Contents lists available at International Journal Press

Journal of Digital Justice and Legal Innovation

journal homepage: https://ijpress.com/jdjli

Volume 1 · Issue 1 · 2026 · pp. 31-41

Interpretive Debt: How Organisations Accumulate Hidden Liabilities When Humans Defer Sensemaking to Automated Systems

Dr. Deepika Rani[1,*] Prem Lata[2]

[1] Advocate and Researcher, Supreme Court of India, New Delhi, India

[2] Research Scholar, Mahatma Jyotiba Phule Rohilkhand University (State Government University), Bareilly, U.P. India

* Corresponding author: drdeepikaraniadvocate@gmail.com

ARTICLE INFO

Article history:

Received: 23-7-2026

Received in revised form: 03-8-2026

Accepted: 04-8-2026

Available online: 05-08-2026

Keywords:

Freedom of expression

Informational privacy

Open justice

Right to be forgotten

ABSTRACT

This paper develops the concept of *interpretive debt*: a latent organisational liability that accumulates when humans defer sensemaking to automated systems (dashboards, recommender systems, risk scores, and AI assistants) embedded in routine decision-making. Drawing on the technical-debt metaphor, Weickean organisational sensemaking, the human-factors tradition on automation, and the latent-conditions paradigm in safety science, I argue that organisations face a structural choice between performing interpretive work and outsourcing it. Each deferral is locally rational and immediately efficient, yet erodes the collective interpretive capacity required to notice when systems are wrong, to socialise newcomers into genuine understanding, and to reconstruct the rationale behind past decisions. Interpretive debt remains invisible until a *repayment event* (an audit, crisis, system failure, or generational turnover) forces it due at once. The paper differentiates interpretive debt from adjacent constructs (cognitive debt, epistemic debt, automation bias), specifies three accumulation mechanisms (deferral, opacity, atrophy), proposes three operationalisations, and outlines a research agenda. The contribution is to name an organisational asset that is currently nameless (interpretive capacity) and to integrate four siloed literatures around its erosion.

1. INTRODUCTION

A compliance team at a mid-sized lender is asked, three years after the fact, to reconstruct why a particular loan was declined. The decision was made by a credit-risk model that has since been retrained twice. The original feature engineering is documented in a wiki page last edited by an analyst who has left the firm. The committee that signed off on the model retired the policy memo when the next version was released. The team can produce the *score* the model assigned; what it cannot produce is the *interpretation* the score once carried: what counted as a borderline case, what considerations a human reviewer would have raised, what the score meant in context. The decision was reached efficiently. The justification has evaporated.

This vignette describes a pattern that, I argue, is widespread but undertheorised. Across sectors (finance, healthcare, logistics, content moderation, scientific research, and increasingly white-collar work of every variety) organisations are weaving automated decision-support systems into the routine grain of

© 2026 International Journal Press · Licensed under CC BY 4.0

operations. Each deferral to a dashboard, a recommender, a risk score, or an AI assistant is locally rational: the system is faster, more consistent, and often more accurate than the human alternative. Yet the cumulative effect of those deferrals is rarely modelled. Something is being given up. The question this paper takes up is: *what*, exactly, is given up, and how should socio-technical scholarship name and analyse it?

My answer is that organisations are accumulating *interpretive debt*. By this I mean a latent collective liability that builds up when humans repeatedly defer sensemaking work to automated systems. The deferral is the proximate transaction; the accumulating stock is a depletion of the organisation's capacity to perform sensemaking when it is required: to notice that an automated output is wrong, to reconstruct why an earlier decision was made, to bring newcomers into genuine understanding rather than mere interface competence, and to render the organisation's reasoning legible to itself and to outsiders. Interpretive debt is *latent*: it is invisible during the long periods of routine operation when systems behave as expected. It becomes visible (sharply and expensively) at *repayment events*, when an audit, a regulatory inquiry, a crisis, a system failure, or a generational turnover forces the organisation to perform interpretation it has not been performing.

The paper makes three contributions. First, it names an organisational asset that the current vocabulary leaves nameless: *interpretive capacity*, understood as the ongoing socio-cognitive practice through which organisational members make sense of operational reality. Second, it offers a debt-structure for analysing this asset's erosion, importing the analytical traction of Cunningham technical-debt metaphor while differentiating the new construct from its closest neighbours: Sculley et al. hidden technical debt in machine-learning systems and the very recent coinages of "cognitive debt" and "epistemic debt" [1, 2, 3, 4, 5]. Third, it integrates four bodies of scholarship (debt metaphors, Weickean sensemaking, automation human factors, and latent-conditions safety theory) that have until now developed largely in isolation.

