

Toward Reconstructing the Concept of Erosion of Independent Professional Judgment (EIPJ) under Algorithmic Management: A Self-Critical

PhD Student. **Dhia Eddine Berbaché***¹, PhD Student. **Djamel Eddine Reguig Berra***², **Rayane Zazi***³

¹ Laboratory of Psychological and Social Studies, Mohamed Khider University of Biskra, (Algeria)
Email : dhiaeddine.berbaché@univ-biskra.dz ; **ORCID**  <https://orcid.org/0009-0002-4424-6419>

² Laboratory of Psychological and Social Studies, Mohamed Khider University of Biskra, (Algeria)
Email : djameleddine.reguigberra@univ-biskra.dz ; **ORCID**  <https://orcid.org/0009-0000-8859-8294>

³ Ferhat Abbas University Sétif_1, (Algeria)
Email : rayane.zazi27@gmail.com ; **ORCID**  <https://orcid.org/0009-0000-2013-4020>

Received: 01/01/2026 ; Accepted: 28/06/2026 ; Published: 08/09/2026

Abstract:

This paper presents a self-critical review of the "organizational cognitive erosion" concept previously proposed by the same author, testing its conceptual validity limits rather than defending it. The paper narrows the claim to a single, well-defined construct termed "Erosion of Independent Professional Judgment" (EIPJ), operationally distinguished from four adjacent constructs, and provides a five-dimensional operational definition along with explicit exclusion criteria. The paper explicitly acknowledges that one of its theoretical anchors remains an unsettled synthesis, and downgrades its reliance on a single study that was subsequently subject to a publisher correction.

Keywords: Erosion of Independent Professional Judgment; Algorithmic Management; Deskilling; Measurement Problem; Self-Critical Review.

1. INTRODUCTION

This paper declares from its opening lines that it constitutes a self-critical review of the "organizational cognitive erosion" concept previously proposed by the author in an earlier work — not a defense of that concept, but rather a systematic test of its conceptual validity limits against the same standards applied to any novel theoretical construct.

The paper proposes narrowing the claim to a single, well-defined construct: the Erosion of Independent Professional Judgment (EIPJ), wherein all subsidiary terms are treated as operational dimensions of a single concept rather than as independent constructs in their own right.

This paper also declares its philosophical position explicitly: it does not proceed from the assumption that human autonomy is an absolute value. The methodological concern advanced here is more circumscribed — not the retreat from reliance on human judgment in tasks where algorithms demonstrably outperform statistically, but rather the irreversible degradation of the underlying capacity for such judgment. This concern aligns with what a recent comprehensive review of the AI risk landscape has termed "enfeeblement," defined as the gradual decline in human capabilities and agency that results not from a sudden loss of control but from an accumulation of individually rational micro-decisions to delegate cognitive tasks to artificial

intelligence — decisions that, taken together, generate a dependency spiral in which humans progressively lose the skills, confidence, and judgment required for independent functioning. (Grey & Segerie, 2025, p. 72)

Four established constructs account for aspects of the competence-degradation phenomenon under automation: Deskilling captures the structural dispossession of procedural knowledge through task fragmentation; skill decay describes the deterioration of a specific skill following interrupted practice; automation complacency refers to a momentary performance state in monitoring tasks; and cognitive offloading denotes the voluntary transfer of mental operations to an external tool.

This paper proposes the adoption of the formal abbreviation EIPJ for the reconstructed construct, and operates through a critically oriented methodology structured along two axes: the first addresses structural gaps in the conceptual architecture, while the second addresses reconstruction, the causal mechanism, and the conditions for the construct's survival and falsification.

2. AXIS ONE: STRUCTURAL GAPS AND THE RATIONALE FOR A NEW CONSTRUCT

2.1. Why Updating Deskilling Is Insufficient Rather Than Building a New Construct

Deskilling, as theorized by Braverman in his foundational treatise and critically examined by Attewell, describes the transfer of procedural knowledge from worker to management through the fragmentation of labor — such that the act of thinking itself is structurally reassigned to management, leaving the worker no opportunity to exercise it regardless of motivational change. (Braverman, 1974, p. 85)

Braverman's central argument rests on the separation of conception from execution: under scientific management principles, the planning, judgment, and decision-making components of a task are systematically extracted from the worker's role and consolidated within managerial or technical hierarchies, while the worker retains only the physically executable residue of the task. This is not, in Braverman's formulation, a gradual erosion of a capacity that remains latent and recoverable — it is a structural redesign of the job itself, such that the cognitive component of the work no longer exists as an available option for the worker to exercise, regardless of motivation, incentive, or attentiveness. Attewell's critical examination nuances this account by noting that the empirical picture is more variable across occupations and historical periods than Braverman's original thesis suggests, with some instances of technical change generating upskilling rather than deskilling, and with the direction and magnitude of the effect depending heavily on labor market conditions, managerial choices, and the specific technology in question. Even accounting for this critical qualification, the core structural insight that both scholars converge upon remains decisive for the argument advanced here: the mechanism Braverman describes operates at the level of job design and organizational structure, not at the level of an individual worker's psychological or cognitive state. (Attewell, 1987)

This paper proposes that EIPJ describes a narrower and more precisely bounded condition than Deskilling: the continued deterioration in the quality of specialized professional judgment despite the theoretical availability of opportunity to exercise it and despite changes in organizational incentive. This distinction matters precisely because a worker subjected to genuine Deskilling in Braverman's sense cannot simply choose, one day, to resume exercising evaluative judgment on the task, since the task, as structurally reconfigured, no longer contains a slot within

which that judgment could be exercised — whereas the condition this paper seeks to isolate is one in which that slot remains structurally present, technically available, and formally permitted, yet the capacity to fill it competently has nonetheless deteriorated through prolonged non-use. Three structural distinctions justify the independent standing of this proposed construct. First, Deskilling describes a structural dispossession of work as a whole that does not recede with motivational change — no increase in incentive or attentiveness restores what has been architecturally removed from the job — whereas EIPJ describes a deterioration in a latent capacity for which the system continues to provide theoretical opportunity, meaning that the possibility of recovery, however difficult in practice, is not foreclosed by design in the way it is under Deskilling. Second, Deskilling targets the loss of procedural-technical skill in its narrow sense — the know-how required to execute a defined sequence of operations — whereas EIPJ targets something more evaluative in character: the capacity for judgment under conditions of ambiguity and uncertainty, where no single correct procedure exists and the worker must weigh incomplete or conflicting information to arrive at an assessment. This is a meaningfully different cognitive demand from the procedural competence at stake in classical accounts of Deskilling. Third, Deskilling is a collective phenomenon affecting an entire labor category as a structural feature of how that category of work has been redesigned, whereas EIPJ operates at the individual level, with variability determined by accumulated professional experience, meaning that two workers occupying formally identical roles within the same algorithmically managed system may nonetheless exhibit markedly different degrees of erosion depending on the depth and recency of their prior independent practice.

Should these distinctions be empirically confirmed, EIPJ acquires its legitimacy as an independent construct, standing apart from Deskilling not merely as a rebranding but as a description of a genuinely distinct phenomenon operating through a different mechanism, at a different level of analysis, with different implications for recoverability. Should they fail to hold — should EIPJ prove, upon closer empirical scrutiny, to collapse into either the structural dispossession described by Deskilling or the practice-interruption dynamics described by Skill Decay — then the construct does not survive as an independent contribution and is instead reintegrated into whichever of those two literatures actually accounts for the observed pattern. This conditional stance is not incidental to the paper's argument; it is precisely what the falsification conditions specified in Axis Two are designed to determine, and it reflects the broader self-critical posture that the paper adopts throughout: the construct is offered as a testable proposal whose survival depends on evidence, not as a conclusion defended in advance of that evidence.

