The Acceleration of Institutionalization: Epistemic Drift and the Infrastructure of Judgment in AI-Augmented Work

Dr Suela Pirushi

Acknowledgments. *The author thanks the professionals who participated in both the observation and field phases of this research. Supporting materials, coding frameworks, and supplementary analyses are registered and openly archived at https://osf.io/am2yr (DOI: 10.17605/OSF.IO/AM2YR).* *Correspondence: suelapirushi@yahoo.com*

Abstract

Organisational theory assumes that institutionalization unfolds slowly, through years of normative sedimentation. This study suggests that, under specific organisational conditions, digital infrastructure may compress the institutionalization of poorly founded knowledge claims to approximately six weeks in AI-augmented professional services. Through a 22-week process study across three organisational units (498 coded claims, 64 interviews, and 15 process-traced claims), it traces a four-stage drift sequence through which AI-generated claims may acquire institutional endorsement through repetition rather than verification. The findings point to three infrastructural mechanisms that interrupt drift at different stages: epistemic surfacing through provenance, anticipatory verification through accountability, and legitimised contestation. A mediation analysis indicated that approximately 71% of the governance effect operated through challenge behaviour, which is consistent with the interpretation that infrastructure does not substitute for professional judgment but creates the conditions under which judgment can be exercised. The study contributes a theory of micro-institutionalization specifying how digital workflows may compress the sedimentation process, and develops the concept of infrastructure as a condition of possibility for situated professional agency under algorithmic co-production. Given the small number of units, findings are interpreted as patterns consistent with the theorised mechanisms rather than as statistically generalisable effects, and replication across settings is recommended.

Keywords: *micro-institutionalization, epistemic drift, governance infrastructure, AI governance, professional services*

Introduction

Organisational theory has traditionally understood professional credibility as depending on either structural controls or individual expertise. Structural accounts emphasise verification procedures, quality systems, and hierarchical review as the mechanisms that ensure knowledge quality. Agentic accounts emphasise professional judgment, domain expertise, and individual epistemic virtue. This dichotomy has produced a productive but unresolved tension in organisation theory, with structuration theory offering an uneasy synthesis that privileges neither pole (Giddens, 1984; Barley & Tolbert, 1997).

The advent of generative artificial intelligence in knowledge-intensive work exposes the limits of both positions. AI co-production creates what we term the paradox of algorithmic credibility: structural controls falter because AI outputs resist conventional verification, being fluent, authoritative, and epistemically opaque; individual expertise falters because the surface cues professionals have historically relied upon to assess credibility carry no epistemic information when AI generates them. Neither more structure nor better people resolve the problem on its own. Yet some organisations maintain quality while others experience progressive erosion. How?

This puzzle has a temporal dimension that existing theory does not readily accommodate. Institutional theory posits that credibility emerges through slow sedimentation, the gradual normative consolidation of practices into taken-for-granted routines over years or decades (Tolbert & Zucker, 1996; Powell & Colyvas, 2008). Yet prior qualitative research across six professional services organisations suggests that under AI co-production, poorly founded claims may acquire institutional endorsement within weeks, through digital workflows that compress the cycle from provisional content to organisational resource. If institutionalization can be accelerated by an order of magnitude, existing theory lacks the conceptual apparatus to explain when credibility erodes, where it can be arrested, and what makes arrest possible.

This study addresses these questions through a 22-week process study across three organisational units in two professional services firms. The design combines process tracing of 15 knowledge claims through their complete organisational lifecycle with embedded theoretical probes that help disentangle the mechanisms through which

infrastructure shapes epistemic practice. The findings make two contributions. First, they develop a theory of micro-institutionalization: the rapid hardening of provisional knowledge into organisational artifacts, enabled by digital infrastructure that may compress the sedimentation process from years to weeks. Second, they describe a mode of organisational action in which infrastructure creates the conditions of possibility for professional judgment without determining its content, offering one specification of the structure–agency relationship under algorithmic co-production.

