims about prevalence, causal effect sizes, or practical feasibility would require separate empirical research. This methodological boundary prevents an attractive normative theory from being misrepresented as a measured intervention.

4. Findings: The Epistemic Substitution Error

The core finding is conceptual: institutions can commit an epistemic substitution error without making a predictive error. To establish this, a conceptual separation is necessary. Predictive adequacy concerns whether available information supports a forecast about a specified outcome. Explanatory adequacy concerns whether the reasons for that forecast or decision are communicable and fit the explanatory question. Substantive permissibility concerns whether a decision satisfies applicable rights, duties, and distributive standards. Participatory legitimacy concerns whether an affected person receives the reason-giving standing appropriate to an institution exercising authority over that person. Success along one dimension does not logically guarantee success along the others.

Consider a deliberately strong case for automation. A model is calibrated, its error rates are equal across relevant groups, and its prediction about an applicant is true. An expert can explain the factors driving the model, and the decision does not violate any pre-existing entitlement. The institution nonetheless refuses to admit the applicant's evidence about a changed circumstance because the workflow accepts only fixed categorical inputs. If the applicant has a legitimate claim to have such evidence considered, the resulting wrong is not a defect in calibration. The defect concerns the authority to determine which reasons may enter the decision at all.

The word "substitution" has a precise meaning here. A proxy may permissibly represent a person for some tasks: a sensor may report attendance, a laboratory measure may quantify blood glucose, and a model may rank approximate risks. The error occurs when the institution moves from representational assistance to normative replacement, treating what the proxy says about the person as exhausting what the person may properly say in relation to the decision. An input can be valid evidence without becoming a full substitute for the subject's standing as a participant in justification.

Proposition 1, accuracy-legitimacy non-entailment: predictive accuracy is insufficient, by itself, to establish procedural legitimacy. Consider two processes with identical predictions and outcomes. One permits relevant person-supplied evidence to enter an accountable review; the other categorically excludes it where an independently justified standing claim exists. The processes can be equal in predictive performance but unequal in the manner in which authority is exercised. This is an independence argument, not a proof that every decision lacking individualized review is wrong; the independent normative claim and its exceptions still require defense.

Proposition 2, the explanation-uptake distinction, states that furnishing understandable reasons for a verdict is not equivalent to giving the affected person standing to contribute reasons to that verdict. An appeal portal containing a model explanation can satisfy the first function while failing the second if no authorized actor can revise the decision or respond to new evidence. Counterfactual explanations, for example, may help applicants understand how altered inputs could change outputs (Wachter et al., 2018); yet guidance about what an applicant should change does not necessarily permit a challenge to what the institution ought to count as relevant.

Proposition 3, the non-substitution principle: when an institution exercises material authority over a person in a matter open to potentially relevant person-supplied reasons, it must provide a practicable, reason-sensitive route through which those reasons can affect the basis of its judgment, unless a proportionate, defensible exception applies. The claimant is not entitled to determine the result or to override better evidence. Standing means an opportunity to contribute to accountable justification. In some contexts, retrospective review can satisfy the principle if it remains timely enough to prevent or remedy material harm.

The normative warrant requires explicit defense. The argument begins from a conditional demand of accountable authority: where an institution authoritatively settles a materially consequential, reasonably contestable matter about a person, it must be able to justify the

grounds of that settlement to the person under rules that treat relevant submissions as evidence rather than interruption. If case-specific information may reasonably affect the application or interpretation of those grounds, categorically excluding it while claiming that an accurate proxy is sufficient undermines that justificatory relationship (Anderson, 2012; Kinchin, 2024; Vredenburg, 2022). This is not a deduction from statistical accuracy, nor a universal legal right. It is a defeasible moral premise about the exercise of authority, and opponents may reject or qualify it. A system can meet the premise through properly empowered human, assisted, or representative review; the right attaches to genuine reason-sensitive uptake, not to a particular staffing model.

These propositions identify an important normative asymmetry. An institution can know a person without hearing that person, while a person can be heard without persuading the institution. Hearing is therefore not a promise of success. It is a promise that a class of reasons will not be excluded solely because the institution already possesses a powerful predictive description. The entitlement is to a fair possibility of affecting justification, not an entitlement to dictate the result.