2.2. Clarifying the Boundaries: Algorithmic Management and Generative Artificial Intelligence

Algorithmic management refers to the use of a computerized system to perform traditional managerial functions such as task allocation, performance evaluation, and monitoring. This definition situates algorithmic management squarely within the institutional apparatus of the organization: it is not a tool that a worker elects to consult when convenient, but a system embedded in the formal architecture of how work is assigned, assessed, and supervised, such that the worker's relationship to it is structured by organizational authority rather than by personal preference. This institutional embeddedness is what gives algorithmic management its distinctive character as a managerial function rather than a mere productivity aid — decisions about what task to perform next, how that performance will be scored, and whether that performance meets an

acceptable threshold are, under algorithmic management, delegated in whole or in part to a computational system operating on behalf of the organization.

This paper proposes that EIPJ target exclusively algorithmic management as an institutionally imposed context performing managerial and evaluative functions, rather than the voluntary personal use of generative AI systems. This boundary is not a minor definitional nicety; it demarcates two categorically different relationships between a worker and an automated system. In the case of algorithmic management, the worker typically has limited or no discretion over whether the system is used, how its outputs are weighted, or what recourse exists when its judgments diverge from the worker's own — the system is, in an important sense, imposed as a condition of continued employment or continued standing within the organization. In the case of voluntary personal use of generative AI, by contrast, the worker retains discretion over whether to consult the system at all, and any resulting change in cognitive habits arises from a series of individually chosen acts of delegation rather than from an externally structured requirement. Conflating these two contexts would obscure a mechanism that this paper considers central to the proposed construct: namely, that the erosion at stake is hypothesized to arise, at least in part, from the involuntary or institutionally compelled character of the worker's exposure to the automated system, not merely from repeated exposure to automation as such.

This distinction carries practical significance, as affirmed by the literature on the "duality of algorithmic management" — a conceptual lens advanced by Meijerink and Bondarouk for interpreting the complex relationship between HR management algorithms and task autonomy and the value accruing to workers subject to such systems, in contrast to the prevailing tendency to present algorithmic management as a phenomenon with predetermined negative outcomes. (Meijerink & Bondarouk, 2021)

The duality framework is instructive here precisely because it resists a simple narrative in which algorithmic management is treated as uniformly corrosive to worker autonomy or capability. Instead, it holds open the possibility that the same underlying technology can, depending on its design and the discretion it affords, either constrain or expand the space within which a worker exercises independent judgment. Adopting this more differentiated view has direct implications for how EIPJ is scoped in this paper: rather than treating algorithmic management as monolithically productive of erosion, the construct is proposed as applying specifically to instances in which the system exhibits a substitutive character — that is, instances in which the institutional design of the algorithmic management system displaces rather than merely supplements the worker's own evaluative role. This qualification prevents the construct from overreaching into contexts where algorithmic management, despite its institutional and managerial character, may in fact preserve or even enhance the worker's opportunity to exercise professional judgment, a possibility the duality literature explicitly keeps in view.

2.3. Table of Distinctions between EIPJ and Adjacent Conceptual Constructs

This paper proposes clarifying this distinction through Table 1 (presented in the Figures and Tables section), using five criteria: the affected level, the primary cause, recoverability, the temporal frame, and the structural intervention point. These five criteria were selected not arbitrarily but because each targets a dimension along which the constructs surveyed in this paper — Deskilling, Skill Decay, Automation Complacency, Algorithmic Dependence, and EIPJ itself — diverge in ways that carry distinct theoretical and practical implications. The affected level criterion distinguishes whether the phenomenon operates on the structure of work itself, on a

discrete skill, on momentary attention, on a behavioral pattern, or on a latent evaluative capacity. The primary cause criterion isolates the mechanism through which the deterioration is hypothesized to arise, since two constructs may describe superficially similar declines in performance while pointing to entirely different generative processes. Recoverability, as elaborated further below, functions as perhaps the single most consequential criterion for separating EIPJ from its nearest conceptual neighbor. The temporal frame criterion situates each construct along a spectrum ranging from momentary lapses measured in minutes to cumulative processes unfolding over years. And the structural intervention point criterion, introduced and defined specifically for the purposes of this paper, supplies the design-level test that determines whether a given construct is even a candidate for the kind of erosion this paper seeks to isolate.

The criterion of "structural intervention point" refers to the design-level condition in which the task structure does not structurally preclude the possibility of manual intervention, even when that possibility has not been exercised for an extended period. This criterion is deliberately formulated at the level of system design rather than at the level of observed worker behavior, because the paper's central concern is with a capacity that deteriorates despite the continued formal availability of an opportunity to exercise it — a concern that only becomes analytically tractable once one can independently verify, at the level of the system itself, that such an opportunity in fact exists in technical, policy, and practical terms. Without this design-level anchor, any observed decline in worker performance would be indistinguishable from a decline attributable to the outright removal of the opportunity to act, which is precisely the condition that characterizes Deskilling rather than EIPJ. This distinction is consistent with a recent quantitative model of AI dependence, which documented across four distinct domains that endoscopists who used AI-assisted colonoscopy for twelve weeks recorded a 21% decline in polyp detection rates upon withdrawal of that assistance, and that pilots whose actual manual flying averaged only five to fifteen minutes per flight failed between 38% and 44% of basic situational awareness tasks when their independent skills were assessed in the absence of autopilot systems. (Park et al., 2026, pp. 1–2)

Both of these empirical illustrations share a common structural feature that is directly relevant to the criterion under discussion: in neither case had the underlying manual capability been formally withdrawn from the professional's role. Endoscopists remain licensed and equipped to detect polyps without algorithmic assistance, and pilots remain certified and technically able to fly manually; the intervention point, in the sense defined by this paper, was structurally present throughout. What had changed was not the availability of the opportunity but the frequency and depth of its actual exercise, which is exactly the configuration that this paper's proposed criterion is designed to capture and that distinguishes it from a scenario of structural dispossession.

The decisive distinguishing criterion between "algorithmic dependence" and EIPJ is "recoverability": if a worker readily recovers independent performance upon motivational change, the more accurate description is dependence rather than erosion. This criterion functions as the paper's primary safeguard against construct proliferation, since Algorithmic Dependence and EIPJ might otherwise appear, on a superficial reading, to describe the same underlying phenomenon — both involve a worker who has come to rely on an automated system and who exhibits diminished independent performance in its absence. The recoverability criterion supplies the empirical test that separates them: dependence, as this paper conceives it, is fundamentally a behavioral and motivational pattern, such that a sufficiently strong incentive, a policy change, or a simple

reminder of the stakes involved is enough to restore the worker's willingness and ability to perform independently, because the underlying capacity itself was never substantially degraded — only its habitual exercise had lapsed. Erosion, by contrast, is proposed in this paper as describing a condition in which the capacity itself has deteriorated at a level that motivational change alone cannot immediately reverse, such that even a highly incentivized worker would require a period of renewed practice, and possibly targeted retraining, before independent performance returns to its prior baseline. This is what renders recoverability the decisive criterion rather than merely one criterion among several: it is the empirical marker that, in principle, allows a researcher to determine after the fact which of the two constructs actually describes a given worker's observed pattern of decline and subsequent response to intervention.

