



ISSN 2954-2757 (Online)

ESENSI: Jurnal Manajemen Bisnis

<https://esensijournal.com/index.php/esensi>

Vol. 29 Issue 1, April (2026)

doi.org/10.56943/esensi.v29i1.490

Legitimacy Debt: The Erosion of Professional Judgment in Human-AI Accounting

Gregorius Rudy Antonio^{1*}

gregorius@staff.ubaya.ac.id

Universitas Negeri Surabaya

*Corresponding Author: Gregorius Rudy Antonio

Email: gregorius@staff.ubaya.ac.id

ABSTRACT

Accounting firms increasingly report that AI adoption improves measurable performance, even as separate evidence points to a decline in the judgment capacity that professional licensure presupposes. This article develops a conceptual explanation for why both can be true at once. Following a theory-synthesis method, it integrates five theoretical traditions, human-AI symbiosis, virtue ethics, institutional theory, governmentality, and sustainability accounting, into a single mechanism. It introduces legitimacy debt, the accumulated gap between the judgment institutions still recognize as belonging to accountants and the judgment those accountants can still exercise, and presents a revised SIGMA model specifying the conditions under which this debt forms, stays concealed, and is eventually repaid. The article traces the mechanism through developmental displacement, in which task substitution erodes the practice that builds professional judgment, and through miscalibration under algorithmic visibility, in which both excessive trust and excessive discounting of AI output produce the same failure of proportionate judgment. Six falsifiable propositions link these mechanisms across individual, organizational, and profession-wide levels, with Indonesia examined as a boundary condition that reveals the double-edged role of collective deliberation. The article offers accounting research a testable construct for a phenomenon current frameworks describe only partially, and reframes integrity as a designable organizational infrastructure rather than an individual disposition, with direct implications for accounting education, firm practice, professional standards, and regulation.

Keywords: Accounting Ethics, Artificial Intelligence, Integrity Infrastructure, Legitimacy Debt, Professional Judgment, SIGMA Framework

Utang Legitimasi: Erosi Penilaian Profesional dalam Kolaborasi Manusia-AI di Bidang Akuntansi

ABSTRAK

Kantor akuntan publik semakin sering melaporkan bahwa adopsi kecerdasan buatan (AI) meningkatkan kinerja yang terukur, meskipun bukti lain menunjukkan penurunan kapasitas penilaian yang menjadi prasyarat lisensi profesi. Artikel ini menyusun penjelasan konseptual mengenai mengapa kedua hal tersebut dapat terjadi secara bersamaan. Dengan menggunakan metode sintesis teori, artikel ini mengintegrasikan lima tradisi teoretis, yaitu simbiosis manusia-AI, etika kebajikan, teori institusional, governmentality, dan akuntansi keberlanjutan, ke dalam satu mekanisme tunggal. Artikel ini memperkenalkan konsep utang legitimasi, yaitu kesenjangan yang terakumulasi antara penilaian yang masih diakui secara institusional sebagai kewenangan akuntan dan penilaian yang masih dapat dilaksanakan secara nyata oleh akuntan tersebut, serta menyajikan model SIGMA yang telah direvisi untuk menjelaskan kondisi-kondisi yang menyebabkan utang ini terbentuk, tetap tersembunyi, dan pada akhirnya harus dilunasi. Artikel ini menelusuri mekanisme tersebut melalui perpindahan perkembangan (developmental displacement), yaitu proses substitusi tugas yang mengikis praktik yang membangun penilaian profesional, serta melalui miscalibrasi akibat visibilitas algoritmik, yaitu kondisi ketika kepercayaan berlebihan maupun ketidakpercayaan berlebihan terhadap keluaran AI sama-sama menghasilkan kegagalan penilaian yang proporsional. Enam proposisi yang dapat diuji menghubungkan mekanisme-mekanisme tersebut pada tingkat individu, organisasi, dan profesi secara keseluruhan, dengan Indonesia dikaji sebagai kondisi batas yang mengungkap peran ganda deliberasi kolektif. Artikel ini menawarkan bagi penelitian akuntansi sebuah konstruk yang dapat diuji untuk fenomena yang selama ini hanya dijelaskan secara sebagian oleh kerangka kerja yang ada, serta memposisikan kembali integritas sebagai infrastruktur organisasi yang dapat dirancang, bukan semata-mata sebagai disposisi individu, dengan implikasi langsung bagi pendidikan akuntansi, praktik kantor akuntan, standar profesi, dan regulasi.

Kata Kunci: *Etika Akuntansi, Infrastruktur Integritas, Kecerdasan Buatan, Kerangka SIGMA, Penilaian Profesional, Utang Legitimasi*

INTRODUCTION

Accounting firms are investing heavily in artificial intelligence, and that investment is beginning to show measurable returns. Firms that adopt AI-based capabilities report fewer restatements, lower operating costs, and, over time, smaller audit teams (Fedyk et al., 2022). A recent synthesis of the accounting literature confirms that AI and machine learning are reconfiguring professional judgment across auditing, financial reporting, and management accounting, shifting accountants away from transaction processing and toward interpretation and oversight (Jones & Free, 2026). At the same time, a separate body of work finds that the opacity and scale of these systems introduce ethical demands that earlier audit technologies did not pose; AI tools can render audit evidence less visible even as they make audit work more efficient, leaving practice more exposed to scrutiny than before (Lehner et al., 2022; Seethamraju & Hecimovic, 2023). These two sets of findings are usually treated as opposite sides of a cost-benefit debate. This article treats them instead as two expressions of one underlying mechanism that has not yet been named.

The behavioral evidence sharpens this puzzle rather than resolving it. Some studies find that professionals defer too readily to AI output; others find the reverse. In an audit experiment involving complex estimates, auditors who received contradictory evidence from their firm's AI system proposed smaller adjustments than auditors who received the same evidence from a human specialist, consistent with algorithm aversion (Commerford et al., 2022). On more structured and measurable tasks, however, the same professionals tend to trust algorithmic output more than human judgment, consistent with algorithm appreciation; recent theoretical work shows that these opposing effects are not contradictions but predictable outcomes of how a task is structured and how errors are observed (Jussupow et al., 2024). The two effects point in opposite directions yet produce the same consequence: judgment that is no longer proportioned to the quality of the underlying evidence. This pattern is more plausibly attributed to a decline in calibration than to a decline in trust. This article terms the cumulative effect of that decline legitimacy debt: the widening distance between the judgment that institutions still formally recognize as belonging to accountants, and the experience, exposure, and calibrated skill that those accountants still retain to exercise it.