4.1 Conditions for Meaningful Uptake

Five conditions make meaningful uptake assessable. Access is a usable, timely route for testimony, correction, or contextual evidence. Admissibility is evaluation under defensible relevance rules rather than exclusion merely because evidence lacks a pre-defined input field. Responsiveness is an explanation that engages the submitted reason, including reasons for rejecting it. Traceability is a record of what was offered, how it was considered, and which actor held responsibility. Revisability is the actual institutional authority to change the outcome or its justificatory basis when new material reasons warrant it.

These are distinct obligations. An online submission box establishes access without establishing admissibility. A generic acknowledgement establishes neither responsiveness nor evidence of evaluation. An audit log may deliver traceability without a pathway to remedy. A formal appeal may nominally allow revision while delays or fees make its exercise inaccessible. A system can therefore score well on conventional transparency measures while failing the deeper standing requirement. Conversely, meaningful uptake can occur through human mediation of an automated tool even when the underlying model cannot be fully interpreted.

For high-stakes, contestable decisions the five conditions form a conjunctive standard: merely excelling at explanation cannot compensate for the absence of meaningful revision. Their procedural intensity varies with stakes, discretion, power asymmetry, and likely irreversibility, as Proposition 4 and Table 2 specify. In routine low-stakes, fixed-rule transactions, simple correction and escalation may suffice. The standard requires a pathway capable of reason-sensitive change, not a predetermined appeal success rate.

Uptake is therefore not measured by the share of appeals that succeed. Evidence of compliance includes records linking submissions to reasons for acceptance or rejection, accessible procedures, genuine remedial powers, and safeguards against selective barriers. A system with few reversals may have adjudicated challenges conscientiously; one with many

reversals may still exclude claimants unable to navigate its procedures. Interpretation requires examination of case records and access disparities. Table 1 summarizes the five conditions.

Table 1. Dimensions of institutional epistemic uptake

Condition	Diagnostic question	Minimal institutional evidence
Access	Can the affected person submit relevant reasons in time?	Accessible, timely channel; supported formats
Admissibility	Are new reasons evaluated under defensible relevance rules?	Published criteria; recorded evaluation
Responsiveness	Does a reviewer address the substance of the submission?	Reasoned response linked to evidence
Traceability	Can the process be reconstructed and responsibility assigned?	Dated record and identifiable decision authority
Revisability	Can a justified challenge alter the decision or its basis?	Authorized remedy and correction pathway

Note. These conditions describe a normative institutional standard; they are not a validated psychometric instrument.

4.2 A Limiting Case: Accurate Decisions Without a Wrong

A serious theory must also identify when non-participation is justified. Imagine an automated public library system that calculates an undisputed overdue charge from an agreed tariff and a reliably recorded return date. The patron has a simple, accessible correction mechanism for erroneous records, the amount is trivial, and the calculation involves no evaluative judgment about personal circumstances. It would be extravagant to require an individualized hearing before every transaction. Here limited procedural automation can be legitimate because the material facts and normative rules are fixed, the stakes are low, and an effective correction route remains available.

Now consider a welfare decision that turns not only on an income threshold but on contested judgments about household composition or capacity to work. The stakes are consequential, case-specific testimony may be relevant, and the data categories may underrepresent lived circumstances. The standing claim is correspondingly stronger. The contrast shows why the theory should not conflate personhood with a universal veto over automated procedures. Rather, the degree of institutional responsibility tracks the authority exerted, contestability of the basis, evidential contribution available, and likely cost of a wrong or unanswerable outcome.

Proposition 4, proportionate uptake: as decision stakes, factual or normative contestability, and institutional power asymmetry increase, the justification demanded for restricting any of

the five uptake conditions becomes more stringent. These factors do not constitute an arithmetic score. A serious potential deprivation or an irreversible sanction is sufficient to trigger the strongest presumptive safeguards when relevant person-supplied reasons are possible. Uncontested fixed-rule, low-impact transactions can use lighter mechanisms, while emergency decisions may postpone, but ordinarily may not extinguish, consequential review.