2.4. Theoretical Grounding of "Latent Capacity" with an Explicit Caveat on Its Degree of Stability

The term "latent capacity" as used in this paper draws on three specific bodies of literature: first, the tacit knowledge literature as developed by Polanyi and applied to organizational knowledge by Nonaka; second, the expertise literature as established by Ericsson and colleagues, which demonstrates that expertise deteriorates when actual practice is interrupted even as explicit theoretical knowledge remains intact; and third, the cognitive reserve literature as theorized by Stern, which indicates that latent capacity refers to a reservoir of mental capacity that can be mobilized when needed, independent of its habitual rate of use. (Polanyi, 1966, pp. 4–25)

Each of these three bodies of literature was selected to supply a distinct component of what the paper means by "latent capacity" rather than to serve as interchangeable restatements of a single idea. The tacit knowledge literature contributes the insight that a significant portion of professional competence is not reducible to explicit, articulable rules, but instead resides in forms of knowing that are acquired through embodied practice and organizational immersion — knowledge that, precisely because it is tacit rather than codified, can be present in a worker's repertoire without being continuously exercised, and can therefore plausibly be subject to a distinct process of decline when the practice that sustains it lapses. The expertise literature contributes a complementary but analytically separate insight: that even highly deliberate, effortfully acquired expertise is not a permanent acquisition but a performance capacity that requires ongoing maintenance through actual engagement with the task, such that the theoretical knowledge underlying that expertise can remain intact in a worker's explicit understanding of the domain while the applied performance capacity itself deteriorates in the absence of continued practice. The cognitive reserve literature contributes a third and again distinct insight: the notion of a mobilizable reservoir of capacity that exists independently of its habitual rate of use, which supplies the paper with a conceptual basis for treating "latency" itself as a coherent property — the idea that a capacity can be present, in some meaningful sense, even during a period in which it is not being actively drawn upon, rather than simply ceasing to exist the moment it goes unused. (Ericsson et al., 1993)

This paper explicitly acknowledges a methodological limit: this synthesis across three bodies of literature is a compositional proposal by the author and not an established concept agreed upon across these three literatures collectively. This acknowledgment is not a peripheral disclaimer but a substantive methodological commitment that runs through the paper's broader self-critical posture: rather than presenting the three-literature synthesis as though it reflected a pre-existing scholarly consensus, the paper is explicit that the act of drawing these three bodies of work together into a single unified notion of "latent capacity" is itself an interpretive move undertaken by the

author, and one that carries a corresponding burden of justification and a corresponding vulnerability to challenge. Polanyi and Nonaka address tacit knowledge that is socially and organizationally transmitted — their unit of analysis is fundamentally interpersonal and institutional, concerned with how knowledge circulates, is embedded in practice, and is passed between individuals within an organizational context. Ericsson addresses individual expertise formed through deliberate practice — his unit of analysis is the single performer engaged in a structured, effortful regimen of skill acquisition, a framework developed primarily in domains such as music and chess where performance can be precisely measured against an objective standard. Stern addresses a neurophysiological reserve measurable through direct neural indicators — his unit of analysis is grounded in observable biological substrates, such as brain volume, synaptic density, or patterns of neural network efficiency, that can in principle be assessed independently of any behavioral report. These are, in other words, three phenomena that operate at fundamentally different levels of analysis: the social-organizational, the individual-behavioral, and the neurophysiological respectively. The paper's decision to draw on all three simultaneously is therefore a deliberate theoretical wager rather than a straightforward citation of settled consensus, and this section's function is to render that wager visible rather than to obscure it behind the appearance of an uncontested synthesis. (Nonaka, 1994)

2.5. Operational Definition of "Independent Professional Judgment" and Its Exclusion Criteria

This paper defines "independent professional judgment" operationally as the capacity to integrate multiple incomplete pieces of information within a specialized task context to produce an assessment or decision that transcends the literal application of a pre-existing written rule, and proposes five operational sub-dimensions for it, all subordinate to the single EIPJ definition: diagnosis; estimation; problem-solving; critical reasoning; and anomaly detection. This operational definition is constructed deliberately to exclude two categories of activity that might otherwise be mistaken for the phenomenon under investigation. It excludes rule-following in the strict sense, since a worker who correctly applies a fully specified written procedure to a fully specified situation is not, on this definition, exercising independent professional judgment at all, regardless of how competently that application is performed — the defining feature of the capacity this paper seeks to isolate is precisely that it is called upon in situations where the available information is incomplete, where the applicable rule is ambiguous or absent, and where the worker must therefore synthesize disparate and imperfect inputs into a coherent assessment. It also implicitly excludes any decision-making process that could, in principle, be fully automated through an exhaustive decision tree, since such a process would by definition not require the integrative, judgment-based synthesis that the definition specifies. The five proposed sub-dimensions are intended to operationalize this general definition across the range of concrete forms that such judgment typically takes in professional practice: diagnosis captures the task of inferring an underlying condition or cause from a set of observable but incomplete symptoms or indicators; estimation captures the task of producing a quantitative or qualitative approximation in the absence of complete or certain data; problem-solving captures the task of devising a course of action to resolve a novel or non-routine difficulty for which no pre-existing solution has been specified; critical reasoning captures the task of evaluating the soundness, consistency, or plausibility of a claim, argument, or proposed course of action; and anomaly detection captures the task of recognizing that a given situation deviates from an expected pattern in a way that warrants further

scrutiny. These five sub-dimensions are explicitly framed as subordinate to the single EIPJ definition rather than as independent constructs in their own right, meaning that they are not proposed as five separate phenomena each requiring its own measurement and validation, but as five distinguishable manifestations of a single underlying capacity, included to give that capacity operational traction across the diverse forms it takes in actual professional practice. (Stern, 2002)

This paper stipulates that professional judgment must be distinguished across occupations, which renders EIPJ a higher-order construct rather than a uniform single-form structure. This stipulation follows directly from the operational definition just given: because the specific content of diagnosis, estimation, problem-solving, critical reasoning, and anomaly detection will necessarily differ substantially depending on the professional domain in question — the diagnostic judgment exercised by a physician interpreting ambiguous clinical presentations is not interchangeable with the diagnostic judgment exercised by a mechanical engineer inferring the cause of a structural failure, even though both fall under the same sub-dimension in the abstract — EIPJ cannot be treated as though it referred to a single uniform skill measurable by a single uniform instrument across all occupations. Instead, EIPJ functions as a higher-order construct, meaning that it specifies a general conceptual structure and a general set of sub-dimensions that must then be operationalized separately, and with attention to domain-specific content, within each professional context to which the construct is applied.

The concept of EIPJ, as defined in this paper, does not encompass: performance decline attributable to fatigue, burnout, or motivational deficits unless capacity deterioration is demonstrated after controlling for these factors. This exclusion criterion is essential to the construct's conceptual integrity, because fatigue, burnout, and motivational deficits are all well-documented sources of performance decline that operate through mechanisms entirely distinct from the sustained non-use of a structurally available intervention point that this paper hypothesizes as the driver of EIPJ, and a construct that failed to exclude these alternative sources would risk attributing to algorithmic management a decline that in fact originates elsewhere. This exclusion is methodologically necessary, as a systematic review has documented that occupational burnout is strongly and consistently associated with performance decline specifically in memory — with burned-out workers exhibiting impairments in both prospective and delayed memory — while evidence of its effects on executive functions and attention remains less consistent across studies. (Renaud & Lacroix, 2023, p 109)

This finding is particularly instructive for the exclusion criterion under discussion, because it indicates that burnout's cognitive footprint is not uniform across all cognitive domains: its effects on memory appear robust and well-replicated, while its effects on the executive and attentional functions more closely associated with the kind of evaluative judgment this paper is concerned with remain comparatively uncertain. This unevenness underscores why any observed decline in a worker's independent professional judgment must be examined with burnout and fatigue explicitly controlled for before that decline can be attributed to EIPJ, since failing to do so risks conflating a well-established memory-specific effect of burnout with the judgment-specific erosion this paper proposes as a distinct phenomenon.