Three features distinguish legitimacy debt from a mere restatement of known problems. It accumulates gradually and is repaid, if at all, only after a long delay. It is concealed by the very success metrics organizations use to justify further automation, since output quality can continue to rise while the underlying capacity to evaluate that output quietly erodes. And it surfaces disproportionately in non-routine engagements, precisely the cases in which codified procedure runs out and professional judgment is what remains.

The underlying issue is not whether an algorithm is biased or opaque. It is whether the human who remains formally responsible for that algorithm's output

still has the capacity to evaluate it. Responsibility without capacity does not amount to accountability; it produces what has been described as a moral crumple zone, a structure in which humans absorb the blame for systemic failures they can no longer meaningfully control (Elish, 2019; Matthias, 2004; Santoni de Sio & van den Hoven, 2018). Four existing literatures approach this phenomenon, and none accounts for it fully. Work on algorithm aversion and appreciation explains a single decision, not a pattern that develops across a sequence of decisions over time. Work on deskilling documents that capability is being lost, but does not explain why quality assurance systems fail to detect the loss. Work on the responsibility gap identifies who should be held to account when an autonomous system errs, but is silent on why the accountable party loses the capacity to assess the error in the first place. Work on institutional decoupling explains why formal structures drift from actual practice, but remains normatively neutral and offers no theory of individual capacity. Legitimacy debt connects all four through a single temporal mechanism, and in doing so accounts for the anomaly that none of them can explain on its own: that measurable performance can rise at the very moment the underlying capacity for professional judgment is declining.

This article accordingly asks two questions. Under what institutional conditions does an increase in AI-driven technical capability reduce the capacity for professional judgment? And how do the resulting gaps accumulate, remain concealed, and eventually come to light? Its primary contribution is to frame legitimacy debt as a temporal stock construct rather than a single event. Its mechanistic contributions include developmental displacement, miscalibration under algorithmic visibility, the cost of calibrated dissent, the accountability-capacity divide, and integrity infrastructure as the cross-level condition that governs whether capacity translates into action. Its normative contribution is to specify moral human-AI symbiosis as the condition under which AI expands *techne* without foreclosing the formation and exercise of *phronesis*.

What this article does not claim should be stated at the outset. It does not oppose the use of AI in accounting. It does not assume that human judgment is inherently superior to algorithmic output. And it does not claim that AI is incapable of producing results that resemble, or even exceed, human judgment at the level of output. Its argument concerns institutional conditions specifically, because those conditions determine whether a gain in technical capability is accompanied by a corresponding loss in the capacity for professional judgment.

RESEARCH METHODOLOGY

Conceptual articles most often fail not from a lack of evidence but from a lack of discipline: theories get chosen without criteria, empirical findings get borrowed without a clear status, constructs get defined in terms of one another, and claims get phrased so loosely that they cannot be wrong. Conceptual contributions generally

take one of a few forms: introducing a new construct, linking constructs that have previously stood apart, or challenging an established explanation (Hulland, 2020; MacInnis, 2011). This article does the first two at once. It introduces a new construct, legitimacy debt, and links five previously separate theoretical traditions to explain how that construct forms. Jaakkola (2020) distinguishes four templates for conceptual work: theory synthesis, theory adaptation, typology, and model. This article is primarily a theory synthesis, with an element of theory adaptation folded in. A typology would not fit, because the phenomenon at hand is a sequence of causes unfolding over time rather than a set of variations in form. A model would not fit either, because the core construct does not yet exist and must first be built. Theory adaptation alone would also fall short, since no single existing theory can be transplanted into this domain and immediately explain the anomaly this article starts from. What the argument needs is a synthesis across traditions, and that synthesis requires exactly one adaptation: carrying the concept of symbiosis from the context of cognitive division of labor into the context of moral capacity formation, which changes what the term means. The consequences of this choice are binding. The article cannot claim to represent the literature as a whole, since it is not a systematic review, and it cannot claim empirical support. It claims only three things: that the mechanism is internally coherent, that its constructs can be told apart from one another, and that its propositions could turn out to be wrong.

Two kinds of theory need to be kept distinct here. Domain theory is the substantive knowledge base for the problem itself; method theory is the conceptual lens brought in from elsewhere to analyze it (Lukka & Vinnari, 2014). The domain theory in this article is knowledge about professional judgment, auditing, and technology adoption in accounting; its theoretical contribution belongs to that domain, not to the borrowed lenses. A single rule governs which theories are included: a tradition earns a place only if it fills an explanatory slot without which the mechanism cannot be stated. Five such slots exist, and Table I sets out the tradition assigned to each, along with what it contributes and where its limits lie.

Table 1 Five Theoretical Traditions, Explanatory Slots, Limitations, and Contributions to the Mechanism

Tradition	Explanatory Slot	Contribution to the Mechanism	Limitation
Human–AI symbiosis (Brynjolfsson & McAfee, 2014; Licklider, 1960)	Technical substrate	Establishes the cognitive division of labor; identifies which tasks are substituted (P1)	Treats moral capacity as inherited; remains silent on its erosion
Virtue ethics and phronesis (Kristjánsson et al., 2021; MacIntyre, 2007; West & Buckby, 2023)	Agency and normativity	Explains what exposure develops and what is lost in its absence (P2)	Moralistic when considered in isolation; does not explain the

Tradition	Explanatory Slot	Contribution to the Mechanism	Limitation
			institutional failure of virtuous agents
Institutional theory (Dimaggio & Powell, 1983; Meyer & Rowan, 1977; Scott, 2013)	Diffusion and legitimacy	Explains adoption pressures and ceremonial governance (P1, P5)	Normatively neutral; lacks a theory of individual capacity
Governmentality (Ahrens & Ferry, 2021; Miller & O’Leary, 1987; Miller & Rose, 1990)	Power and visibility	Explains asymmetries in the costs of dissent and subjectification (P3)	Does not provide a constructive basis for resistance
Sustainability accounting (Gray, 1992)	Telos and accountability horizon	Shows why sustainability mandates increase the demand for phronesis (P5)	Vulnerable to instrumental capture

Source: Analysis by the researchers (2026)

The same criterion governs exclusion. Several traditions common in the AI-and-accounting literature, including agency theory, technology-adoption models, and stakeholder theory, are left out because none adds explanatory power to the mechanism formulated here; including them would add authority without adding explanation, which is the definition of literature stacking.