4.3 Application Thresholds and the Exception Test

The trigger for enhanced uptake is conjunctively structured but not numerical. First ask whether an institution exercises consequential authority over an identified person. Next ask whether the evidential record, interpretation of a policy, or inference supporting the outcome is reasonably contestable and could be informed by that person. At intake, a claim is sufficiently material when, if supported, it could reasonably affect the decision rule, its application, or the adequacy of the institutional reasons; a claimant need not prove likely reversal before being allowed to submit it. When both criteria hold, all five conditions are presumptively required. Serious irreversibility, vulnerability, or marked power asymmetry intensifies this presumption rather than creating a numerical cutoff. The institution bears the burden of documenting any restriction. Table 2 specifies normative defaults, not universal legal entitlements.

Table 2. Proportionate institutional uptake across decision contexts

Decision class	Threshold trigger	Minimum justified process
High-stakes, contestable	Threat to basic needs, status, livelihood, or similarly serious interests; materially disputable grounds	All five conditions, with reasoned review and an effective remedy before irreversible harm where practicable
Moderate, context-sensitive	Nontrivial consequences; some plausible individualized evidence or discretionary interpretation	Accessible evidence submission and recorded evaluation; responsive review and escalation when a material challenge is raised
Low-stakes, fixed-rule	Trivial reversible effect; transparent agreed rule; no material case-specific discretion	Low-friction data correction, notice of the rule, and an effective route to rectify mistakes
Urgent provisional action	Imminent significant harm makes prior consultation genuinely impracticable	Documented temporary restriction, independent post-decision review within an actionable window, and restoration or compensation where warranted

Note. These thresholds guide normative assessment. They are not numerical cutoffs, validated legal rules, or proof of empirical effects.

The five conditions also have distinct application thresholds. Access is timely only if submission or assisted representation occurs while a material remedy remains possible; a post-hoc portal is inadequate when an irreversible sanction has already taken effect. Admissibility is triggered by a non-frivolous, potentially material submission under the relevant decision rules, not by an unlimited entitlement to introduce irrelevant claims. Responsiveness must identify the substance accepted or rejected and its bearing on the grounds of decision; a standardized explanation may suffice for materially identical low-impact cases but cannot substitute for addressing a distinctive high-stakes challenge. Traceability requires a case-linked record and identifiable responsible authority in consequential cases, while a simpler correction log may suffice for trivial transactions. Revisability requires an actor or accessible referral pathway empowered to change an erroneous outcome or justification; where the original reviewer cannot amend a governing rule, the limits of that authority and any legitimate further route must be explained.

These requirements concern the quality of available procedures, not rigid service-level deadlines, appeal quotas, or automatic entitlements to policy exceptions. In moderate-stakes cases the institution may use triage, aggregate replies to genuinely identical issues, and escalation when a new material reason arises. In high-stakes cases, triage cannot end in an unreviewable machine denial of a non-frivolous challenge. The rule against compensating away a missing essential condition remains strongest precisely when an outcome is difficult to reverse.

A proportional exception is defensible only if five requirements are met together: (1) a legitimate aim such as preventing imminent harm or protecting another person's rights; (2) a demonstrable necessity, with ordinary workload or institutional convenience insufficient on its own; (3) use of the least restrictive feasible alternative, including assisted submissions or redacted evidence where appropriate; (4) a recorded reason, a responsible reviewer, and protection against systematically unequal restrictions; and (5) an actionable time limit, independent reconsideration where feasible, and a remedy capable of repairing material harm. Failure of any requirement defeats the exception in a high-stakes case. The remedy need not guarantee reversal, but must be more than a notice issued after consequences are irreversible.

The exception test also has a stopping rule. If facts and rules are genuinely fixed, the result is readily reversible, the burden is minimal, and error correction works, an individualized hearing is not ordinarily required. If any of those assumptions becomes contested, the institution must reopen the question of uptake rather than treat its original automation category as conclusive.

5. Stress Tests and Objections

The following objections test whether reason-sensitive uptake is a defensible institutional duty rather than an unconditional right to individualized attention.