Theoretical Framework

The Structure–Agency Impasse in Knowledge Governance

The debate between structural and agentic accounts of organisational knowledge quality has deep roots. Structural perspectives emphasise how formal procedures, hierarchical review, and verification systems constrain action to produce reliable outcomes (Scott, 2001; Adler & Borys, 1996). Agentic perspectives emphasise how professional expertise, situated judgment, and epistemic virtue enable practitioners to navigate uncertainty (Emirbayer & Mische, 1998; Bandura, 2001). Structuration theory attempted to transcend this dichotomy by proposing that structure and agency are mutually constitutive (Giddens, 1984), but the framework's abstract recursiveness has proved difficult to operationalise in empirical research on knowledge quality (Barley & Tolbert, 1997).

Emirbayer and Mische (1998) offered a more analytically tractable conception of agency as a temporally embedded process with three dimensions: the iterational, in which actors draw on habitual patterns from the past; the practical-evaluative, in which actors exercise situated judgment in response to present demands; and the projective, in which actors orient toward future possibilities. This framework is valuable because it specifies that the quality of agency depends on which temporal orientation is activated in a given situation. When practitioners rely on iterational habits, processing content the way they always have, they trust surface cues. They are exercising agency, but in a mode poorly suited to the epistemic novelty of AI co-production. The question becomes: what activates the practical-evaluative orientation that enables situated judgment?

Institutional Theory and the Temporal Assumption

Institutional theory provides powerful tools for understanding how practices become taken for granted, but it embeds a temporal assumption that AI co-production may disrupt. Tolbert and Zucker (1996) theorised institutionalization as a three-stage process: habitualization, objectification, and sedimentation, unfolding over extended periods as normative consolidation gradually transforms provisional practices into durable routines. Powell and Colyvas (2008) subsequently called for attention to the micro-foundations of institutional processes, arguing that institutionalization operates through everyday interactions and interpretive work, but their framework retained the assumption that sedimentation requires sustained normative reinforcement.

Digital infrastructure may compress this process. Barley (1986) showed that new technology triggers structural change by altering the conditions under which actors interact, and Suchman (1987) demonstrated that the relationship between plans and situated action is reshaped by technological mediation. Orlikowski (2000) theorised technology as an "occasion for structuring," a catalyst that accelerates the crystallisation of provisional practices into organisational routines. Star and Ruhleder (1996) demonstrated that infrastructure shapes practice by becoming invisible: once embedded, it recedes from awareness and its classificatory assumptions become taken for granted (Bowker & Star, 2000). Under AI co-production, digital document systems, shared templates, and knowledge bases may transform a single instance of content into an organisational resource that carries implicit institutional endorsement. If this compression is empirically real, institutional theory needs a concept for the rapid hardening of practices into artifacts under digital conditions, what we will call micro-institutionalization.

The Paradox of Algorithmic Credibility

AI co-production creates a distinctive epistemic condition in which the standard organisational toolkit, more structure or better people, may fail on its own. Structure falters because conventional verification presupposes that the object of verification carries markers of its epistemic origins: an author, a reasoning chain, a source. AI-generated content carries none of these; it produces what Bender and colleagues

(2021) termed stochastic outputs, fluent text without corresponding epistemic grounding. Individual expertise falters because the surface cues professionals rely upon to assess credibility, coherence, precision, professional formatting, are precisely the features AI produces most convincingly, yet they carry no epistemic information (Ananny & Crawford, 2018).