Combining method theories from different traditions is not a neutral act, since each carries its own ontological and epistemological commitments, and doing so without reflection produces a synthesis that looks rich but is actually incoherent (Modell et al., 2017). The sharpest tension in this article sits at a single point and needs to be named rather than smoothed over. Virtue ethics is agentic and teleological, presupposing a self-forming subject oriented toward virtue (MacIntyre, 2007); governmentality is anti-essentialist and treats the subject as an effect of power relations, so that an exhortation to cultivate virtue can itself be read as a subtler form of subjugation. This article does not resolve that philosophical tension, and readers hoping for a resolution will not find one. What it does instead is work from the narrowest premises that both traditions share: that the capacity to judge is formed through practice within institutions, that institutions can support or undermine the conditions for that practice, and that acting against dominant pressure is possible, even at a cost. Virtue ethics calls this third premise virtue; the Foucauldian tradition calls it counter-conduct, the refusal to be governed by a given rationality (Senellart et al., 2009). Every mechanism in this article rests on these three shared premises and no stronger ones, so a reader who rejects one tradition outright can still evaluate the mechanisms, since no proposition here depends on the metaphysical truth of either side.

Six assumptions underlie the argument, stated openly so that each can be challenged individually. Professional judgment in non-routine cases is a capacity

rather than a stock of knowledge; it forms through practice and erodes when unused (A1). Ranking accounting tasks by codifiability produces an order that tracks how easily AI can substitute for them (A2). Organizations replace human deliberation with system output whenever that output is equivalent and cheaper, absent a clear institutional barrier (A3). Professional licenses, fiduciary duty, and legal liability remain with humans and do not transfer along with system output (A4). Quality assurance and inspection regimes evaluate documented evidence rather than undocumented capability (A5). AI capability continues to rise, so the boundary of what counts as codifiable keeps shifting, and the framework has to remain usable as that boundary moves (A6). A4 carries the most weight, and it is an institutional and legal claim rather than a computational one. A1 and A5 are the most exposed to disconfirmation, and both are testable; Section 10 sets out what follows if either fails.

The primary unit of analysis is the individual professional inside an organization, even though the constructs invoked here operate at different levels, and mixing levels is a common error in this literature: deskilling is often treated as an individual matter when organizational design in fact drives it, while decoupling is often treated as organizational when individual decisions in fact produce it. One rule holds throughout this article as a result: every proposition states the level at which its mechanism operates, and every transition between levels states the mechanism of that transition. The move from macro to meso runs through legitimacy pressure; from meso to micro, through organizational design, incentives, and the cost of dissent; from micro back to meso, through the repetition of behavior that hardens into routine; and from meso back to macro, through change in professional norms and standards. Legitimacy debt itself sits at the meso-macro level: it accumulates value across many commitments but is realized within a single one, and it is this separation between accumulation and realization that keeps it from collapsing into individual bias.

Empirical literature plays only three roles in this article. As premises, established findings must be explainable by the model, including that AI investment improves measurable audit quality and that auditors discount evidence generated by AI systems relative to evidence from human specialists (Commerford et al., 2022; Fedyk et al., 2022). As boundary conditions, findings constrain where the propositions apply, including the finding that whether a task is objective or subjective determines the direction of calibration error (Jussupow et al., 2024). And as institutional precedent, findings suggest that the mechanisms proposed here have already operated in this field around earlier technologies (Ahyaruddin et al., 2023; Harun et al., 2020). What empirical literature does not do in this article is confirm the framework: a conceptual argument is not validated by citing studies designed to test something else, and citing a correlation because it happens to reinforce a preferred conclusion is rhetoric rather than reasoning. Wherever empirical work appears in this article, its role is stated explicitly.

The article's central theoretical construct is a metaphor, and metaphors are not free of theoretical commitment; using one requires specifying how closely the correspondence holds and where it breaks down, or the result sounds deep without explaining anything (Cornelissen, 2005). Four features of debt do the explanatory work here. Small transactions accumulate into a larger total, and each one looks individually repayable. Because no reporting mechanism tracks the accumulation, the debt stays invisible until it is due. Each additional small transaction raises the eventual cost of recovery. And because restoring a single generation's capacity takes years rather than a fixed number of transactions, delay multiplies the cost of repair rather than merely postponing it. Non-routine events call the debt due at a time chosen by neither party, which points to features of debt beyond accumulation and delay: there is no creditor, no contract, no due date, and no interest rate, and no single payment discharges it, since rebuilding capacity takes a span of years rather than a single settlement. As a structure for explaining why these costs build up and why they are deferred, the metaphor is useful; it is not a substitute for making the underlying argument, and none of the article's conclusions rest on the metaphor rather than on the mechanism it illustrates.

Claims in this article fall into three kinds and are held to different standards accordingly. Definitional claims, such as the claim that calibrated dissent is not simply a frequency override, are judged by how clearly and distinctly the terms are defined. Causal claims, the propositions P1 through P6, are judged by whether they are consistent with existing evidence, and each is paired with a condition under which it could be shown false. Normative claims, addressed in Section 9, are judged by whether they follow logically from the mechanism, and are offered as design principles rather than as proven policy. At least three alternative explanations could in principle account for everything discussed here, and each is stated up front rather than left for a reviewer to raise. The first holds that what looks like decline is really just a change in the kind of skill accountants hold, as traditional skills give way to oversight skills; this framework predicts a shortfall specifically in non-routine cases that oversight skill cannot resolve, while the alternative predicts no such shortfall. The second holds that structured training and simulation can substitute for lived experience; if that is correct, P2 fails, for the reasons set out in Section 7. The third holds that judgment itself may simply be unnecessary once a system is reliable enough, since any loss of human capability is offset by a gain in efficiency; this article does not dispute that a system can be reliable, but points to A4: as long as licensing, fiduciary duty, and legal liability remain with humans, human judgment capacity remains a precondition for anyone to bear that liability at all. This third alternative, then, is not a refutation of the framework but a proposed change to the institution, and it needs to be argued for openly as a proposal rather than smuggled in as a statement of fact.