- **Oracle or perfect-prediction objection.** Accuracy concerns an outcome or forecast, not whether the institution has discharged independent duties of justification. A perfectly predicted event does not entail that affected persons can be excluded when contestable reasons are relevant.
- **Manipulation objection.** A right to submit reasons does not guarantee credibility or success. Fair corroboration, documented relevance rules, and proportionate sanctions for fraud are preferable to closing the process to everyone.
- **Expertise objection.** Affected persons are not sovereign over technical conclusions. Their standing is strongest for first-person facts, circumstances, values, and consent, while specialized claims remain open to expert assessment.
- **Cost and workload objection.** Triage by seriousness and material new evidence, collective representation, and standardized corrections can control cost. Workload alone does not justify eliminating remedy for severe, contestable decisions.
- **Unequal capacity objection.** Formal appeal rights can advantage articulate or well-resourced claimants. Translation, supported submissions, disability accommodations, and audits of access disparities are necessary (Young, 2000; Medina, 2013).
- **Privacy and third-party rights objection.** Uptake need not disclose protected material. Summaries, independent reviewers, and redacted evidence can preserve responsive consideration while respecting others' rights.
- **Merely psychological voice objection.** Perceived fairness matters (Tyler, 2006; Langer et al., 2019), but this principle rests on accountable reason-giving, not claimant satisfaction. A respectful hearing that cannot change any reason is inadequate, and a justified rejection is permissible.

These replies preserve a demanding but limited principle: the legitimacy of an appeal depends on relevant reasons being genuinely considered under accountable authority, not on automatic deference to claimants. Administrative design must protect both applicants and people competing for scarce opportunities.

6. Applications Across Social Institutions

The framework concerns a structural feature of institutional authority rather than a particular industry. In public administration, a risk score may appropriately trigger investigation, but cannot replace assessment of a citizen's relevant account in a contestable eligibility decision. A standing audit must distinguish a correction to factual input, a challenge to an inference, and an objection to the policy that transforms a score into a benefit or sanction. These are different sorts of reasons, and each may require a different authorized reviewer.

In employment, predictive screening may undervalue nonstandard qualifications, interrupted careers, or work acquired outside conventional credentials. Studies of automated interviews address stakeholder acceptance and perceptions under different stakes (Langer et al., 2019), but they do not establish that every applicant is entitled to a personal interview. The narrower

issue is whether a consequential selection process arbitrarily closes a feasible route for material, relevant context to reach a responsible decision-maker.

In education, dropout prediction may sensibly prompt assistance. It becomes normatively troubling when a risk category conclusively determines access to support or admission without a mechanism for considering temporary hardship, caregiving, or changed goals. Comparable failures may arise in housing allocations, professional licensing, or credit appeals when categorical records are treated as sufficient despite reasonable individualized challenges. Each domain has different authority, confidentiality, and redress obligations; none establishes a blanket entitlement to override objective evidence.

Academic publishing illustrates the distinction between epistemic assistance and justificatory closure. Automated peer review can flag inconsistencies, check reporting, and improve editors' triage. A substitution error would arise in a consequential editorial decision if a machine-generated characterization became a final verdict that authors could not correct when it misread an argument or methodological design. Independent editorial responsibility and a reason-sensitive response pathway matter more than whether any particular reviewer was human or computational.

These are diagnostic examples, not allegations that specific journals, welfare agencies, employers, or platforms act this way. Technical audits remain essential (Kroll et al., 2017; Raji et al., 2020), but uptake audits ask a further question: can the person affected by a representation introduce a reason that an authorized institutional actor must actually assess?

7. Operationalization and Falsifiable Research Agenda

The primary empirical unit is a decision episode, from initial classification through communication, possible challenge, and final justification. Researchers should record the availability and use of the five conditions, whether submitted reasons were evaluated, the competence of the reviewer to grant remedy, case outcomes, and differential procedural burdens. Measured procedural compliance must be distinguished from perceived legitimacy; neither is equivalent to prediction accuracy.