The paper proceeds as follows. Section 2 reviews the four scaffolding literatures and positions interpretive debt against adjacent constructs. Section 3 defines the construct formally and specifies its locus, object, and structure. Section 4 specifies three mechanisms by which interpretive debt accumulates (deferral, opacity, and atrophy) and develops the repayment-event mechanism that makes the construct distinctive. Section 5 turns to implications: three operationalisation paths, mitigation strategies derived from Zuboff automate/informate distinction and from High-Reliability Organisation theory, and a research agenda [6, 7]. Section 6 acknowledges boundary conditions and limitations. Section 7 concludes.

2. THEORETICAL FOUNDATIONS

2.1 The technical-debt metaphor and its extensions

The metaphor originates with Cunningham, who used it to describe the trade-off in shipping imperfect code: "shipping first-time code is like going into debt …the danger occurs when the debt is not repaid" [1]. Fowler extended the metaphor with a 2×2 quadrant (deliberate/inadvertent × prudent/reckless) that has become the dominant taxonomy in software engineering practice [8]. The metaphor's analytic power lies in four features: (a) the deferral is *locally rational*, often deliberately so; (b) the deferred work *accrues interest*, compounding the cost of later remediation; (c) repayment requires *deliberate effort*, not the passage of time; and (d) under-servicing produces *systemic*, not merely local, dysfunction. These features are what makes "debt" more than a colourful synonym for "problem."

Sculley et al. "Hidden Technical Debt in Machine Learning Systems" was the first major extension of the metaphor into algorithmic territory [2]. Their argument is that machine-learning systems incur ML-specific debts (boundary erosion, entanglement (encapsulated in the "Changing Anything Changes Everything" principle), undeclared consumers, hidden feedback loops, data dependencies, and changes in the external world) and that "it is dangerous to think of these quick wins as coming for free" [2, p. 2503]. Sculley et al. locate the debt in code, data pipelines, and configurations. The construct developed in this paper relocates the debt to *human and organisational* interpretive capacity, while preserving the four analytic features of Cunningham's original.

Three further extensions appearing in 2025–2026 sit close enough to interpretive debt that explicit differentiation is required. Sankaranarayanan proposes *epistemic debt* as the cognitive cost of "outsourcing the intrinsic cognitive load required for schema formation" in generative-AI-scaffolded novice programmers, producing "fragile experts [5]." Storey defines *cognitive debt* as a team-level property ("the erosion of shared understanding across a software system over time") and locates it specifically in "people" alongside technical debt in code and intent debt in externalised knowledge [4]. Kosmyna et al. provide an electroencephalographic study popularising "cognitive debt" at the individual neural level: large-language-

© 2026 International Journal Press · Licensed under CC BY 4.0

model-assisted essay writers exhibited the weakest network connectivity, with carry-over effects on subsequent unaided writing [3]. Critiques of debt metaphors apply to all extensions, including this one. Dudycz argues that the metaphor "labels the absence of tests as technical debt, [and so] we postpone the real issue": the metaphor is communicative, not actuarial [9]. I adopt this stance: interpretive debt is a diagnostic frame, not a literal balance sheet.

Table 1 summarises the construct's position in relation to its neighbours.

Table 1. *Differentiation of interpretive debt from adjacent constructs*

Construct	Origin	Locus	Object	Repayment mechanism
Technical debt	Cunningham [1]	Codebase	Code quality	Refactoring
Hidden ML technical debt	Sculley et al. [2]	ML pipeline	Models, data, configs	Pipeline rebuild
Epistemic debt	Sankaranarayanan [5]	Individual novice learner	Schema formation	Metacognitive scaffolding
Cognitive debt (team)	Storey [4]	Developer team	Shared mental model of code	Shared comprehension work
Cognitive debt (neural)	Kosmyna et al. [3]	Individual brain	Neural engagement	Unaided practice
Automation bias	Mosier et al. [10]	Decision-maker	Vigilance	(Not theorised as debt)
Interpretive debt	This paper	Organisational sensemaking capacity	Interpretation of operational decisions	Forced by audit, crisis, failure, or turnover

2.2 Sensemaking as the depletable asset

The second pillar is Weick's program on organisational sensemaking. Weick defines sensemaking as the ongoing accomplishment through which people "structure the unknown" by extracting cues, forming plausible accounts, and acting back into the environment [11]. Sensemaking is grounded in identity, retrospective, enactive, social, ongoing, focused on extracted cues, and driven by plausibility rather than accuracy [11]. The collapse of sensemaking in the Mann Gulch fire showed that when a group's interpretive frame falls apart, structure and identity collapse together; the Tenerife air disaster made the same point for tightly coupled operational settings [12, 13]. Maitlis and Christianson consolidated three decades of subsequent work, defining sensemaking as "the process through which people work to understand issues or events that are novel, ambiguous, confusing, or in some other way violate expectations" (p. 57), and showed how related constructs (sensegiving, sensebreaking, and prospective sensemaking) extend the basic apparatus [14, 15, 16, 17].