RESULT AND DISCUSSION

The Mechanism of Developmental Displacement

Aristotle distinguishes *techné*, skill judged by output, from *phronesis*, skill judged by fit to a situation (Aristotle, 2009). Current AI already matches or exceeds human performance on output-level tasks, but licensure, fiduciary duty, legal liability, and professional standing remain human (Fedyk et al., 2022). Output parity therefore does not mean parity in agency or accountability: humans stay answerable for decisions they no longer substantively make. *Phronesis* resists transfer because it forms only through embedded practice and carries motivational components that a system's output cannot replicate for the next person (Kristjánsson et al., 2021).

Codifiability, how easily a task becomes verifiable code, drives substitution, reinforced by cost pressure and isomorphic pressure to adopt AI (Dimaggio & Powell, 1983). Generative and agentic systems now push this boundary into tasks that once required judgment: assessing drafts, weighing evidence, building going-concern arguments. This matters because *phronesis* develops only through deliberate practice on judgment-requiring tasks, and those are exactly the tasks junior professionals lose first (Ericsson et al., 1993). Each substitution looks locally rational; the aggregate effect does not. This article terms that aggregate effect developmental displacement: the unintended loss of capacity-building exposure as a byproduct of individually sound decisions. Because *phronesis* is a capacity, not a store of knowledge, its erosion does not appear in measured output, since output is what gets measured. Alternative readings are worth naming: capacity might shift rather than shrink, or the change may be a different skill type rather than a loss. This article specifies the conditions under which its own prediction holds instead.

Adoption is driven as much by the need to appear legitimate as by efficiency (Dimaggio & Powell, 1983; Scott, 2013; Suchman, 1995). Firms build governance apparatus, policies, committees, model inventories, while substantive oversight capacity often stays thin underneath, the precondition for decoupling (Meyer & Rowan, 1977). The same pattern has already played out in this field with a different imported technology: coercive pressure drove IPSAS institutionalization in Indonesia ahead of implementation, and local government accountability practice was adopted for legitimacy while formal structure and actual practice stayed apart (Ahyaruddin et al., 2023; Harun et al., 2020). AI systems also log collaborative work in fine detail, reframing professionals as reviewers of output meant to assist them (Ahrens & Ferry, 2021; Miller & O'Leary, 1987; Miller & Rose, 1990). This creates an asymmetry: departing from a system's recommendation requires explanation; following it does not. This article calls that the cost of justified dissent, and it rises with measured accuracy even when accuracy is not the right basis for the specific case.

The behavioral literature shows two apparently opposite patterns as one problem. On measurable tasks, people over-trust algorithmic output; on subjective tasks, especially after seeing a system err, they discount it, and auditors receiving AI-generated contrary evidence propose smaller adjustments than those receiving the same evidence from a human specialist (Commerford et al., 2022; Jussupow et al., 2024). Both directions produce the same failure: judgment disproportionate to the evidence. What is needed is calibrated dissent, the ability to accept, question, or reject a recommendation in proportion to evidence quality, uncertainty, and case specifics, an ability no aggregate statistic can substitute for.

Integrity Infrastructure and the SIGMA Model

Codes of ethics define integrity as uprightness and honesty, too thin a definition for the weight the concept carries (Rogošić & Perica, 2023; West & Buckby, 2023). This article defines it independently: the disposition and action that sustains adherence to professional values under conflicting pressure. Integrity converts capacity into action; without it, capacity stays potential. It is not the same as override frequency, since overrides can be wrong, and frequent overriding under an aversion regime signals miscalibration rather than virtue. What matters is whether a decision to accept or reject aligns with the evidence in that case.

Personal integrity needs institutional channels to act through. This article introduces integrity infrastructure: the organizational structures, protected review time, competent reviewers, escalation rights, retaliation protection, documented disagreement, partner role-modeling, that let integrity operate rather than remain aspirational. This is distinct from AI governance quality, which concerns model control rather than conditions for human input; a firm can have a full model inventory and ethics committee while having no reviewer with time or authority to reject system output. The IESBA Code of Ethics revision, effective December 2024, names technology as an ethical risk source and requires managing complexity in judgment, but standards can only demand, not supply, time or protection, a structural problem recurring across AI accountability literatures (IESBA, 2023; Morley et al., 2021).

This article replaces an earlier, less testable “integrity as keystone” formulation with a mechanistic one: integrity is disposition at the individual level and infrastructure at the organizational level, connected by a transmission mechanism carrying capacity (micro) into dissent (meso) and design (meso) into behavior (micro). The relationship is interactive, not additive: capacity without infrastructure goes unused, infrastructure without capacity produces uncalibrated dissent. This claim fails if no interaction effect appears, or if governance quality predicts dissent just as well.

Legitimacy debt is the accumulated gap between the judgment institutions still recognize as belonging to accountants and the judgment they can still exercise. It has three dimensions: capacity (lost through developmental displacement),

authority (the accountability-capacity gap between formal responsibility and actual standing), and exposure (declining experience available to the next generation). It conceals itself because the same system that displaces deliberation generates documentation that looks like diligence, so inspection regimes assessing documentary traces never see the signal that should trigger correction. It is also deferred, showing no deficit in routine work until a non-routine case makes it visible. Sustainability assurance sharpens this: it expands accountability to non-contractual and intergenerational claims that computational methods cannot resolve for standing or materiality, so firms expanding sustainability assurance while replacing deliberation will show the widest gap between disclosure volume and assessment depth (Erin & Bamigboye, 2022; Gray, 1992).