7.1 Worked Hypothetical Scenarios

Scenario A, social assistance. An agency uses a validated earnings model to flag a household as ineligible for a temporary benefit. The claimant submits evidence that the apparent earnings belong to a former household member who moved out before the relevant assessment period. If the application permits attachment of this evidence but the reviewer must deny any appeal inconsistent with the original score, access exists while admissibility and revisability fail. A compliant pathway records the new evidence, tests it under published household-composition rules, gives a case-specific response, and empowers a reviewer to correct the determination where warranted. The model may remain statistically excellent throughout. This hypothetical illustrates why a submission portal, standing alone, is not adequate uptake.

Scenario B, professional credential review. An applicant is denied registration because automated extraction fails to recognize supervised practice completed under a renamed institution. The applicant provides a contemporaneous registration record. Admissibility requires that the record not be discarded merely because its label differs from training categories; traceability requires preserving the submission and the reviewer's reason for accepting or discounting it. If a professionally accountable panel can amend the decision after verification, the five conditions may be met even if the appeal ultimately fails for another legitimate requirement. This case shows that meaningful participation does not imply preferential treatment.

Both scenarios are deliberately constructed; no administrative outcome or effectiveness rate is claimed. They yield straightforward audit probes: Was the relevant claim heard before the decision became materially irreversible? Did a reviewer evaluate its substance? Could an authorized actor amend the record or remedy the result? A yes to data-submission access alone is insufficient.

7.2 A Boundary Case: Consequential Decisions Under a Genuinely Fixed Rule

Scenario C, a time-limited educational bursary. Assume a published, nondiscretionary rule requires a completed application by a fixed date and that the applicant concedes the deadline and the accuracy of the submitted record. Losing the bursary may be consequential, but the administrative reviewer has no lawful discretion to disregard the agreed rule. Epistemic standing does not automatically include a right to rewrite that rule or to obtain an exception unavailable to other applicants. The institution should nevertheless communicate the operative criterion, admit evidence of clerical error or unequal rule application if alleged, respond to any such material challenge, and identify an authorized policy-review channel where one exists. If the application is undisputedly late under a genuinely fixed rule, a reasoned refusal need not constitute substitution. If the institution instead falsely treats contestable evidence about filing time as irrelevant merely because its model labels the file late, the diagnosis changes. This paired case guards against conflating severe consequences with unlimited adjudicative discretion.

Across the three scenarios, the diagnostic turns on counterfactual institutional capacity rather than claimant success: would a credible and decision-relevant submission be examined by someone empowered to make a legally or institutionally permitted correction? A process may fail even when its model is accurate; it may pass even when a claim is ultimately rejected. Researchers must therefore separately code the availability of a challenge, the materiality of its content, the reviewer's authority, and the reasons given for non-revision.

7.3 Empirical Evaluation Strategy

A preregistered vignette experiment could hold benefit-screening accuracy and final decision constant while manipulating four interfaces: no challenge, an explanation only, evidence submission with no answer, and evidence submission with authorized reason-sensitive review. A separate manipulation should vary decision severity. The expectation is that genuine uptake improves procedural legitimacy judgments beyond what explanation alone contributes, although this is a hypothesis rather than a demonstrated effect.

To isolate the mechanism, researchers should independently manipulate outcome favorability, accuracy, politeness, access, and reviewers' power to revise. Otherwise apparent uptake effects could reflect improved outcomes or warmer language. Samples spanning institutional and cultural settings, preregistered exclusion criteria, and tests for unequal access would be needed before generalizing. Recent work relating technical fairness, voice, and algorithmic legitimacy underscores why these constructs should not be conflated (Kinchin, 2024; Perdana, 2026).

Qualitative process tracing could examine de-identified appeal records or authorized interviews, coding whether a challenge introduced a material new reason, whether a reviewer engaged it, what authority the reviewer held, and whether the result or the recorded justification changed. Two independent coders, adjudication rules, and negative-case analysis would strengthen validity. Neither high nor low reversal rates alone establish legitimate uptake.