The concept further excludes: deterioration in average routine performance on tasks in which the algorithm genuinely and demonstrably outperforms, and mere preference for system use when the system is available. These two additional exclusions serve a similar boundary-setting function. The first prevents the construct from being invoked simply because a worker's unassisted

performance on a given task falls below the performance of the algorithmic system, since in cases where the algorithm genuinely and demonstrably outperforms unassisted human judgment, a decline in the worker's relative standing against that algorithm reflects the algorithm's superior capability rather than any deterioration in the worker's own underlying capacity. The second prevents the construct from being invoked on the basis of a worker's behavioral preference for using the system when it is available, since such a preference, standing alone, indicates only a pattern of choice and carries no direct implication regarding whether the worker's independent capacity has in fact deteriorated — a worker may strongly prefer to use an available algorithmic tool for reasons of convenience or efficiency while nonetheless retaining, if called upon, a fully intact capacity for independent judgment.

2.6. The Phenomenon of Apparent Success: How EIPJ Remains Invisible

This paper proposes that EIPJ may remain invisible so long as the algorithmic system functions normally, manifesting only when the system fails in a rare critical situation. This proposition carries an important implication for how the construct might be detected, studied, or even suspected in the first place: under ordinary operating conditions, an organization relying on a well-functioning algorithmic management system would observe nothing untoward, since the system itself continues to produce acceptable outcomes, the worker continues to be evaluated favorably against whatever performance metrics the organization tracks, and no visible signal exists to indicate that the worker's own independent capacity to perform the task without algorithmic assistance has quietly deteriorated in the background. The erosion, on this account, is not merely difficult to observe — it is structurally rendered invisible by the very success of the system that is hypothesized to be producing it, because the routine functioning of the algorithm continuously substitutes for the exercise of independent judgment, thereby simultaneously preventing that judgment from being called upon and preventing its absence from being noticed. It is only in the rare instance in which the algorithmic system itself fails, malfunctions, or encounters a situation outside its operational parameters — precisely the kind of critical, high-stakes moment in which independent human judgment would be most urgently needed as a fallback — that the underlying erosion becomes observable, and by that point, the observation arrives at the worst possible moment, when the consequences of the worker's diminished capacity are most likely to be severe.

This observation acquires particular significance in light of what Bainbridge observed four decades ago in the context of industrial automation, when she demonstrated that a human operator left to intervene only rarely in exceptional circumstances progressively loses the very skill required for such intervention — without noticing this loss until the system fails at a critical moment. (Bainbridge, 1983, p. 776)

Bainbridge's original formulation, developed in the context of industrial process control systems, identified what has since come to be recognized as one of the central paradoxes of automation: the more reliable and capable an automated system becomes, the less frequently a human operator is called upon to intervene manually, and the less frequently that operator intervenes, the more that operator's own manual skill deteriorates through simple disuse — with the perverse consequence that the operator is least prepared to intervene competently at precisely the moment when intervention is most critically needed, namely when the highly reliable system finally does fail. This dynamic is instructive for the argument advanced in the present paper for two related reasons. First, it supplies a historical and empirically grounded precedent for the

general mechanism this paper proposes, demonstrating that the phenomenon of skill erosion masked by system reliability is not a novel speculative concern invented for the purposes of this paper but a documented feature of human-automation interaction observed across a substantial span of industrial history. Second, and more specifically relevant to the present section's argument, Bainbridge's finding directly substantiates the claim that such erosion characteristically goes unnoticed by the very operator experiencing it: the operator in Bainbridge's account does not consciously perceive their own skill declining in real time, precisely because the opportunities to test that skill against reality have become so infrequent that no ordinary occasion exists on which the decline could be registered — the loss surfaces only retrospectively, and only at the moment of system failure, which is exactly the pattern of invisibility this paper proposes as characteristic of EIPJ under algorithmic management more broadly.

2.7. The Measurement Problem: Five Differentiated Approaches

This paper proposes five measurement approaches: first, actual performance assessment in a controlled simulation task; second, general cognitive tests; third, expert evaluation of real decisions; and fourth, indirect behavioral indicators. The fifth approach — proposed here as a necessary addition absent from prior formulations — targets specifically the measurement of the capacity to detect rare system-failure events. Each of these five approaches is offered as a distinct response to a measurement problem that the invisibility phenomenon described in the preceding section makes especially acute: if EIPJ, by its very nature, does not manifest during ordinary system operation and surfaces only at the moment of rare critical failure, then any attempt to measure it must find some way of assessing a capacity that is, under normal working conditions, never actually called upon and therefore never directly observed in the natural course of a worker's job. The five approaches proposed here can be understood as five different strategies for making that hidden capacity observable without simply waiting for a real system failure to occur.

The first approach, actual performance assessment in a controlled simulation task, addresses this problem by artificially recreating, in a controlled setting, the conditions under which independent judgment would need to be exercised — presenting the worker with a task scenario stripped of algorithmic assistance and measuring the accuracy, speed, and quality of the resulting independent performance against a recorded baseline. The value of this approach lies in its direct engagement with actual performance rather than with a proxy for performance, but it carries the corresponding limitation that a simulated task, however carefully constructed, may not fully replicate the cognitive and situational demands of a genuine unassisted decision made under real professional stakes. The second approach, general cognitive tests, offers a complementary strategy by assessing underlying cognitive functions — such as reasoning, working memory, or attention — that are theorized to underlie the specific capacity for professional judgment, without requiring the worker to perform the actual professional task itself; this approach trades some domain specificity for greater standardization and ease of administration across different occupational contexts. The third approach, expert evaluation of real decisions, takes a different tack by having qualified peers or supervisors review actual decisions the worker has made in the course of real work, assessing the quality of the judgment exercised in situations that did arise organically rather than in an artificially constructed scenario; this approach benefits from ecological validity but depends heavily on the availability of a sufficient number of real cases in which the worker's independent judgment, rather than the algorithm's output, can be clearly identified and isolated for evaluation. The fourth approach, indirect behavioral indicators, relies on observable proxies for

underlying judgment capacity — such as patterns of hesitation, rates of unprompted deviation from algorithmic recommendations, or the frequency with which a worker seeks additional information before accepting a system-generated output — on the premise that such behavioral patterns may correlate with, even if they do not directly measure, the latent capacity of interest.

The fifth approach, which this paper explicitly proposes as a novel addition rather than a restatement of an existing measurement strategy, targets specifically the measurement of the capacity to detect rare system-failure events. This approach is singled out as a necessary addition precisely because of the argument developed in the preceding section: since EIPJ is hypothesized to become consequential above all at the moment when the algorithmic system itself fails or produces an anomalous output, a measurement strategy confined to ordinary task performance — however well-designed — risks entirely missing the dimension of the construct that matters most in practice, namely the worker's ability to recognize that something has gone wrong in the first place. A worker might perform adequately on a general simulation task or score well on a cognitive test while nonetheless having lost, through prolonged reliance on a system that rarely errs, the specific vigilance required to notice the rare occasion on which that system does err. By proposing anomaly-detection capacity as a distinct, fifth measurement target — rather than assuming it is adequately captured by the other four approaches — this paper seeks to ensure that any empirical program built around the EIPJ construct does not inadvertently measure everything except the dimension of the phenomenon that the theoretical argument in Section 2.6 identifies as most critical.