Table 2 Legitimacy Debt Distinguished from Adjacent Concepts

Concept	Concise Definition	Unit of Analysis	Level	Mechanism	Outcome	Temporal Dimension	Limitation of Prior Concept	Contribution of Legitimacy Debt
Automation bias	Tendency to accept system output without adequate verification	Decision	Micro	Cognitive bias	Error in a single decision	Point in time	Does not explain accumulation or concealment	Positions it as one pathway toward capacity erosion
Algorithm aversion (Commerford et al., 2022; Dietvorst et al., 2015)	Tendency to discount algorithmic advice	Decision	Micro	Sensitivity to model error	Disproportionate rejection	Point in time	Appears contrary to automation bias	Unifies both as miscalibration
Algorithmic conformity	Adjustment of judgment to system recommendations due to institutional incentives	Decision/behavior	Micro-meso	Cost of dissent	Reduced deliberation	Flow	Explains behavior, not capacity	Links it to the loss of capacity
Professional deskilling	Loss of skills due to automation	Individual/job	Micro-meso	Nonuse of skills	Declining competence	Gradual	Does not explain why quality systems fail to detect it	Adds mechanisms of concealment and licensing burden
Responsibility gap (Matthias, 2004)	Absence of any party who can appropriately be held accountable for a learning system	Sociotechnical system	Macro	Unpredictability	Responsibility vacuum	Structural	Treats capacity as a state rather than something that erodes	Explains how that capacity is lost
Moral crumple zone (Elish, 2019)	Humans absorb blame for systems they cannot control	Human role	Micro-meso	Burden without control	Unfair responsibility	Structural	Descriptive of the final outcome	Explains the process through which it develops over time
Institutional decoupling	Formal structures are separated	Organization	Meso	Legitimacy compliance	Ceremonial compliance	Gradual	Normatively neutral; lacks an account of	Links decoupling to the erosion of

Concept	Concise Definition	Unit of Analysis	Level	Mechanism	Outcome	Temporal Dimension	Limitation of Prior Concept	Contribution of Legitimacy Debt
(Meyer & Rowan, 1977)	from actual practice						individual capacity	judgment capacity
Legitimacy debt	Accumulated gap between licensed judgment and the capacity to exercise it	Professionals within engagements; firms; profession	Meso—macro, temporal	Displacement + miscalibration + authority gap, concealed through documentation	Vulnerability to failure in non-routine engagements	Accumulated and deferred stock	—	Explains improving measured performance while capacity declines

Source: Analysed by the researcher (2026)

The revised SIGMA model treats its five dimensions as conditions rather than traits. Symbiosis (S) is the aspirational relationship where AI extends techne without displacing phronesis. Integrity infrastructure (I) is the transmission mechanism above; note that model control can be high even where I is low, and that combination increases debt rather than reducing it. Moral sustainability orientation (M) raises the demand on phronesis by giving standing to non-contractual and intergenerational claims. Accountability alignment (A) is the fit between formal responsibility and actual authority, information, and time; where it fails, the accountability-capacity gap opens. information and review time on one side and professional capability on the other; where that alignment fails, an accountability-capacity gap opens.

SIGMA pathway diagram — reproduced from the original figure. Main pathway: codifiability raises substitution pressure; substitution reduces capacity-building exposure; capacity loss combines with visibility-driven calibration error to reduce calibrated dissent; falling dissent widens the accountability-capacity gap; the gap accumulates into legitimacy debt; debt produces vulnerability in non-routine engagements. Two paths interrupt this: integrity infrastructure slows displacement and lowers the cost of dissent; algorithmic documentation delays detection. Table 3 efines the constructs and their observable indicators and Table 4 summarizes the six propositions before they are stated individually.

Figure 1. Legitimacy debt in human-AI accounting work: a cross-level process model

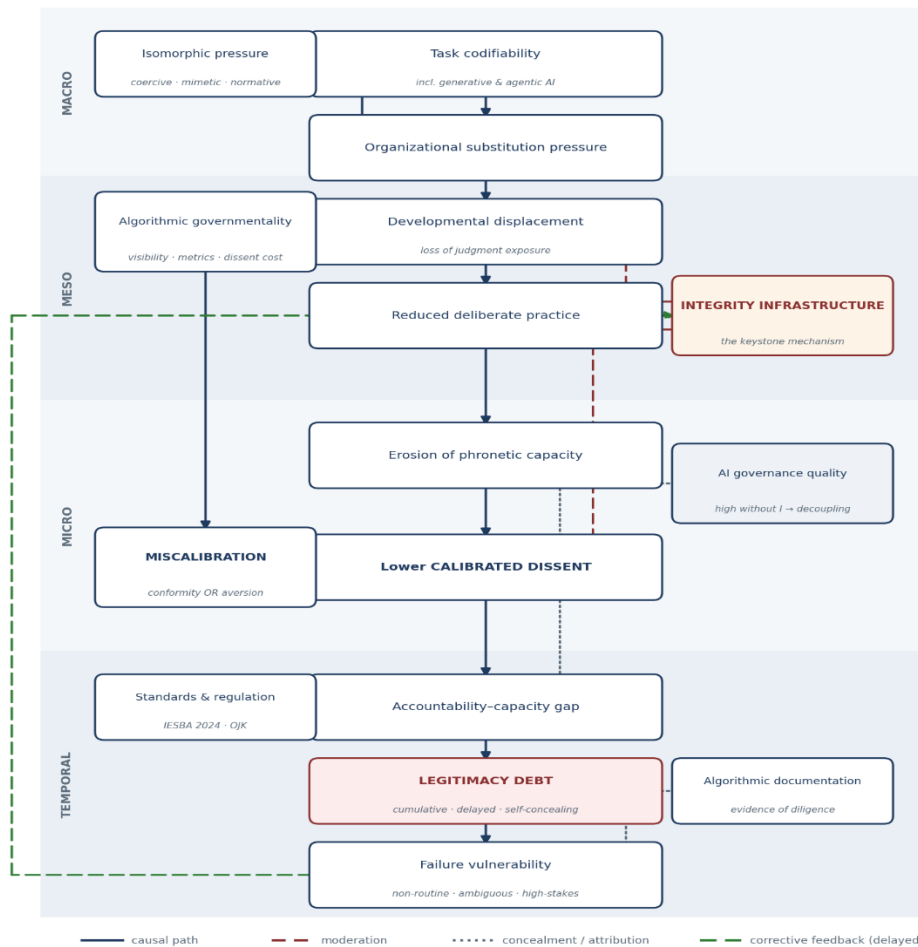


Figure 1 Legitimacy Debt in Human-AI Accounting Work: A Cross-Level Process Model

Source: Processed by the researcher (2026)