A discriminant-validity test should include negative controls: incorrect predictions with effective review, accurate predictions with exclusion of relevant reasons, and adverse but fully rule-governed decisions after responsive hearing. The first is a model-performance problem without necessary substitution; the second is the theory's focal case; the third can be procedurally legitimate despite an unwelcome outcome. Independent coders should assess the facts and institutional powers before learning the study authors' preferred classification. If the framework labels all adverse automated decisions as substitution, or fails to distinguish recorded consideration from mere receipt of evidence, its operational specification is too broad and should be revised.

A process audit should examine the five conditions separately from calibration, error parity, and model documentation. For high-stakes cases, evidence of a complete failure of revisability should not be averaged away by high scores on other dimensions. Table 3 translates four empirical hypotheses into proposed study designs; it is not a validated index.

Four researchable hypotheses follow. First, reason-sensitive review increases perceived procedural legitimacy after holding predictive performance fixed. Second, explanation alone produces weaker gains than meaningful response and revisability. Third, supported submission channels reduce procedural disparities relative to unsupported appeals. Fourth, the effect of better prediction on exclusion is moderated by the availability of uptake pathways. These behavioral claims may be empirically false even if the normative non-substitution argument survives; theory and observation therefore remain separately revisable.

Table 3. Empirical hypotheses and proposed tests

Hypothesis	Predicted comparison	Recommended evidence
H1	Genuine uptake vs. no uptake, holding accuracy constant	Factorial vignette experiment
H2	Responsive review vs. explanation-only interface	Randomized interface experiment

Hypothesis	Predicted comparison	Recommended evidence
H3	Assisted access vs. unassisted access among vulnerable claimants	Field evaluation with equity measures
H4	Model improvements with vs. without standing safeguards	Longitudinal institutional process audit

Note. These are proposed tests; the paper reports no original empirical measurements.

8. Discussion: Legitimacy Beyond Instrumental Optimization

The central theoretical move is to reject a hidden substitution of questions: from “Is the institution justified in deciding this way?” to “Does the prediction perform well?” Predictive performance can be a powerful reason in favor of using a tool. It cannot by itself settle whether the institution is entitled to close its justificatory process against people who bear its consequences. The difference is not anti-scientific. It recognizes that questions about authority are partly about which evidence should count, how uncertainty is distributed, and who may challenge the public use of power.

The originality boundary requires care. Kinchin (2024) already establishes the salience of voice in algorithmic procedural justice. Perdana (2026) maps correctability alongside substantive fairness, while Ibitoye et al. (2026) theorize the restructuring of human judgment and Wong (2026) defends ownership of public judgments by accountable offices. This article does not claim that institutional voice, the gap between technical and procedural fairness, or a right to contest algorithms is new. Its proposed addition is the narrow substitution-error diagnosis, independent of predictive error, together with conjunctive uptake conditions and a contextual exception test focused on person-supplied reasons. Whether that analytical specification is useful beyond established correctability theory is an open research question.

The error also differs from testimonial injustice (Fricker, 2007). Identity prejudice is not required: a neutral algorithmic policy that categorically refuses all case-specific evidence may commit substitution, whereas prejudicial disbelief can occur within a procedure that is, in other respects, open to evidence. Similarly, not every procedural injustice is substitution; arbitrary delay, unequal fees, or political exclusion may be wrong without a proxy replacing someone’s standing as a contributor to reasons. The relations are ones of possible overlap, not conceptual identity.

Nor does a right to explanation necessarily secure uptake. Explanations are directed from institution to subject; the error arises when the reverse channel is closed to potentially relevant reasons. Miller (2019) emphasizes the social nature of explanation and Vredenburg (2022) develops its normative significance. The present standard asks, additionally, whether the institution is capable of changing its reasons in light of a legitimate submission. A responsive rejection can pass that test. An informative but unreviewable decision can fail it.

The framework also changes the meaning of human oversight. A person who rubber-stamps machine outputs adds little normative value. What matters is whether an appropriately

empowered and accountable actor can receive relevant evidence, reconsider governing standards when authorized, and record why a challenge succeeds or fails. The reviewer need not be capable of intuitively simulating every computation. The institutional capacity for correction is conceptually distinct from the personal ability to reproduce a model's internal calculations. This is consistent with broader critiques of transparency as an exclusive governance ideal (Ananny & Crawford, 2018; Creel, 2020).