What makes this scaffolding indispensable for interpretive debt is Weick earlier insight that new technologies present "equivocal" stimuli requiring ongoing sensemaking [18]. Sensemaking is performed, not stored. It is a practice, not a stock. Every moment in which interpretation could have been performed and was not is a moment in which the organisation has *failed to exercise the practice that produces its interpretive capacity*. From this it follows that interpretive capacity, like any practised skill, is depletable through disuse, and from this in turn it follows that one can theorise its erosion as a debt.

2.3 Automation, complacency, and the substitution myth

The third pillar is the human-factors literature on automation. Bainbridge identified two ironies: the designer who treats the operator as unreliable and seeks to eliminate them ends up leaving the operator the *hardest* residual tasks, and automation simultaneously *requires* high skill (for rare crucial interventions) while *destroying* it (because operators no longer practise) [19]. Christoffersen and Woods generalised this as the "substitution myth": automation does not substitute for, but transforms, human roles [20]. Parasuraman and Riley classified the resulting pathologies (use, misuse, disuse, abuse); Parasuraman and Manzey integrated the evidence into a unified model of complacency and bias, finding that "automation complacency is found in both naive and expert participants and cannot be overcome with simple practice" [21, 22]. Mosier et al. showed in high-tech cockpits that automation cues are used "as a heuristic replacement for vigilant information seeking and processing": that is, deferral is the very mechanism that produces interpretive atrophy [10]. Endsley "automation conundrum" captures the dynamic in situation-awareness language: as automation becomes more reliable, situation awareness declines, leaving humans less able to take over when needed [23].

© 2026 International Journal Press · Licensed under CC BY 4.0

The construct developed here generalises this tradition beyond safety-critical control rooms to the wider socio-technical fabric of organisational decision-making: to dashboards, recommender systems, risk scores, and AI assistants used in routine, non-safety-critical settings.

2.4 Latent conditions and the repayment-event mechanism

The fourth pillar supplies the *latency* mechanism that gives interpretive debt its distinctive shape. Reason's [24] "resident pathogens" metaphor, restated in his 2000 BMJ paper [25] as "latent conditions" (which "may lie dormant within the system for many years before they combine with active failures and local triggers to create an accident opportunity" [25]) establishes the analytical structure adopted here. Perrow had earlier argued that complex, tightly coupled systems incubate normal accidents: failures that emerge from the system's structure rather than from any single broken component [26]. Vaughan documented how the normalisation of deviance produced the Challenger launch decision: each individual judgement was locally rational, each accommodation defensible in context, yet the cumulative trajectory was catastrophic [27]. Dekker restated this as "drift into failure": failure that emerges "opportunistically, non-randomly, from the very webs of relationships that breed success [28]." Weick and Sutcliffe showed how High-Reliability Organisations resist this drift through five mindfulness practices (preoccupation with failure, reluctance to simplify, sensitivity to operations, commitment to resilience, deference to expertise) [7].

The latent-conditions literature gives interpretive debt the property that distinguishes it from cognitive debt, epistemic debt, and automation bias: interpretive debt is *not directly observable in routine operations*. It manifests only when an external trigger (what I will call a repayment event) forces the organisation to perform interpretation it has not been performing. Bridging the safety-science vocabulary to the wider socio-technical setting allows the latent-conditions logic to be extended from catastrophic accidents in safety-critical industries to ordinary organisations facing audits, regulatory inquiries, customer escalations, model retirements, and staff turnover.

3. DEFINING INTERPRETIVE DEBT

Drawing the four pillars together, I propose the following working definition:

> *Interpretive debt is a latent organisational liability that accumulates when human members of an organisation systematically defer sensemaking work to automated decision-support systems, eroding the collective capacity to (i) recognise when those systems are wrong, (ii) reconstruct the rationale for prior decisions, and (iii) socialise new members into substantive understanding rather than mere interface competence. The debt becomes manifest only at repayment events, when the organisation is required to perform interpretation it has not been practising.*

Three features deserve emphasis.

Locus. Interpretive debt resides in the organisation's collective interpretive capacity: distributed across individuals, artefacts, documentation, training practices, and the conversational routines through which operational decisions are framed and contested. This locus differs from technical debt (which resides in code), Sculley et al. hidden ML debt (which resides in pipelines and data dependencies), Storey cognitive debt (which resides in developer teams' mental models of their own code), Sankaranarayanan epistemic debt (which resides in individual learners' schemas), and Kosmyna et al. cognitive debt (which resides in individual neural engagement) [2, 4, 5, 3]. It is closest in spirit to Walsh and Ungson organisational memory, but where Walsh and Ungson identified five storage bins (individuals, culture, transformations, structures, ecology), interpretive debt names the *practice* through which those stored resources are activated to make sense of new situations [29]. Where organisational memory is a stock, interpretive capacity is the practised flow through which the stock is rendered usable.