2.8. The Gerlich Reference: An Explicit Repositioning as Secondary Rather Than Central

The Gerlich study documented, in a survey of 666 participants, a statistically significant negative association between frequent use of generative AI tools and self-reported critical thinking skills, mediated by cognitive offloading. (Gerlich, 2025, p. 6)

On its face, this finding might appear to offer direct empirical support for the general thrust of the argument developed in this paper, since it links AI reliance to a self-reported decline in a cognitive capacity closely related to the kind of evaluative judgment at stake in the EIPJ construct, and does so through the specific mediating mechanism of cognitive offloading — the voluntary transfer of mental operations to an external tool that this paper has already identified, in Section 1, as one of the four established constructs adjacent to but distinct from EIPJ. A less self-critical treatment of the literature might therefore have been tempted to place the Gerlich study at the theoretical center of the paper's empirical case.

This paper explicitly repositions this reference methodologically: given that a formal publisher correction was issued regarding an error in one of its analytical tables, this paper downgrades the reference to the status of a secondary supporting source. (Gerlich, 2025, p. 252)

This repositioning is undertaken deliberately and transparently rather than passed over in silence, in keeping with the self-critical methodology the paper announces at the outset. The existence of a formal publisher correction — issued specifically in relation to an error in one of the study's analytical tables — is treated here as sufficient grounds to reduce the evidentiary weight the paper places on the study, even though the correction does not necessarily invalidate the study's overall conclusions in their entirety. This cautious response reflects a broader methodological stance adopted throughout the paper: rather than assuming that a corrected source remains fully reliable simply because its central finding was not formally retracted, the paper treats the correction as a signal warranting reduced reliance, downgrading the Gerlich study from what might otherwise

have served as a central empirical pillar to the status of a secondary supporting source — one that can still be cited as consistent with the paper's broader argument, but that is no longer treated as capable of bearing the full evidentiary weight of establishing the EIPJ mechanism on its own. Moreover, beyond the correction itself, the Gerlich study's context of voluntary personal use of generative AI tools, rather than institutionally imposed algorithmic management, already placed it at some remove from the specific phenomenon this paper is concerned with, as established in Section 2.2 — a further reason for treating it as supporting rather than central evidence.

The study by Wood and colleagues remains the closest available empirical reference to institutional algorithmic management, having documented in its extensive field study of digital platform economy workers that workers face continually shifting rules and regulations whose rationale they cannot understand, along with evaluations issued by automated systems that they can neither contest nor appeal. This study occupies a markedly different evidentiary position than the Gerlich study, precisely because it examines algorithmic management in the institutionally imposed sense that this paper's construct specifically targets, rather than the voluntary personal use of AI tools that characterizes the Gerlich study's context. The specific features documented by Wood and colleagues — rules that shift without transparent rationale, and evaluations issued by automated systems against which workers have no meaningful avenue of contestation or appeal — are directly relevant to the causal mechanism this paper proposes, since a worker operating under these conditions is structurally positioned in exactly the kind of substitutive algorithmic-management relationship that this paper hypothesizes as the antecedent of EIPJ, even though the Wood et al. study itself was not designed to measure judgment erosion directly and therefore cannot, on its own, confirm the specific causal pathway proposed here. Table 2 (presented in the Figures and Tables section) delineates the distinctions among the evidential contexts of these three sources: the Gerlich study, the Wood et al. study, and the algorithm aversion studies. (Wood et al., 2018, p. 60)

This table serves an important function within the paper's overall self-critical architecture, since it makes explicit, in a single comparative format, precisely what kind of evidential support each of the three sources can and cannot offer — distinguishing, for instance, between a source that can only establish the plausibility of a possible mechanism and a source that offers direct evidence of the institutional context the paper is actually concerned with — so that no reader of the paper could reasonably come away with the impression that any one of these three sources, individually or collectively, amounts to a confirmed empirical demonstration of the EIPJ construct as such.

2.9. Evidence That May Weaken the Hypothesis: Competing Explanations and Reverse Hypotheses

This paper carefully distinguishes between two fundamentally different categories of competing evidence: the first encompasses competing explanations for the same observed decline — agreeing on the existence of the decline while disagreeing on its cause; the second encompasses an independent reverse-outcome hypothesis that does not agree on the existence of a decline in the first place. This paper carefully distinguishes between two fundamentally different categories of competing evidence: the first encompasses competing explanations for the same observed decline — agreeing on the existence of the decline while disagreeing on its cause; the second encompasses an independent reverse-outcome hypothesis that does not agree on the existence of a decline in the first place. This distinction is treated with particular care because the two categories pose qualitatively different

threats to the EIPJ construct and therefore call for different kinds of empirical response. Competing explanations, belonging to the first category, do not challenge the paper's core observation that a decline in independent professional judgment occurs under sustained algorithmic management; rather, they challenge the paper's proposed account of why that decline occurs, offering alternative causal mechanisms that could, if correct, redirect explanatory credit away from the sustained-non-use pathway this paper hypothesizes. Reverse hypotheses, belonging to the second category, pose a more fundamental challenge, since they do not merely offer an alternative cause for an agreed-upon effect but instead dispute the existence of the effect itself, raising the possibility that what the paper interprets as erosion may in some contexts be better characterized as no decline at all, or even as a form of enhancement. Distinguishing these two categories explicitly, rather than treating all counter-evidence as a single undifferentiated threat, allows the paper to specify more precisely what kind of empirical finding would count against which part of its argument.

The first category includes the algorithm aversion phenomenon demonstrated by Dietvorst and colleagues, whereby many individuals resist relying on algorithms even when those algorithms demonstrably outperform human judgment statistically. (Dietvorst et al., 2014, p. 116)

This finding is relevant as a competing explanation because it points toward a countervailing psychological tendency that runs in the opposite direction from the substitutive reliance this paper's causal mechanism presupposes: if workers are, as a general psychological matter, disposed to resist algorithmic recommendations even in the face of demonstrated statistical superiority, then the depth and consistency of substitutive reliance required to drive the EIPJ pathway may be less pervasive across worker populations than the paper's core argument might otherwise suggest, and any observed decline in independent judgment among a given worker population might need to be explained through some mechanism other than straightforward substitutive dependence on the algorithm.

It also includes the problem of reverse causation, which proposes that individuals with lower baseline cognitive independence are those who are drawn to or directed toward cognitively intensive algorithmic roles. This alternative explanation is particularly consequential because it inverts the paper's assumed causal direction entirely: rather than algorithmic management causing a decline in independent judgment over time, this account holds that workers who already possessed comparatively lower independent judgment capacity before ever encountering the algorithmic system are the ones more likely to end up in, or be selected into, roles characterized by heavy algorithmic management, whether through self-selection based on personal comfort with such systems or through organizational sorting practices that channel workers toward roles matching their pre-existing capabilities. Under this account, any cross-sectional association between algorithmic management and lower independent judgment would reflect a pre-existing selection effect rather than a genuine erosion process unfolding over the course of the worker's exposure to the system. This objection can only be resolved through a longitudinal design that measures the worker's independent competence prior to entry into the algorithmic environment, since only a design of this kind — one that establishes each worker's baseline capacity before algorithmic exposure begins and then tracks that capacity over time — can distinguish a genuine within-person decline from a between-person selection effect that merely mimics the appearance of decline in cross-sectional data.