Accountable institutional listening must not romanticize testimony. Administrators also owe consistency, confidentiality, stewardship, and fairness to competing claimants. They can reject unsupported claims under public standards and limit participation for proportionate reasons. The threshold and exception test is therefore constitutive of, not ancillary to, the non-substitution principle. Where relevant context is absent and a low-stakes rule is fixed, individualized hearings may add administrative cost without increasing legitimacy.

The article's contribution is thus a mid-level normative architecture rather than a universal legal doctrine. It identifies a kind of illegitimate exclusion that existing technical measures may miss, states the circumstances in which that exclusion matters, and derives procedural tests that empirical researchers can examine. It also creates a critical standard for otherwise impressive algorithmic institutions: a system may be accurate, equitable on reported metrics, and transparent in operation, yet still fail to treat affected persons as participants in the justification of decisions governing their lives.

9. Limitations

Three limitations are important. First, this is a conceptual argument supported by literature and constructed cases, not an empirical estimate of institutional prevalence or policy effectiveness. Its operational hypotheses require preregistered testing and engagement with actual organizational constraints. Second, the boundary between contestable and exhaustively codifiable decisions is itself socially contested; different legal systems and professions assign different procedural entitlements. The proposed normative thresholds therefore require contextual interpretation rather than automatic conversion into a compliance score.

Third, the theory says relatively little about conflicts among claimants with incompatible entitlements or scarce resources. Hearing every applicant may improve the legitimacy of the process without resolving distributive injustice. Conversely, public participation may sometimes impede urgent interventions. The framework provides a rebuttable presumption in favor of reason-sensitive uptake, not an absolute lexical priority over all other values. Comparative legal research, participatory design, and field studies could clarify how the five conditions should be institutionalized across jurisdictions and sectors.

A further limitation concerns novelty and moral grounding. Prior work already establishes that voice, contestation, and corrective review can matter. The proposed diagnostic is a structured synthesis, not the invention of procedural justice. Its most difficult normative assumption is that particular exercises of institutional power generate independent entitlements to reason-sensitive participation. Jurisdictions and moral traditions may disagree about that entitlement, especially in private contractual decisions. Comparative institutional

ethics and empirical studies of borderline cases could test the framework's practical discrimination without pretending that empirical popularity alone proves its moral premises.

A further boundary concerns competent authority and second-order challenges. A first-line reviewer may be entitled to correct a disputed factual input but not to alter a legislatively or contractually binding eligibility criterion. The framework should not mistake that limited competence for an automatic substitution error, provided the limits and any available escalation route are candidly stated. Whether a decision rule itself is substantively unjust remains a separate normative question. This division between first-order factual challenge, discretionary interpretation, and policy-level contestation will require further institutional and comparative legal analysis.

10. Conclusion

An institution need not predict wrongly in order to judge wrongly. Accuracy answers whether a forecast tracks its target; procedural legitimacy also asks whether those subject to institutional authority retain an appropriate standing to contribute and challenge reasons. The epistemic substitution error occurs when a proxy's account of a person is allowed to replace that standing without sufficient justification. Its core pathology is not the use of artificial intelligence, but the conversion of instrumental epistemic success into unwarranted decisional closure.

The non-substitution principle developed here requires access, admissibility, responsiveness, traceability, and revisability where institutional power, consequential stakes, and contestable reasons warrant them. Exceptions remain possible, particularly for routine low-stakes transactions and urgent decisions accompanied by effective later review. The goal is not maximal participation in every automated transaction. It is a defensible allocation of reason-giving authority so that better prediction can improve institutions without making those institutions unanswerable to the people they govern.

The broader philosophical implication is that knowing about someone and being answerable to that person are different relationships. Human beings can be objects of reliable classification and still be neglected as participants in justification. A legitimate algorithmic institution must therefore preserve not only the capacity to reach good conclusions, but also the capacity to learn why a conclusion may need to be reconsidered.

Ethics Approval

No human participants were recruited, no identifiable personal data were collected, and no intervention was conducted. Formal research ethics approval was therefore not required for this conceptual philosophical study.