Object. The object of interpretation is the operational decision: the loan approval, the triage recommendation, the content moderation call, the resource allocation, the diagnostic flag, the strategic forecast. What is at stake is not whether the organisation can *produce* a decision (automated systems will produce decisions reliably) but whether the organisation can produce a *reasoned account* of the decision when asked, and whether members can recognise when a decision is wrong. The relevant capacity, in Polanyi sense, includes tacit knowledge that "we can know more than we can tell" (p. 4), which makes interpretive capacity especially fragile under delegation: tacit practices that are not exercised cannot be externalised in Nonaka and Takeuchi sense, and therefore cannot be re-socialised to newcomers [30, 31].

© 2026 International Journal Press · Licensed under CC BY 4.0

Structure. Interpretive debt preserves the four analytic features of Cunningham's original debt metaphor. Deferrals are *locally rational*: the system is faster and often more accurate. The debt *accrues interest*: each unexercised interpretive practice makes the next interpretive demand harder to meet, because the supporting tacit knowledge thins. Repayment requires *deliberate effort*: the debt cannot be paid down by waiting, only by re-exercising the practice (or by Zuboff informating-style interventions that re-render the system's reasoning legible) [6]. Under-servicing produces *systemic* dysfunction: the manifestation is not local error correction but the collapse of the organisation's capacity to give account of itself.

It is useful to distinguish interpretive debt from three near neighbours with which it might be conflated. Against *automation bias*: automation bias names a moment-to-moment cognitive tendency at the individual level, whereas interpretive debt names an accumulating organisational property whose effects compound over time [10, 22]. Against *deskilling* in the Bravermanian tradition: deskilling is a labour-process claim about the systematic transfer of craft knowledge into machinery and managerial procedures; interpretive debt is an interpretive-capacity claim, applicable equally to white-collar professionals managing AI advisors and to operatives managing physical machinery [32]. Against *cognitive debt*: these constructs locate the erosion in mental models or neural engagement; interpretive debt locates it in the collective sensemaking practice of the organisation, and crucially incorporates the latency/repayment structure absent from the cognitive-debt formulations [3, 4].

4. MECHANISMS OF ACCUMULATION AND REPAYMENT

I specify three mechanisms by which interpretive debt accumulates, and a fourth mechanism by which it is forced due. Figure 1 represents the cycle schematically.