It is worth being precise about what is, and is not, unique to AI here. Drift-like dynamics are not new: consultants have long recycled prior reports, legacy templates outlive their evidentiary basis, and human-authored content propagates through organisations without re-verification. What AI changes is not the existence of these dynamics but their conditions and pace. Three features are distinctively algorithmic. First, AI produces fluent, authoritative prose with no recoverable provenance, severing the link between a claim and any human who could be asked to account for it. Second, it does so within a single drafting session and at volume, compressing the window for reflection that human authorship ordinarily imposes. Third, its outputs reproduce exactly the surface markers professionals use as quality heuristics, so the cues that normally trigger scrutiny instead suppress it. The phenomenon this study documents is therefore best understood as a general organisational vulnerability, propagation of unverified content, that AI sharply accelerates and partially conceals, rather than as a wholly novel pathology. Recent scholarship on responsible AI and algorithmic accountability situates this concern within a broader governance agenda (Crawford, 2021; Dignum, 2019; Mitchell et al., 2019).

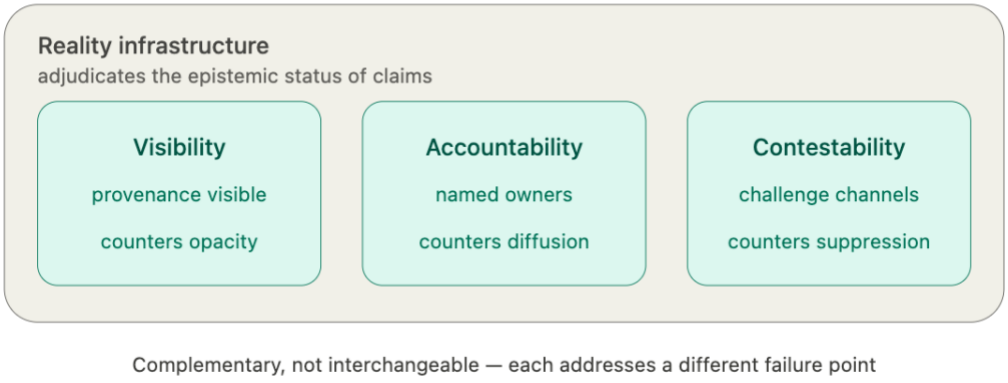
Prior qualitative research identified organisations that maintained quality under these conditions and organisations that experienced progressive erosion. The differentiating factor was not the sophistication of their structural controls or the expertise of their professionals but the configuration of what was termed reality infrastructure: background systems that structure how knowledge is produced, evaluated, and authorised. The present study uses longitudinal process tracing and embedded theoretical probes to examine how this infrastructure operates in time, and how it relates to the exercise of professional judgment.

Reality Infrastructure: Three Dimensions

To avoid the construct becoming over-inclusive, we define reality infrastructure narrowly as the configuration of background systems that adjudicate the epistemic status of knowledge claims, and we specify it along three distinct, separately manipulable dimensions (illustrated in Figure 2). Visibility infrastructure makes the provenance and verification status of a claim apparent at the point of use, so that an evaluator can see how a claim was produced. Accountability infrastructure assigns identifiable ownership for the evidentiary quality of a claim or section, so that responsibility is locatable rather than diffuse. Contestability infrastructure provides legitimate, low-cost channels for challenging a claim, so that doubt can be voiced and acted upon. Each dimension addresses a different failure point: visibility addresses opacity, accountability addresses diffusion of responsibility, and contestability addresses the social suppression of challenge. The three are complementary rather than interchangeable, a point the probe design was constructed to examine.

Figure 2

Reality Infrastructure: Three Dimensions



Note. Reality infrastructure adjudicates the epistemic status of claims through three separately implementable dimensions. Visibility infrastructure (provenance visible) counters opacity. Accountability infrastructure (named owners) counters diffusion of responsibility. Contestability infrastructure (challenge channels) counters suppression of challenge. The three dimensions are complementary rather than interchangeable.

Method

Research Design and Sites

The study employed a longitudinal process design across three organisational units within two professional services firms over 22 weeks. The first 12 weeks comprised baseline observation under existing conditions; the final 10 weeks introduced governance mechanisms as theoretical probes designed to disentangle how different dimensions of infrastructure shape epistemic practice.