Informed Consent

Not applicable. This theoretical study did not involve human participants or identifiable personal information.

Funding

None.

Conflict of Interest

The author declares no conflicts of interest.

Data Availability

No empirical dataset was generated or analysed. The argument is based on published literature and explicitly identified hypothetical cases; bibliographic sources are provided in the reference list.

Author Contributions

Sole author: conceptualization, methodology, validation, writing, review and editing, and final approval. Generative artificial intelligence assistance is disclosed above.

References

- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989. <https://doi.org/10.1177/1461444816676645>
- Anderson, E. (2012). Epistemic justice as a virtue of social institutions. *Social Epistemology*, 26(2), 163–173. <https://doi.org/10.1080/02691728.2011.652211>
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732. <https://doi.org/10.15779/Z38BG31>
- Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of Machine Learning Research*, 81, 149–159. <https://proceedings.mlr.press/v81/binns18a.html>
- Burrell, J. (2016). How the machine “thinks”: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1–12. <https://doi.org/10.1177/2053951715622512>
- Creel, K. (2020). Transparency in complex computational systems. *Philosophy of Science*, 87(4), 568–589. <https://doi.org/10.1086/709729>
- Dotson, K. (2011). Tracking epistemic violence, tracking practices of silencing. *Hypatia*, 26(2), 236–257. <https://doi.org/10.1111/j.1527-2001.2011.01177.x>
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198237907.001.0001>
- Ibitoye, A. O., Ajenifuja, O. F., Abdul Kareem, R. S., Adeyemo, D. D., & Famutimi, R. F. (2026). Decision substitution and the reorganisation of judgement in AI-mediated society. *AI & Society*. <https://doi.org/10.1007/s00146-026-03349-3>

- Kinchin, N. (2024). “Voiceless”: The procedural gap in algorithmic justice. *International Journal of Law and Information Technology*, 32, eaae024. <https://doi.org/10.1093/ijlit/eaae024>
- Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165(3), 633–705. <https://pennlawreview.com/2017/02/23/accountable-algorithms/>
- Langer, M., König, C. J., & Papathanasiou, M. (2019). Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment*, 27(3), 217–234. <https://doi.org/10.1111/ijsa.12246>
- Medina, J. (2013). *The epistemology of resistance: Gender and racial oppression, epistemic injustice, and the social imagination*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199929023.001.0001>
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Perdana, A. (2026). Technical fairness and social legitimacy: A conceptual mapping of algorithmic fairness, procedural justice, and explainable AI. *AI and Ethics*, 6, Article 491. <https://doi.org/10.1007/s43681-026-01342-6>
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372873>
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68). Association for Computing Machinery. <https://doi.org/10.1145/3287560.3287598>
- Tan, K. H. (2026). Human agency under algorithmic governance: A humanities framework for law, management and public life. *International Journal of Law Management & Humanities*, 9(4), 559–571. <https://doi.org/10.63108/IJLMH.12537>
- Theoharakis, V., & Mylonopoulos, N. (2026). AI epistemic disengagement and consumer dependence: An augmentation-substitution framework. *Psychology & Marketing*, 43(7), 1675–1686. <https://doi.org/10.1002/mar.70136>
- Tyler, T. R. (2006). *Why people obey the law*. Princeton University Press. <https://doi.org/10.2307/j.ctv1j66769>
- Vredenburg, K. (2022). The right to explanation. *Journal of Political Philosophy*, 30(2), 209–229. <https://doi.org/10.1111/jopp.12262>
- Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law &*

Technology, 31(2), 841–887.

<https://jolt.law.harvard.edu/assets/articlePDFs/v31/Counterfactual-Explanations-without-Opening-the-Black-Box-Sandra-Wachter-et-al.pdf>

Wong, S.-M. (2026). Office and unowned architecture: Public judgment in the age of algorithmic governance. *Ethics and Information Technology*, 28(4), Article 48.
<https://doi.org/10.1007/s10676-026-09919-9>

Young, I. M. (2000). *Inclusion and democracy*. Oxford University Press.
<https://doi.org/10.1093/0198297556.001.0001>