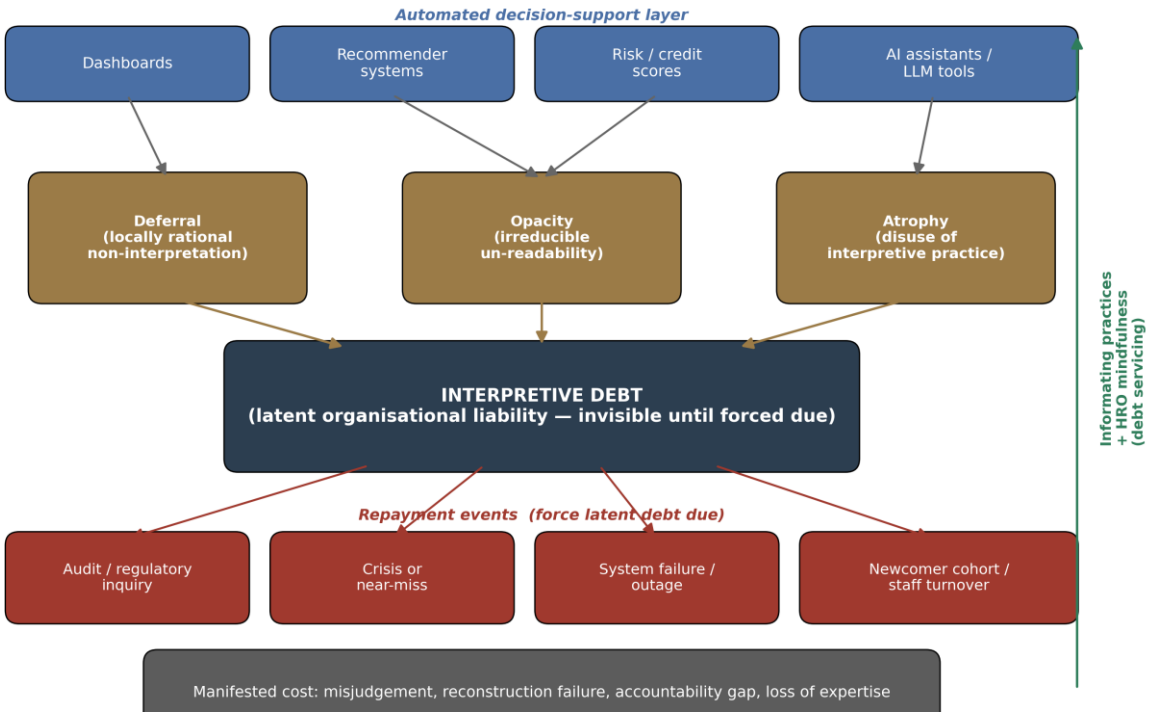

Figure 1. The interpretive debt cycle

Automated decision-support systems become the targets of organisational deferral. Three accumulation mechanisms (deferral, opacity, and atrophy) feed a latent stock of interpretive debt that remains invisible during routine operations. Repayment events (audit, crisis, system failure, generational turnover) force the debt due, manifesting as misjudgement, reconstruction failure, accountability gaps, and loss of organisational expertise. Informating practices combined with High-Reliability Organisation mindfulness constitute the servicing mechanism by which the debt can be paid down [6, 7].

© 2026 International Journal Press · Licensed under CC BY 4.0

4.1 Deferral: locally rational non-interpretation

The first mechanism is the deferral itself. Each time an organisational actor consults a dashboard, accepts a recommender's ranking, applies a risk score, or asks an AI assistant for a draft, an act of interpretation that could have been performed is not performed. The substitution is rarely complete (the actor still reads the dashboard, still glances at the ranking, still skims the draft) but the *generative* interpretive work is offloaded. The cognitive-offloading literature frames this as the use of external action to "alter the information processing requirements of a task so as to reduce cognitive demand" [33]. Sparrow et al. "Google effects on memory" study showed that when people expect future access to information, they remember *where to find it* rather than *what it is*: a microcosm of the broader pattern in which the locus of interpretation shifts from the human to the system [34]. Each deferral is locally rational and often productivity-enhancing. The cumulative effect is the slow draining of the practice through which interpretive capacity is maintained.

4.2 Opacity: irreducible un-readability

The second mechanism is opacity, which differs from deferral in that the deferring actor *could not* have performed the interpretation even if they had wanted to. Burrell distinguishes three forms of opacity: intentional corporate or state secrecy, technical illiteracy (most users cannot read code), and the opacity intrinsic to high-dimensional machine-learning representations [35]. The third form is the most consequential for interpretive debt because it is irreducible by openness or by code literacy: even with full source-code access and statistical fluency, no human reads a 175-billion-parameter model the way an auditor reads a ledger. Ananny and Crawford argue persuasively that transparency-as-accountability has hard limits ("seeing without knowing") and that algorithmic accountability requires alternative, networked forms of governance [36]. Pasquale characterised the resulting environment as a "Black Box Society" in which consequential decisions are shrouded by technical opacity and legal trade-secrecy [37]. The regulatory response, including the EU GDPR provisions sometimes glossed as a "right to explanation," has been argued to fall short [38]. Governance scholarship on algorithmic agents similarly foregrounds the overt and covert challenges that automation poses to justice and the future of work [39]. Opacity thus compounds the deferral mechanism: even an organisation that wishes to re-perform interpretation may find that the relevant inputs and reasoning are technically inaccessible.

4.3 Atrophy: disuse of interpretive practice

The third mechanism is atrophy. Interpretive practice, like any practice, decays through disuse. Bainbridge anticipated this point for control-room operators; Casner et al. documented its empirical manifestation in commercial pilots, finding that while basic instrument-scanning skills are reasonably retained under automation, *cognitive* flying skills (navigational awareness, troubleshooting, position tracking) degrade significantly [19, 40]. The probable-cause finding of the Asiana 214 accident report (the flight crew's "inadequate monitoring of airspeed" following an "unintended deactivation of automatic airspeed control") is now a textbook real-world repayment event for accumulated automation-induced interpretive debt [41]. In medicine, Goddard et al. documented automation-bias effects from clinical decision-support systems including both commission (following bad advice) and omission (missing problems automation did not flag) errors [42]. The generalisation to ordinary white-collar work is straightforward: analysts who outsource summary writing to LLMs cannot easily resume detailed close reading; managers who rely on dashboards lose the practice of reading raw operational data; reviewers who defer to recommender rankings lose the practice of forming independent judgements about the full candidate set. Empirical studies of AI adoption in the information technology sector have likewise documented effects on employee performance and reliance [43].

4.4 Repayment events

The mechanism that gives interpretive debt its distinctive analytical bite is the *repayment event*. A repayment event is any external trigger that requires the organisation to perform interpretive work it has not been performing. Four classes are particularly important.