The three units were selected for maximum theoretical leverage. Unit A, a 12-person business development team at a global management consultancy, operated in a context of advanced AI integration with existing but unevenly applied governance. Unit B, a nine-person team at a technical advisory firm, operated with minimal governance and high AI dependence. Unit C, an 11-person team at the same advisory firm, shared Unit B's organisational context but received a different theoretical probe. Both organisations had participated in prior qualitative research, providing continuity of access and baseline understanding of their epistemic practices.

Each unit received a different governance mechanism as a theoretical probe: Unit A received provenance labelling requirements that made the epistemic status of claims visible at the point of review; Unit B received accountability assignment procedures that designated named individuals as epistemic owners of defined document sections; Unit C received all three mechanisms: provenance, accountability, and contestability through a rotating reviewer role and anonymous concern-logging system. Each unit's 12-week baseline served as its own comparative referent for assessing mechanism effects.

Participants

The study involved 32 professionals across the three units (12, 9, and 11), spanning junior, mid-level, and senior roles in consulting and technical advisory. Roles were identified narratively through observation and interview (analysts, senior associates, partners) rather than through a structured taxonomy. Demographic details, including age, years of professional experience, and AI-usage frequency, were not collected via structured questionnaire; the study relied on observational and interview-based role identification. We note this as a limitation and recommend that future studies

incorporate systematic demographic collection. The 64 interviews reported below were drawn from these participants across three waves, so most participants were interviewed more than once.

Claim Selection and Coding

Two trained coders independently classified every substantive claim in tracked proposals along three dimensions: evidentiary quality, provenance status, and accountability assignment. Evidentiary quality was the basis for tracing selection, and was coded on a four-level scale with worked examples (summarised in Table 2): strong (verifiable cited primary source or original analysis), adequate (plausible secondary source or a clear reasoning chain), weak (apparent specificity but no identifiable source, or surface plausibility only), and unsupported (demonstrably inaccurate, internally contradictory, or without evidential basis). The full framework, including the provenance and accountability dimensions and all inclusion and exclusion rules, is in the registered Online Supplement (Section A).

The selection rule for process tracing is stated explicitly: of all claims classified "weak" or "unsupported" during the first four weeks of baseline observation, the 15 that appeared earliest and could be reliably traced through successive documents were selected for the full protocol. Claims were therefore selected on evidentiary quality, not provenance: the traced set was not exclusively AI-generated, although most traced claims entered without documented provenance (13 of 15), consistent with the origin blindness described in the findings. To guard against coder bias, all claims were classified independently by two coders, with substantial inter-rater reliability assessed on a random 20% subsample (evidentiary quality $\kappa = .82$; provenance $\kappa = .86$; accountability $\kappa = .88$; overall $\kappa = .84$); disagreements were resolved through discussion with the lead researcher.

Data Collection

Four complementary data streams operated continuously across both phases. Claim-level coding classified every substantive claim in tracked proposals, 498 claims across 38 proposals, by evidentiary quality, provenance status, and accountability assignment. Longitudinal interviews at three time points, Weeks 1, 12, and 22, yielded 64 interviews

across hierarchical levels, capturing within-person shifts in epistemic awareness and practice. Interviews lasted approximately 45–60 minutes, were audio-recorded with participant consent. Analysis followed an iterative thematic approach, with codes developed inductively from field observations and validated against the claim-level process tracing. Direct observation of 23 proposal-preparation sessions and 11 review meetings provided access to the tacit, embodied dimensions of epistemic practice that are invisible in retrospective accounts; the researcher attended as a non-participant observer, and contemporaneous field notes were expanded into fuller memos after each session.

The study's analytical core was process tracing of the 15 weakly supported claims. Each was tracked through successive documents, recording every instance of reuse, modification, verification, or challenge, providing the temporal precision needed to validate the theorised drift process and identify the points at which infrastructure interrupted it.

Ethics

The study received ethical approval through the doctoral research ethics review process. All participants gave informed consent; participation was voluntary and confidential; organisational and individual identities were pseudonymised; and data were stored securely under access restriction in line with institutional data-protection policy and the UK General Data Protection Regulation. Interview transcripts cannot be shared owing to confidentiality agreements.