Table 3 Construct Definitions, Levels, and Observable Indicators

Construct	Definition	Level	Empirical indicators
Task codifiability	The extent to which a task can be expressed as a machine-learnable, verifiable pattern	Meso	Proportion of AI-assisted procedures; classification of task types
Developmental exposure	Repeated exposure to consequential judgment tasks with feedback	Micro–meso	Hours on judgment tasks by grade; mentoring ratio; composition of junior tasks
Phronetic capacity	The ability to recognize, weigh, and act on non-routine situations involving ambiguity, value	Micro	Performance on embedded ambiguous cases; quality of

Construct	Definition	Level	Empirical indicators
	conflict, incomplete information, and public-interest consequences		reasoning rather than answer conformity
Individual professional integrity	The disposition and action to maintain commitment to professional values under opposing pressure	Micro	Behavior in costly vignettes; actual escalation
Integrity infrastructure	Structures that enable integrity to be enacted substantively	Meso	Review-time budgets; reviewer grade and competence; escalation pathways; non-retaliation protection; quality review of overrides
AI governance quality	Completeness and effectiveness of model oversight	Meso–macro	Model inventory; validation; monitoring; follow-up
Algorithmic conformity / miscalibration	A response to system recommendations disproportionate to the quality of the evidence	Micro	Acceptance rate; review time; judgment changes following recommendations
Calibrated dissent	The ability to accept, question, or reject recommendations in proportion to evidence, model uncertainty, materiality, and case characteristics	Micro	Alignment of decisions with case truth, not override frequency
Accountability–capacity gap	The difference between formal responsibility and authority and actual capacity, time, information, and competence	Meso	Ratio of sign-off responsibility to substantive review hours; signatory competence regarding the system
Legitimacy debt	Accumulated gap between licensed judgment and the capacity to exercise it	Meso–macro, temporal	Performance in ambiguous tail cases; restatements; going-concern failures; depth of sustainability assurance
Moral sustainability orientation	The extent to which non-contractual and intergenerational claims	Meso–macro	Depth of assurance relative to disclosure volume

Construct	Definition	Level	Empirical indicators
	receive substantive standing		

Source: Processed by the researcher (2026)

Table 4 Summary of Propositions, Mechanisms, Levels, Moderators, and Empirical Designs

#	Proposition	Mechanism	Level	Primary moderator	Empirical design	Falsifying condition
P 1	Codifiability → substitution of deliberation	Local rationality + isomorphic pressure	Meso	Engagement economics; model reliability; regulatory capacity	Audit software logs; composition of hours by grade	Substitution does not increase despite high codifiability
P 2	Substitution → developmental displacement → erosion of phronetic capacity	Loss of deliberate practice	Meso→micro	Replacement training; mentoring; rotation	Cohort study; embedded ambiguous cases	Post-AI cohorts are not weaker on ambiguous cases
P 3	Algorithmic visibility → miscalibration (conformity or aversion)	Cost of justified dissent	Macro→micro	Task subjectivity; opacity; published accuracy	2×2 vignette experiment; alignment with case truth	Calibration persists despite high visibility
P 4	Integrity infrastructure moderates capacity → dissent	Cross-level transmission	Meso×micro	Hierarchy; partner role modeling; engagement economics	Cross-office field study; interaction test	Additive rather than interactive effects; or AI governance quality predicts equally well
P 5	The gap accumulates into legitimacy debt; delayed realization	Accumulation + concealment through documentation	Meso–macro, temporal	Portfolio composition; inspection regime	Longitudinal archives; lag structure t+2...t+5	No delayed effect; failures instead decline
P 6	Collective deliberation	Psychological safety vs.	Meso	Hierarchy; dissent	Within-office team	Deliberation improves

#	Proposition	Mechanism	Level	Primary moderator	Empirical design	Falsifying condition
	strengthens infrastructure only when minority dissent is protected	consensus pressure		documentation protocols	comparison ; ethnography	calibration without procedural protection

Source: Processed by the researcher (2026)

1. P1. Task Substitution

Higher codifiability and measurability raise the likelihood that AI output replaces human deliberation, reinforced by isomorphic pressure to adopt (Dimaggio & Powell, 1983). Once output is equivalent at lower cost, substitution needs no justification; not substituting does. Agentic systems now push codifiability into judgment-requiring tasks, strengthening this proposition. Test: audit software logs of hours by grade and AI-assisted procedure share, before and after rollout. Refuted if substitution fails to rise despite high codifiability.

2. P2. Developmental Displacement

Substituting tasks that once provided developmental exposure reduces deliberate practice and weakens phronetic capacity, hitting juniors hardest, since capacity forms through repeated high-stakes exposure with feedback, not instruction, and entry-level work is both the source of that exposure and the cheapest to automate (Ericsson et al., 1993). Effects are seniority-dependent, so damage stays hidden for a generation. Test: cohort comparison on planted ambiguous cases, pre- and post-adoption. Refuted if post-AI cohorts show no weakness. The sharpest threat: if simulated training builds capacity as effectively as real exposure, this proposition fails.

3. P3. Miscalibration Under Algorithmic Visibility

Transparency and conformity-rewarding metrics raise the perceived cost of dissent; direction depends on task type, conformity on objective tasks, aversion on subjective ones, but both reduce calibrated dissent (Commerford et al., 2022; Jussupow et al., 2024). Test: 2×2 vignette experiment (advice source × task nature), scored against case truth rather than override frequency. Refuted if calibration holds despite high visibility and metric pressure.

4. P4. Integrity Infrastructure as Transmission Mechanism

Infrastructure moderates the capacity-to-dissent relationship by supplying time, competence, authority, and protection; capacity without it goes unused, infrastructure without capacity produces uncalibrated dissent. This predicts a complementarity, not a main effect. Test: cross-office field study of the capacity × infrastructure interaction on dissent quality. Refuted if effects are additive, or if governance quality predicts dissent equally well.

5. P5. Accumulation of Legitimacy Debt

As responsibility stays human while capacity, authority, and review capacity decline, the gap accumulates as legitimacy debt, realized mainly in non-routine engagements and delayed by documentation serving as evidence of diligence. Test: archival relationship between substitution intensity at t and tail failures at $t+2$ through $t+5$. Refuted if no delayed effect appears, or if failures fall as substitution rises.