First, *audits and regulatory inquiries*: the lender being asked to explain a three-year-old decision, the hospital being asked to justify a triage call, the platform being asked to defend a moderation choice. Wachter et al. analysis of the GDPR shows that explanation obligations are weaker in law than is sometimes assumed; nonetheless, the operational and reputational pressure to give accounts is real [38]. Second, *crises and near-misses*: moments at which the organisation's account of what it is doing is

© 2026 International Journal Press · Licensed under CC BY 4.0

challenged by the world. Vaughan account of the Challenger decision shows how an organisation that has normalised deferral struggles to reconstruct an integrated interpretation under crisis pressure [27]. Third, *system failures and outages*: the moments at which the automated system is unavailable and the human side must perform the interpretation that has been delegated. Fourth, *generational turnover and newcomer cohorts*: the slow, predictable repayment event in which the people who *did* once perform the interpretation leave, and their tacit knowledge cannot be transmitted because the practice has not been exercised by anyone for years [30, 31].

The defining feature of a repayment event is that it forces *cumulative* unperformed interpretation due at one moment. This is what distinguishes interpretive debt from mere automation bias or skill atrophy: the latency means the debt is invisible in routine operations, and the forcing function means it manifests sharply rather than gradually. The lending vignette with which this paper opened is the audit-and-inquiry case in its purest form: the decision was efficient when made; the interpretation is required years later and is not available. The Asiana 214 case illustrates the system-failure case: when automatic airspeed control was unintentionally deactivated, the residual human task (the interpretation of glide slope and airspeed in unaided conditions) proved beyond the practised capacity of the crew [41]. The third and fourth classes, generational turnover and crisis-induced reconstruction, are typically slower to manifest but no less consequential. Together they suggest that interpretive debt has a characteristic *temporal signature*: long periods of routine operation in which the debt is paid no attention, punctuated by short, expensive periods in which the organisation is asked to do interpretive work it can no longer do.

5. IMPLICATIONS AND RESEARCH AGENDA

5.1 Operationalising interpretive debt

A reasonable objection to the construct is non-falsifiability: until interpretive debt can be observed in specific organisations, it remains a rhetorical figure. I propose three operationalisation paths.

First, *audit-response time and quality*: how long does it take, and how completely, for an organisation to reconstruct the rationale for an automated decision when challenged? Differences across firms, controlling for system complexity, are an empirical proxy for differences in accumulated debt. Second, *newcomer time-to-competence*: how long does it take a newly hired analyst, clinician, or operator to develop a substantive working understanding of operational decisions: not merely interface competence with the supporting systems? When this duration lengthens despite stable training programmes, the organisation may be servicing less interpretive debt than it accrues. Third, *reconstruction fidelity*: when a sample of past automated decisions is presented to current staff for de novo justification, what fraction can be reconstructed with reasoning that matches the original context? This metric, adapted from Endsley situation-awareness assessment tradition, provides a directly measurable indicator of interpretive capacity [44].

5.2 Mitigation: from “automate” to “informate”

Zuboff distinction between deploying information technology to *automate* (substitute machines for human agency) versus to *informate* (to “render events, objects, and processes visible, knowable, and shareable in a new way” (pp. 9–10)) supplies the most useful prescriptive frame [6]. Interpretive-debt-aware deployment is informating deployment. Concretely, this implies that every automated decision-support layer should be paired with practices that re-render the system’s reasoning legible, contestable, and pedagogically tractable. Three classes of practice deserve particular emphasis. *Shadow exercises*, in which staff are required periodically to produce decisions without the automated tool and then compare their reasoning with the system’s output, exercise the interpretive practice that routine deferral would otherwise atrophy. *Rotating audit duties*, in which a sample of past automated decisions is regularly assigned to current staff for de novo justification, exercise the reconstruction capacity that the audit-class repayment event will eventually demand. *Documentation for re-interpretation*, in which the documentation surrounding automated decisions is oriented towards a future reader trying to reconstruct context (not merely towards present compliance) pre-builds the bridges across which future reasoning will need to travel.

Weick and Sutcliffe five HRO principles (preoccupation with failure, reluctance to simplify interpretations, sensitivity to operations, commitment to resilience, and deference to expertise) translate directly into interpretive-debt servicing practices [7]. Reluctance to simplify interpretations and sensitivity to operations in particular function as anti-deferral norms: they require the organisation continually to re-exercise the practice of asking what the system’s outputs actually mean. Lee and See calibrated-trust framework supplies the individual-level analogue: appropriate reliance, neither over- nor under-trust,

© 2026 International Journal Press · Licensed under CC BY 4.0

requires a level of interpretive capacity that interpretive-debt-burdened organisations may have lost [45]. Across all of these prescriptions, the underlying point is that the *practice* of interpretation must be exercised to be retained, and that organisational design can either build that exercise into the routine or (as is presently the default) allow it to be quietly stripped away.