The second category encompasses the possibility that certain applications of algorithmic management enhance learning rather than impede it, through redistribution of tasks between the

human and the system. This possibility represents the more fundamental of the two categories of competing evidence discussed in this section, because it does not merely offer an alternative cause for an agreed decline but proposes instead that, under certain configurations of task allocation between human and algorithm, the net effect on the worker's independent judgment may run in the opposite direction from the one this paper hypothesizes, with the algorithmic system functioning to sharpen rather than dull the worker's capacity by concentrating the worker's attention specifically on the tasks best suited to human judgment. Recent analytical models of human–AI task allocation, empirically tested on image classification tasks, demonstrate that the optimal task configuration assigns simpler tasks entirely to automation, retains human–machine collaboration for moderately difficult tasks, and reserves the most demanding tasks for humans without any assistance. (Fügener et al., 2026, pp. 545–550)

This finding is instructive for the reverse-hypothesis category because it suggests that, under an optimally designed allocation scheme, a worker's independent judgment might be exercised disproportionately on the most cognitively demanding subset of tasks — precisely the tasks most likely to sustain and even strengthen evaluative judgment through continued challenging practice — while the simpler tasks most amenable to skill atrophy through disuse are the ones handed over entirely to automation. If this pattern of task allocation holds in a given algorithmic management context, the substitutive character that this paper's construct presupposes as a precondition for erosion may simply not be present, since the system would not be displacing the worker's opportunity to exercise judgment on demanding tasks but rather concentrating that opportunity more intensively than an unaided work environment might otherwise allow. This possibility stands as a genuine boundary condition on the EIPJ hypothesis, one that the paper does not attempt to explain away but instead incorporates directly into its account of when the proposed erosion pathway is and is not expected to operate.

3. AXIS TWO: RECONSTRUCTION, CAUSAL MECHANISM, AND CONDITIONS FOR CONSTRUCT SURVIVAL AND FALSIFICATION

3.1. The Narrowed Operational Definition

This paper defines Erosion of Independent Professional Judgment (EIPJ) as a measurable decline — confirmed through converging independent indicators — in the accuracy, speed, and breadth of independent professional solutions, subject to two conditions: the decline must be established through pre- and post-measurement relative to a recorded baseline, and it must persist following an intervention that alters the organizational incentive. This narrowed operational definition represents the culmination of the conceptual groundwork laid throughout Axis One, translating the various distinctions, exclusions, and boundary conditions developed in the preceding sections into a single, testable formulation. Each element of this definition is deliberately restrictive rather than expansive, reflecting the self-critical methodology the paper adopts throughout: the requirement that the decline be "measurable" and "confirmed through converging independent indicators" rules out reliance on any single measurement approach or any purely subjective self-report, instead demanding that a genuine instance of EIPJ be corroborated across multiple independent sources of evidence before it can be credibly claimed. The three dimensions specified — accuracy, speed, and breadth of independent professional solutions — together capture distinct facets of what a deterioration in professional judgment would actually look like in practice: accuracy addresses whether the worker's independent conclusions remain correct; speed addresses whether the worker can still arrive at those conclusions within a

reasonable timeframe without the crutch of algorithmic assistance; and breadth addresses whether the worker retains the capacity to generate a sufficiently wide range of possible solutions or interpretations, rather than defaulting narrowly to whatever the algorithmic system would have suggested. A definition confined to accuracy alone might overlook a worker who arrives at correct conclusions only after considerably longer deliberation than before, or one who arrives at a correct conclusion but has lost the capacity to consider alternative possibilities along the way — both of which would plausibly count as meaningful erosion under this paper's broader conception even if accuracy per se remained superficially intact.

The two subsidiary conditions attached to this definition function as methodological safeguards against precisely the kinds of false positives that the paper's exclusion criteria and competing-explanation analysis, developed earlier in Sections 2.5 and 2.9, were designed to guard against. The requirement that the decline be established through pre- and post-measurement relative to a recorded baseline directly addresses the reverse-causation concern discussed in Section 2.9, since only a design that captures the worker's competence before substantial algorithmic exposure, and again afterward, can distinguish a genuine within-person decline from a pre-existing difference between workers who happened to end up in more heavily algorithmically managed roles. The requirement that the decline persist following an intervention that alters the organizational incentive directly operationalizes the recoverability criterion introduced in Section 2.3 as the decisive test separating EIPJ from mere Algorithmic Dependence: if a worker's performance rebounds once the organization changes the incentive structure surrounding manual judgment — for instance, by rewarding independent decisions or penalizing uncritical reliance on the system — then, by this paper's own definitional criteria, the observed pattern does not qualify as EIPJ at all, but is more accurately described as dependence, a behavioral and motivational pattern rather than a genuine deterioration of latent capacity. Only a decline that survives this incentive-based intervention, remaining resistant to correction through motivational means alone, meets the threshold this paper sets for EIPJ proper.

The term "substitutive character" is understood operationally in this paper to mean that the system is designed such that a structurally viable manual intervention point exists but has not been exercised for an extended period. This operational definition of substitutive character connects directly to the structural intervention point criterion introduced earlier in Section 2.3 as one of the five distinguishing criteria in Table 1, and it is essential to the overall EIPJ framework because it is precisely this substitutive character — rather than algorithmic assistance in general — that this paper hypothesizes as the necessary precondition for the erosion pathway to operate. A system that merely supplements a worker's judgment without displacing the occasions on which that judgment would otherwise be exercised does not, on this definition, possess substitutive character, and would not be expected to generate the kind of erosion this paper's construct describes; it is specifically the combination of continued structural availability and sustained practical non-use that defines the substitutive relationship at the heart of the EIPJ mechanism.

The availability of this structural intervention point is assessed through three indicators proposed in this paper: technical availability, policy availability, and practical availability. These three indicators are proposed as jointly necessary, rather than individually sufficient, tests of whether a genuine structural intervention point exists in a given algorithmic management context, since a system could satisfy any one or two of these indicators while failing the third, and such a partial satisfaction would arguably not constitute the full structural availability the paper's

construct requires. Technical availability addresses whether the underlying infrastructure and tools needed for manual intervention remain physically or digitally present and functional within the system, independent of whether workers are permitted or encouraged to use them. Policy availability addresses whether the organization's formal rules, procedures, and governance documents explicitly permit — rather than prohibit or discourage — a worker's recourse to manual, unassisted judgment in place of or in addition to the algorithmic recommendation. Practical availability addresses whether, beyond the formal technical and policy conditions, the actual working environment realistically affords a worker sufficient time, information access, and institutional support to exercise manual intervention without incurring undue cost or penalty, since a system that is technically functional and formally permitted but practically unworkable — for instance, one in which manual review would require far more time than a worker's performance targets allow — cannot be said to genuinely offer a structural intervention point in the full operational sense this paper intends.

3.2. The Theoretical Causal Mechanism

This paper presents the proposed causal pathway for EIPJ in Figure 1 (presented in the Figures and Tables section), which illustrates the basic causal pathway from substitutive-character algorithmic management to the Erosion of Independent Professional Judgment. This causal pathway serves as the theoretical backbone connecting the conceptual groundwork laid out across the preceding sections of Axis One to the empirically testable definition just articulated in Section 3.1: it specifies not merely that EIPJ is hypothesized to occur under conditions of substitutive-character algorithmic management, but the sequence of intervening steps through which that antecedent condition is theorized to produce the ultimate outcome of judgment erosion. Rendering this pathway explicitly in figure form, rather than leaving the causal logic implicit across scattered sections of prose, serves an important methodological function for a paper whose central purpose is self-critical scrutiny of its own theoretical claims: a causal pathway that is made fully visible, with each proposed link stated openly, is thereby also made fully available for challenge, testing, and potential falsification, in a way that a causal argument left diffuse or merely implied across the discussion would not be.