6. P6. Collective Deliberation and Protected Dissent

Deliberative practice strengthens infrastructure only with procedural protection for minority dissent, epistemic diversity, and real consequences for disagreement; without these, it reinforces miscalibration through ceremonial consensus. The mechanism runs through psychological safety, though strong hierarchy and consensus pressure can suppress the same voices calibration depends on (Edmondson, 1999). Test: within-office comparison of teams with and without a formal dissent procedure, psychological safety as mediator. Falsified if deliberation improves calibration without procedural safeguards.

Boundary Conditions: The Indonesian Context

Indonesia functions here as a boundary condition, not an illustration of a universal claim or a cultural reservoir assumed to offer automatic protection. Adoption pressure is strong and largely coercive-mimetic; institutional capacity is uneven, so pressure regularly outruns oversight capacity, the exact condition P5 predicts and the pattern already documented in IPSAS institutionalization and local government accountability (Ahyaruddin et al., 2023; Harun et al., 2020). Collective deliberation is also an established repertoire, and this is the theoretically interesting feature: it can distribute judgment and lower the cost of dissent, or it can reinforce hierarchy and collapse into ceremonial consensus that erases minority voices. Cultural values can function as instruments of control rather than moral resources, as shown directly in Indonesian organizations (Efferin & Hopper, 2007). The resulting formulation extends beyond Indonesia: collective deliberation strengthens infrastructure only when paired with procedural protection for dissent, epistemic diversity, and real consequences for disagreement.

Implications for Practice, Education, and Regulation

If legitimacy debt arises because deliberation gets replaced by outcomes, interventions must target the conditions for deliberation, not the content of ethics instruction. For education, additive responses, an analytics course, an AI module, train techne, which is no longer scarce; what curricula need instead is AI-assisted case analysis with the tools students will actually use, adversarial review scoring reasoning against plausible-but-wrong output, genuinely ambiguous cases requiring defended judgment before peers, and escalation simulations with real social

discomfort and protection for minority positions (Kristjánsson et al., 2021). For firms, the design question is what it costs to declare system output wrong: protected review time, competency-based sign-off, a dissent protocol demanding reasoning rather than punishing deviation, retaliation protection, quality review of overrides rather than counting them, and partner role-modeling. For standard-setters, override competence, evaluating AI output, identifying reasoning failures, building a substantive basis for departure, needs treatment as a testable skill distinct from AI literacy; the IESBA revision already moves this direction, though it supplies no time or protection on its own, making firm-level governance a necessary complement (IESBA, 2023). For regulators: regulate on evidence of feasible oversight rather than documentation; treat override rate as diagnostic rather than as proof of anything, since an extremely low rate on complex work should prompt scrutiny of review capacity rather than be read as success; and align sustainability assurance requirements with the judgment demand they create.

CONCLUSION

This article set out to answer a question the accounting profession has largely avoided: not how much work AI can perform, but what happens to the capacity that justifies professional licensure once the tasks that built that capacity no longer exist, while formal responsibility for outcomes remains with the human professional. Working through a synthesis of five theoretical traditions, human-AI symbiosis, virtue ethics, institutional theory, governmentality, and sustainability accounting, this article introduced legitimacy debt as the construct that names this gap, and the revised SIGMA model as the architecture that specifies the conditions under which the gap forms, stays hidden, or gets repaid. Six testable propositions traced the mechanism from task substitution through developmental displacement, miscalibration under algorithmic visibility, and accumulation into a debt that surfaces only in non-routine, high-stakes engagements.

The article's central contribution is to reframe an apparent contradiction in the existing literature, rising measured performance alongside eroding professional judgment, as the predictable output of a single, nameable mechanism rather than as two unrelated trends. This reframing carries direct implications for practice. It shifts the object of governance from documentation and compliance toward integrity infrastructure: the organizational conditions, protected review time, competency-based sign-off, protection against retaliation, that determine whether professional capacity, where it still exists, is actually exercised. For accounting education, firms, standard-setters, and regulators alike, the implication is the same: interventions aimed at preserving ethical conduct will fail if they target individual disposition alone and ignore the structural conditions for dissent.

This framework is conceptual, and its propositions await empirical testing; the constructs proposed here, phronetic capacity and calibrated dissent in particular,

do not yet have validated measurement instruments, and the claims regarding collective deliberation were developed from a single institutional context that requires comparative testing before wider generalization. Future research should prioritize the cohort and longitudinal designs specified in the propositions above, particularly the lagged relationship between substitution intensity and tail-risk failures, since that lag structure is where this framework predicts the debt will surface first. Until such evidence accumulates, the argument advanced here stands as a call to treat the erosion of professional judgment as a governable institutional risk, not as an unavoidable cost of technological progress.