5.3 Research agenda

Four lines of empirical work follow. (a) Case studies of repayment events (audits, crises, system failures, generational handovers) tracing how organisations succeeded or failed at reconstructing interpretation, with interpretive debt as the explanatory variable. (b) Comparative studies of organisations deploying functionally similar automated systems but with divergent informating practices, testing whether informating organisations accumulate less debt for equivalent automation intensity. (c) Longitudinal observation of newcomer cohorts in heavily automated work settings, measuring time-to-substantive-competence as a debt indicator. (d) Cross-validation of the construct against the HRO five-principles framework: whether organisations scoring high on HRO mindfulness measures show measurably lower interpretive debt by the operationalisations above.

6. BOUNDARY CONDITIONS AND LIMITATIONS

Three boundary conditions delimit the construct. First, interpretive debt presupposes that the relevant decisions *require* interpretation: that there is meaningful work for human sensemaking to do. For genuinely routine, low-stakes, high-frequency decisions (which constitute a substantial fraction of automated workflows), debt accumulation may be acceptable because the latent risk is small. The construct's force is greatest where decisions are consequential, contested, or accountable to outsiders. Second, interpretive debt presupposes that the automated system is *meaningfully opaque* to the deferring human. Where systems are genuinely transparent and their reasoning verifiable on demand, deferral may carry no debt; this is the limiting case at which good informating practice approaches zero accumulation. Third, the construct presupposes that *organisations care about accounting for their decisions*: for legal, professional, ethical, or operational reasons. Where no such accountability obligation exists, interpretive debt is a private cost the organisation may rationally choose to bear.

Two limitations of the present paper should be acknowledged. First, the construct is at this stage conceptual; the three operationalisation paths in §5.1 are proposals, not validated measures. Second, three of the most directly comparable adjacent constructs are very recent and the comparative literature is therefore still settling [3, 4, 5]. Reviewers may reasonably take a different view of where the boundaries between cognitive, epistemic, and interpretive debt should lie. I have argued that the locus (organisational sensemaking capacity), the object (operational decisions), and the repayment-event mechanism distinguish interpretive debt sufficiently to justify its independence; this argument remains to be tested against further empirical and conceptual work.

7. CONCLUSION

This paper has developed *interpretive debt* as a conceptual frame for a phenomenon that current socio-technical scholarship can describe only piecewise. When organisations defer sensemaking work to automated systems (when dashboards substitute for reading the operations, when recommenders substitute for surveying the field, when risk scores substitute for considering the case, when AI assistants substitute for drafting and arguing) each individual deferral is locally rational. The cumulative effect is the slow depletion of the organisation's collective capacity to make sense of itself: to know when the systems are wrong, to reconstruct why past decisions were taken, and to bring new members into substantive rather than merely procedural competence. This depletion remains invisible until a repayment event (audit, crisis, system failure, or turnover) forces the organisation to perform interpretation it has not been practising.

The contribution is to name an organisational asset that is otherwise nameless (interpretive capacity), to give it a debt structure that supports analytical work (deferral, opacity, atrophy as accumulation mechanisms; the repayment event as the forcing function), to differentiate it from the closest adjacent constructs (cognitive debt, epistemic debt, automation bias, deskilling), and to integrate four scholarly traditions that have rarely been read together. The construct's empirical purchase remains to be established; the conceptual case is that the move from automating to informating, in Zuboff's sense, requires a vocabulary for what is at stake when the move is not made, and interpretive debt is offered as that vocabulary.

© 2026 International Journal Press · Licensed under CC BY 4.0

References

[1] Cunningham, W. (1992). The WyCash portfolio management system. *Addendum to the Proceedings of the Conference on Object-Oriented Programming Systems, Languages, and Applications (OOPSLA '92)*, 29–30. Association for Computing Machinery. https://doi.org/10.1145/157709.157715

[2] Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 28, pp. 2503–2511). Curran Associates.

[3] Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task* (arXiv preprint No. 2506.08872). arXiv. https://doi.org/10.48550/arXiv.2506.08872

[4] Storey, M.-A. (2026). *From technical debt to cognitive and intent debt: Rethinking software health in the age of AI* (arXiv preprint No. 2603.22106). arXiv. https://doi.org/10.48550/arXiv.2603.22106

[5] Sankaranarayanan, S. (2026). Mitigating "epistemic debt" in generative AI-scaffolded novice programming using metacognitive scripts. In *Proceedings of the 13th ACM Conference on Learning at Scale (L@S '26)*. Association for Computing Machinery. https://arxiv.org/abs/2602.20206

[6] Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. Basic Books.

[7] Weick, K. E., & Sutcliffe, K. M. (2015). *Managing the unexpected: Sustained performance in a complex world* (3rd ed.). Jossey-Bass.