No link in this causal pathway has been jointly confirmed in a single study as of the writing of this paper, and this constitutes an explicit constraint acknowledged by the paper. This acknowledgment is a direct continuation of the self-critical posture the paper has maintained throughout, and it carries particular weight at this specific juncture in the argument, since Figure 1 represents the paper's most complete and integrated theoretical claim — the point at which the various pieces of evidence, definitions, and distinctions assembled across Axis One are drawn together into a single proposed causal sequence. Precisely because this is the paper's most ambitious theoretical claim, it is also the claim most in need of an explicit evidentiary caveat: while individual segments of the proposed pathway can each draw some degree of support from the disparate literatures discussed earlier — the structural availability literature informing the substitutive-character premise, the Bainbridge-derived invisibility mechanism discussed in Section 2.6, the practice-interruption dynamics documented in the expertise and cognitive-reserve literatures discussed in Section 2.4 — no single existing study has traced the pathway in its entirety, from a documented instance of substitutive-character algorithmic management, through the hypothesized intervening mechanism, to a confirmed and independently verified decline in professional judgment meeting the criteria specified in Section 3.1. The pathway, in other words,

is assembled from empirically grounded but individually partial pieces of evidence rather than validated as an integrated whole by any single empirical investigation, and the paper is explicit that this assembly constitutes a theoretical proposal awaiting comprehensive empirical test, not a finding already established by the literature it draws upon. This constraint is offered not as a weakness to be minimized but as a precise specification of exactly what remains to be demonstrated — a specification that in turn defines the empirical agenda the construct requires before it can move from a proposed causal mechanism to a confirmed one.

3.3. Level of Analysis and Individual Differences

EIPJ operates at the individual level, while algorithmic management constitutes an antecedent organizational-level variable — rendering the framework inherently cross-level in nature and necessitating analytical models that reflect this nesting, such as Hierarchical Linear Models, which were designed specifically to analyze nested-structure data in which individual observations are nested within higher-order units such as teams or organizations, thereby avoiding the statistical error arising from violating the assumption of independence of observations required by conventional regression analyses. (Renaud, Lacroix, 2023, p. 127)

This observation about levels of analysis is not a peripheral methodological footnote but a structural feature of the EIPJ framework that follows directly from the individual-level operational definition established in Section 3.1, in which EIPJ was specified as a decline observed within a given worker relative to that worker's own recorded baseline, and from the antecedent identified throughout Axis One as an organizational-level condition — namely, the presence or absence of a substitutive-character algorithmic management system, which is typically implemented uniformly across an entire team, department, or organization rather than varying idiosyncratically from one individual worker to the next within that unit. This mismatch in levels — an organizational-level cause and an individual-level effect — is precisely the configuration that conventional single-level regression techniques are ill-equipped to handle, because workers nested within the same organizational unit share exposure to the same algorithmic management system and are therefore unlikely to constitute genuinely independent observations in the statistical sense that ordinary regression assumes; treating them as independent when they are not risks generating artificially inflated confidence in the resulting statistical estimates. Hierarchical Linear Models address this problem directly by explicitly modeling the nested structure of the data, allowing individual-level variation in EIPJ to be estimated simultaneously with organizational-level variation in the presence and character of algorithmic management, and thereby providing an analytical framework appropriately matched to the cross-level theoretical claim the paper advances — a claim that could not be adequately tested using a methodology that implicitly assumed a single, undifferentiated level of analysis.

This paper proposes that accumulated deep professional experience may function as a protective factor that slows or entirely prevents the proposed pathway — a proposition that does not invalidate the construct but rather narrows the research population most susceptible to it. This proposition introduces an important qualification to the causal pathway outlined in the preceding section, one that operates specifically at the individual level within the broader cross-level framework just described: even among workers exposed to an identical organizational-level condition of substitutive-character algorithmic management, the paper anticipates that individual differences in accumulated professional experience will produce meaningfully different outcomes, with more experienced workers potentially retaining their capacity for independent judgment for

a longer period, or even indefinitely, relative to less experienced colleagues operating under the same algorithmic system. The theoretical basis for this proposed protective effect connects back to the latent capacity literatures discussed earlier in Section 2.4, since a worker who has accumulated a substantial reservoir of deliberately practiced expertise and tacit, organizationally embedded knowledge prior to extensive algorithmic exposure may possess a deeper and more resilient foundation of latent capacity than a worker whose professional judgment was never given the opportunity to develop fully before algorithmic management became the dominant mode of task performance. Framing this proposition explicitly as a matter of narrowing the susceptible research population, rather than as evidence against the construct itself, is methodologically significant: it allows the paper to accommodate the plausible empirical possibility that experienced professionals show little or no measurable erosion under algorithmic management without treating such a null finding, among that particular subgroup, as a falsification of the construct as a whole. Instead, the construct's claims are appropriately understood as most applicable to workers with comparatively limited pre-existing depth of independent practice, while the question of precisely how much accumulated experience is sufficient to confer meaningful protection remains, as the paper's concluding section later specifies, an open matter for future empirical inquiry. (Raudenbush & Bryk, 2002)

3.4. Repositioning Algorithmic Transparency and the Ethics of Classification

This paper stipulates a methodological distinction: the question of the effectiveness of algorithmic transparency as a proposed remedial intervention is a question concerning the bounds of treatment rather than a gap in the conceptual construct itself. This distinction is introduced to preempt a potential conflation that could otherwise weaken the paper's overall argumentative structure: a reader might reasonably ask whether the mere existence of remedial interventions — such as making an algorithmic management system's decision logic more transparent to the workers subject to it — undermines the case for EIPJ as a genuine, independently standing construct, on the reasoning that if a simple design fix can prevent or reverse the phenomenon, then perhaps the phenomenon was never a deep structural problem to begin with. This paper explicitly forecloses that line of reasoning by relocating the question to a different conceptual register entirely. Whether transparency, or any other proposed intervention, proves effective at slowing, preventing, or reversing EIPJ is treated here as an empirical question about the boundaries and limits of possible treatment for a phenomenon whose existence and definition have already been established through the construct-building work carried out across Sections 2 and 3.1 through 3.3; it is not treated as a question bearing on whether the construct itself is coherently defined or conceptually warranted in the first place. A medical analogy is instructive here in spirit, though not offered by the paper itself: the discovery that a given condition responds to a particular treatment does not retroactively call into question whether that condition constitutes a genuine, independently identifiable phenomenon — treatability and conceptual validity are simply different questions, addressed through different kinds of evidence, and this paper insists on keeping them analytically separate rather than allowing evidence bearing on the former to be mistaken for evidence bearing on the latter.