REFERENCES

- Ahrens, T., & Ferry, L. (2021). Accounting and accountability practices in times of crisis: a Foucauldian perspective on the UK government's response to COVID-19 for England. *Accounting, Auditing & Accountability Journal*, 34(6), 1332–1344. <https://doi.org/10.1108/AAAJ-07-2020-4659>
- Ahyaruddin, M., Yusoff, M. N. H. bin, & Zainuddin, S. A. binti. (2023). *Institutional Isomorphism, Accountability, and Local Government Performance in Era of Public Governance: A Conceptual Framework* (pp. 623–633). https://doi.org/10.1007/978-3-031-27296-7_57
- Aristotle. (2009). *The Nicomachean Ethics* (L. Brown & D. Ross (eds.)). Oxford University Press.
- Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age*. W. W. Norton & Company.
- Commerford, B. P., Dennis, S. A., Joe, J. R., & Ulla, J. W. (2022). Man Versus Machine: Complex Estimates and Auditor Reliance on Artificial Intelligence. *Journal of Accounting Research*, 60(1), 171–201. <https://doi.org/10.1111/1475-679X.12407>
- Cornelissen, J. P. (2005). Beyond Compare: Metaphor In Organization Theory. *Academy of Management Review*, 30(4), 751–764. <https://doi.org/10.5465/amr.2005.18378876>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
- Dimaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48, 147–160. <https://doi.org/10.2307/2095101>
- Edmondson, A. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly*, 44(2), 350–383. <https://doi.org/10.2307/2666999>
- Efferin, S., & Hopper, T. (2007). Management control, culture and ethnicity in a Chinese Indonesian company. *Accounting, Organizations and Society*, 32(3), 223–262. <https://doi.org/10.1016/j.aos.2006.03.009>
- Elish, M. C. (2019). Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction. *Engaging Science, Technology, and Society*, 5, 40–60. <https://doi.org/10.17351/ests2019.260>
- 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>
- Erin, O. A., & Bamigboye, O. A. (2022). Evaluation and analysis of SDG reporting: evidence from Africa. *Journal of Accounting & Organizational Change*, 18(3), 369–396. <https://doi.org/10.1108/JAOC-02-2020-0025>
- Fedyk, A., Hodson, J., Khimich, N., & Fedyk, T. (2022). Is artificial intelligence improving the audit process? *Review of Accounting Studies*, 27(3), 938–985. <https://doi.org/10.1007/s11142-022-09697-x>
- Gray, R. (1992). Accounting and environmentalism: An exploration of the challenge of gently accounting for accountability, transparency and sustainability. *Accounting, Organizations and Society*, 17(5), 399–425. [https://doi.org/10.1016/0361-3682\(92\)90038-T](https://doi.org/10.1016/0361-3682(92)90038-T)
- Harun, H., Eggleton, I. R. C., & Locke, S. (2020). An integrated model of IPSAS institutionalisation in Indonesia: a critical study. *International Journal of Public Sector Management*, 34(2), 155–170. <https://doi.org/10.1108/IJPSM-07-2020-0195>
- Hulland, J. (2020). Conceptual review papers: revisiting existing research to develop and refine theory. *AMS Review*, 10(1–2), 27–35. <https://doi.org/10.1007/s13162-020-00168-7>
- IESBA. (2023). *Final Pronouncement: Technology-related Revisions to the Code*. The International Federation of Accountants (IFAC). <https://www.ethicsboard.org/publications/final-pronouncement-technology-related-revisions-code>
- Jaakkola, E. (2020). Designing conceptual articles: four approaches. *AMS Review*, 10(1), 18–26.
- Jones, S., & Free, C. (2026). What accountants need to know about artificial intelligence and machine learning: a review and call for future research. *Journal of Accounting Literature*, 48(5), 133–158. <https://doi.org/10.1108/JAL-12-2025-0693>
- Jussupow, E., Benbasat, I., & Heinzl, A. (2024). An Integrative Perspective on Algorithm Aversion and Appreciation in Decision-Making. *MIS Quarterly*, 48(4), 1575–1590. <https://doi.org/10.25300/MISQ/2024/18512>
- Kristjánsson, K., Fowers, B., Darnell, C., & Pollard, D. (2021). Phronesis (Practical Wisdom) as a Type of Contextual Integrative Thinking. *Review of General Psychology*, 25(3), 239–257. <https://doi.org/10.1177/10892680211023063>
- Lehner, O. M., Ittonen, K., Silvola, H., Ström, E., & Wührleitner, A. (2022). Artificial intelligence based decision-making in accounting and auditing: ethical challenges and normative thinking. *Accounting, Auditing & Accountability Journal*, 35(9), 109–135. <https://doi.org/10.1108/AAAJ-09-2020-4934>
- Licklider, J. C. R. (1960). Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, HFE-1(1), 4–11. <https://doi.org/10.1109/THFE2.1960.4503259>
- Lukka, K., & Vinnari, E. (2014). Domain theory and method theory in management accounting research. *Accounting, Auditing & Accountability Journal*, 27(8), 1308–1338. <https://doi.org/10.1108/AAAJ-03-2013-1265>
- MacInnis, D. J. (2011). A Framework for Conceptual Contributions in Marketing. *Journal of Marketing*, 75(4), 136–154. <https://doi.org/10.1509/jmkg.75.4.136>

- MacIntyre, A. C. (2007). *After virtue: A study in moral theory*. University of Notre Dame Press.
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175–183. <https://doi.org/10.1007/s10676-004-3422-1>
- Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, 83(2), 340–363. <https://doi.org/10.1086/226550>
- Miller, P., & O’Leary, T. (1987). Accounting and the construction of the governable person. *Accounting, Organizations and Society*, 12(3), 235–265. [https://doi.org/10.1016/0361-3682\(87\)90039-0](https://doi.org/10.1016/0361-3682(87)90039-0)
- Miller, P., & Rose, N. (1990). Governing economic life. *Economy and Society*, 19(1), 1–31. <https://doi.org/10.1080/030851490000000001>
- Modell, S., Vinnari, E., & Lukka, K. (2017). On the virtues and vices of combining theories: The case of institutional and actor-network theories in accounting research. *Accounting, Organizations and Society*, 60, 62–78. <https://doi.org/10.1016/j.aos.2017.06.005>
- Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a Service: A Pragmatic Operationalisation of AI Ethics. *Minds and Machines*, 31(2), 239–256. <https://doi.org/10.1007/s11023-021-09563-w>
- Rogošić, A., & Perica, I. (2023). Affective professional commitment and accounting ethics principles: examining the mediating role of the code of ethics. *Economic Research-Ekonomska Istraživanja*, 36(1), 480–498. <https://doi.org/10.1080/1331677X.2022.2077791>
- Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*, 5. <https://doi.org/10.3389/frobt.2018.00015>
- Scott, W. R. (2013). *Institutions and organizations: Ideas, interests, and identities* (Fourth Edition). SAGE Publications, Inc.
- Seethamraju, R., & Hecimovic, A. (2023). Adoption of artificial intelligence in auditing: An exploratory study. *Australian Journal of Management*, 48(4), 780–800. <https://doi.org/10.1177/03128962221108440>
- Senellart, M., Ewald, F., & Fontana, A. (2009). *Security, Territory, Population*. Palgrave Macmillan UK. <https://doi.org/10.1057/9780230245075>
- Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *The Academy of Management Review*, 20(3), 571. <https://doi.org/10.2307/258788>
- West, A., & Buckby, S. (2023). Professional judgement in accounting and Aristotelian practical wisdom. *Accounting, Auditing & Accountability Journal*, 36(1), 120–145. <https://doi.org/10.1108/AAAJ-09-2020-4949>