[8] Fowler, M. (2009, October 14). Technical debt quadrant. *MartinFowler.com*. https://martinfowler.com/bliki/TechnicalDebtQuadrant.html

[9] Dudycz, O. (2024, June 13). Tech debt doesn't exist, but trade-offs do. *Architecture Weekly*. https://www.architecture-weekly.com/p/tech-debt-doesnt-exist-but-trade

[10] Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. D. (1998). Automation bias: Decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology*, *8*(1), 47–63. https://doi.org/10.1207/s15327108ijap0801_3

[11] Weick, K. E. (1995). *Sensemaking in organizations*. Sage.

[12] Weick, K. E. (1993). The collapse of sensemaking in organizations: The Mann Gulch disaster. *Administrative Science Quarterly*, *38*(4), 628–652. https://doi.org/10.2307/2393339

[13] Weick, K. E. (1990a). The vulnerable system: An analysis of the Tenerife air disaster. *Journal of Management*, *16*(3), 571–593. https://doi.org/10.1177/014920639001600304

[14] Maitlis, S., & Christianson, M. (2014). Sensemaking in organizations: Taking stock and moving forward. *Academy of Management Annals*, *8*(1), 57–125. https://doi.org/10.5465/19416520.2014.873177

[15] Gioia, D. A., & Chittipeddi, K. (1991). Sensemaking and sensegiving in strategic change initiation. *Strategic Management Journal*, *12*(6), 433–448. https://doi.org/10.1002/smj.4250120604

[16] Pratt, M. G. (2000). The good, the bad, and the ambivalent: Managing identification among Amway distributors. *Administrative Science Quarterly*, *45*(3), 456–493. https://doi.org/10.2307/2667106

[17] Stigliani, I., & Ravasi, D. (2012). Organizing thoughts and connecting brains: Material practices and the transition from individual to group-level prospective sensemaking. *Academy of Management Journal*, *55*(5), 1232–1259. https://doi.org/10.5465/amj.2010.0890

[18] Weick, K. E. (1990b). Technology as equivoque: Sensemaking in new technologies. In P. S. Goodman, L. S. Sproull, & Associates (Eds.), *Technology and organizations* (pp. 1–44). Jossey-Bass.

© 2026 International Journal Press · Licensed under CC BY 4.0

[19] Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8

[20] Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research* (Vol. 2, pp. 1–12). Elsevier.

[21] Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

[22] Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

[23] Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, *59*(1), 5–27. https://doi.org/10.1177/0018720816681350

[24] Reason, J. (1990). *Human error*. Cambridge University Press.

[25] Reason, J. (2000). Human error: Models and management. *BMJ*, *320*(7237), 768–770. https://doi.org/10.1136/bmj.320.7237.768

[26] Perrow, C. (1984). *Normal accidents: Living with high-risk technologies*. Basic Books.

[27] Vaughan, D. (1996). *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. University of Chicago Press.

[28] Dekker, S. (2011). *Drift into failure: From hunting broken components to understanding complex systems*. Ashgate.

[29] Walsh, J. P., & Ungson, G. R. (1991). Organizational memory. *Academy of Management Review*, *16*(1), 57–91. https://doi.org/10.5465/amr.1991.4278992

[30] Polanyi, M. (1966). *The tacit dimension*. Doubleday.

[31] Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press.

[32] Braverman, H. (1974). *Labor and monopoly capital: The degradation of work in the twentieth century*. Monthly Review Press.

[33] Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

[34] Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, *333*(6043), 776–778. https://doi.org/10.1126/science.1207745

[35] 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

[36] 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

[37] Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

[38] Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, *7*(2), 76–99. https://doi.org/10.1093/idpl/ipx005

[39] Pandey, J. K., & Kumar, R. (2025). Governing the algorithmic agent: Confronting overt and covert challenges to justice and the future of work. *International Journal of Law Management & Humanities*, *8*(4), 1974–1984. https://doij.org/10.10000/IJLMH.1110648

[40] Casner, S. M., Geven, R. W., Recker, M. P., & Schooler, J. W. (2014). The retention of manual flying skills in the automated cockpit. *Human Factors*, *56*(8), 1506–1516. https://doi.org/10.1177/0018720814535628

[41] National Transportation Safety Board. (2014). *Descent below visual glidepath and impact with seawall, Asiana Airlines Flight 214, Boeing 777-200ER, HL7742, San Francisco, California, July 6,*

© 2026 International Journal Press · Licensed under CC BY 4.0

2013 (Aircraft Accident Report NTSB/AAR-14/01). https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1401.pdf

[42] Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, *19*(1), 121–127. https://doi.org/10.1136/amiajnl-2011-000089

[43] Lalitha Kumari, P., & Pandey, J. K., et al. (2021). Impact of artificial intelligence on employees' performance in information technology sector. *AIP Conference Proceedings*, *2587*(1). https://doi.org/10.1063/5.0150403

[44] Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

[45] Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

© 2026 International Journal Press · Licensed under CC BY 4.0