The proposal to classify workers in research designs also raises a serious ethical problem, which this paper acknowledges explicitly: if organizations deploy such classification for dismissal, exclusion, or promotion denial, the research measurement instrument is converted into a discriminatory tool. This acknowledgment addresses a distinct but related concern that arises

directly from the measurement approaches proposed earlier in Section 2.7, since any of the five measurement strategies discussed there — particularly those involving direct performance assessment or expert evaluation of real decisions — would, if implemented within an actual organizational research design, necessarily generate some form of individual-level classification of workers according to their assessed degree of independent judgment erosion. The ethical hazard this paper identifies is that a measurement instrument developed for the scholarly purpose of validating or testing the EIPJ construct could, once it exists and is applied within a real organization, be repurposed by that organization for entirely different and potentially harmful ends — using a worker's classification along the EIPJ dimension as grounds for adverse employment decisions such as dismissal, exclusion from certain roles, or denial of promotion, rather than as a basis for the kind of supportive, remedial response that the research context originally intended. This risk is not merely hypothetical or incidental to the construct's development; it represents a genuine tension between the scientific value of measuring the phenomenon precisely enough to study it and the practical danger that any sufficiently precise measurement instrument, once created, becomes available for uses well beyond the researcher's original intent or control.

This concern is consistent with the conclusions of a recent editorial issued by a European research network specializing in occupational health and safety, which concluded that the harmful effects of algorithmic management systems can be mitigated when such systems are designed with transparency and fairness, and when workers are afforded genuine influence over the operation of the system itself. (Bowdler et al., 2026, p. 1–2)

This editorial's conclusions lend external support to the broader concern articulated in this section, in that they connect the mitigation of algorithmic management's harmful effects specifically to design features — transparency, fairness, and genuine worker influence over the system — that stand in direct contrast to the kind of unilateral, top-down classification scenario this paper warns against. A measurement and classification regime imposed on workers without their input or genuine influence, deployed for purposes of adverse organizational action rather than supportive intervention, would run precisely counter to the conditions this editorial identifies as necessary for mitigating algorithmic management's harms, reinforcing this paper's insistence that any practical application of the EIPJ construct within a real organizational setting must be accompanied by deliberate ethical safeguards rather than treated as a straightforward extension of the construct's scientific validation.

4. CONCLUSION

This paper has examined the principal structural gaps in the prior conceptual construction, explicitly distinguished EIPJ from four adjacent constructs, addressed the question of why a new construct is warranted rather than an update to Deskilling, and grounded the latent capacity concept across multiple bodies of literature with an explicit caveat regarding the degree of stability of that grounding.

EIPJ does not claim to account for all effects of algorithmic management on workers; rather, this paper proposes a narrowly scoped explanatory framework for a specific condition. The construct is classified in its present stage as an exploratory theoretical construct that has not yet been subjected to adequate longitudinal testing.

EIPJ retains its independent standing, according to the conditions advanced in this paper, only if future studies demonstrate empirical differentiation from Skill Decay in the pattern of recovery, and if the effect persists following changes in organizational incentive.

All evidence cited in this paper derives from Western — American and European — contexts. The phenomenon may differ substantially in cultures characterized by high power distance, where acceptance of algorithmic direction may be higher by default, potentially accelerating the proposed pathway. This constitutes a fundamental limit on any generalization beyond those contexts.

This paper proposes a future research agenda organized around three specific lines of inquiry: whether EIPJ manifests with equal intensity across different professional domains such as medicine, law, and engineering; whether accumulated deep professional experience functions as a genuine protective factor; and whether EIPJ can be reversed once established, through training programs specifically targeting anomaly detection.

5. Figures and Tables:

Table 1: Distinctions between EIPJ and Adjacent Conceptual Constructs

Construct	Affected Level	Primary Cause	Recoverability	Temporal Frame	Structural Intervention Point
Deskilling	Procedural structure of work as a whole	Task fragmentation and separation of thinking from execution	Very low; requires redesign of work	Cumulative, long-term	Structurally absent; work has been redesigned
Skill Decay	A specific individual skill	Interruption of direct practice	High; through resumption of practice	Weeks to months	Present; not used for a short period
Automation Complacency	Momentary attention and monitoring	Excessive confidence in system reliability	Very high; through reactivation of attention	Minutes to a single session	Present; the problem is its momentary non-use
Algorithmic Dependence	Behavioral usage pattern	Preference for system use to conserve effort	High; through motivational change alone	Variable; begins immediately	Present; latent capacity untested
EIPJ (this paper)	Latent capacity for independent judgment (exploratory theoretical synthesis)	Sustained non-use of a technically, policy-wise, and practically available intervention point	Relatively low; not immediately restored by motivational change alone	Months to years (operationally specified per domain)	Present in the tripartite sense, but not actually exercised for a sufficient period

Source: Author's own elaboration.

Table 2: Differentiation of Empirical Evidence Contexts.

Dimension	Gerlich Study (secondary source)	Wood et al. Study	Algorithm Aversion Studies
Nature of context	Voluntary personal use	Institutionally imposed organizational management	Individual high-stakes predictive decision
Sample type	General participants	Digital platform workers	Laboratory experiment participants
Study design	Cross-sectional correlational (with subsequent formal publisher correction)	Direct qualitative field study	Controlled laboratory experiment
Permissible inference	Supporting evidence for the plausibility of a possible mechanism only	Direct evidence of institutional algorithmic management	Pathway-attenuating evidence, conditional on task type

Source: Author's own elaboration.

Figure 1: The Proposed Basic Causal Pathway for EIPJ



Source: Author's own elaboration.

Bibliography List :

1. Attewell, P. (1987). The deskilling controversy. *Work and Occupations*, 14(3), 323–346.
2. Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779.
3. Bowdler, M., Lahti, H., Jelenko, M., Buresti, G., & Valtonen, T. (2026). Algorithmic management and psychosocial risks at work: An emerging occupational safety and health challenge. *Scandinavian Journal of Work, Environment and Health*, 52(1), 1–5. <https://doi.org/10.5271/sjweh.4270>
4. Braverman, H. (1974). *Labor and monopoly capital: The degradation of work in the twentieth century*. Monthly Review Press.
5. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2014). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 143(6), 114–126. <https://doi.org/10.1037/xge0000033>
6. Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363–406. <https://doi.org/10.1037/0033-295X.100.3.363>
7. Fügner, A., Walzner, D. D., & Gupta, A. (2026). Roles of artificial intelligence in collaboration with humans: Automation, augmentation, and the future of work. *Management Science*, 72(1), 538–557. <https://doi.org/10.1287/mnsc.2024.05684>.
8. Gerlich, M. (2025). Correction: AI tools in society. *Societies*, 15(9), 252
9. Grey, M., & Segerie, C.-R. (2025). *The AI risk spectrum: From dangerous capabilities to existential threats* (arXiv:2508.13700). arXiv. <https://arxiv.org/abs/2508.13700>
10. Meijerink, J., & Bondarouk, T. (2021). The duality of algorithmic management: Toward a research agenda on HRM algorithms, autonomy and value creation. *Human Resource Management Review*, 100876. <https://doi.org/10.1016/j.hrmr.2021.100876>
11. Nonaka, I. (1994). *A dynamic theory of organizational knowledge creation*. *Organization Science*, 5(1), 14–37.
12. Park, J., Kim, M., & Han, S. (2026). *The enrichment paradox: Critical capability thresholds and irreversible dependency in human–AI symbiosis*. arXiv. <https://arxiv.org/abs/2603.24391>
13. Polanyi, M. (1966). *The tacit dimension*. Doubleday.
14. Raudenbush, Stephen W., and Anthony S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Thousand Oaks, CA: Sage Publications, 2002
15. Renaud, C., & Lacroix, A. (2023). Systematic review of occupational burnout in relation to cognitive functions. *International Journal of Stress Management*, 30(2), 109–127.
16. Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, 8(4), 448–460. <https://doi.org/10.1017/S1355617701020240>
17. Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2018). Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. *Work, Employment and Society*, 33(1), 56–75. <https://doi.org/10.1177/0950017018785616>