Analytical Approach

Analysis combined process tracing with comparative mechanism analysis. Each claim's trajectory was assessed against the theorised four-stage model. Pre-probe and post-probe comparison estimated mechanism effects. Mediation analysis tested whether challenge behaviour mediated the relationship between infrastructure and drift (full specification in Supplement D).

A note on quantification and inference. This is a process-tracing study in which the analytical priority is the mechanism and sequence of drift rather than its population frequency. The proportions reported below, the drift rates per unit and the share of the

governance effect associated with challenge behaviour, are descriptive summaries of the coded claim corpus. The small number of organisational units (three) means the design cannot support unit-level inferential generalisation without overstating precision. We therefore report the underlying claim corpus alongside the proportions, base inferential claims on the proposal-level model described below, and interpret cross-unit differences as patterns consistent with the theorised mechanisms rather than as statistically estimated effects.

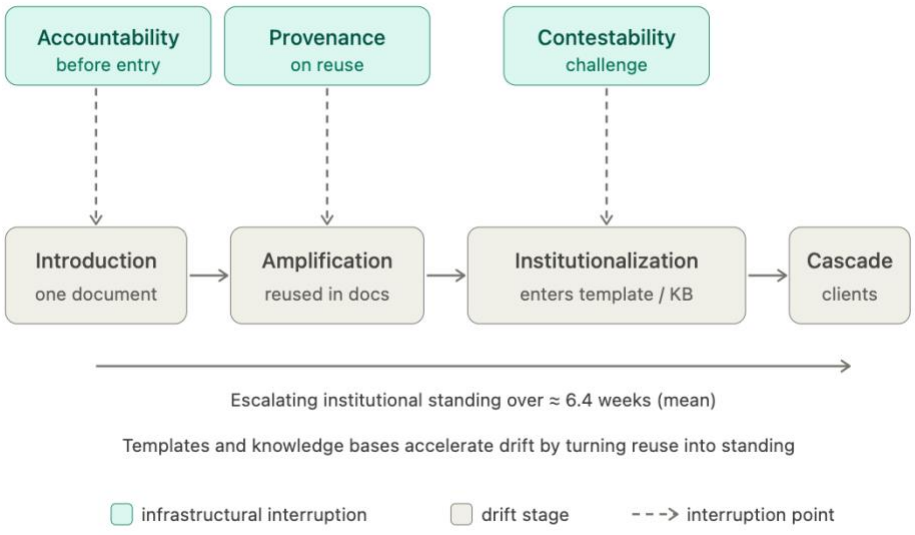
Findings

The Four-Stage Drift Process

Process tracing of the 15 weakly supported claims during the baseline period traced a four-stage sequence through which AI-generated content acquired institutional endorsement (Figure 1). Thirteen of 15 claims followed the full sequence. The mean lifecycle from introduction to cascade was 6.4 weeks. The stages and their empirical patterns are summarised in Table 1.

Figure 1

The Four-Stage Epistemic Drift Process with Infrastructural Interruption Points



Note. The four-stage drift process shows how AI-generated claims progress from introduction (single document) through amplification (reused across documents),

institutionalization (enters template or knowledge base), and cascade (reaches clients). Each governance mechanism interrupts drift at a different stage: accountability before entry, provenance on reuse, contestability at the challenge point. Mean lifecycle from Introduction to Cascade: 6.4 weeks. Solid arrows indicate stage progression; dashed arrows indicate interruption points.

A Claim's Journey: The Digital Health Market Estimate

One claim's trajectory through the four stages illustrates the social accomplishment of drift, how epistemic status changes not through deliberate decision but through the accumulation of reasonable micro-actions.

In Week 2, a junior analyst at Unit A used a generative AI model to draft a market overview for a digital health proposal. The model produced a market sizing estimate: "The global digital health market is estimated at $230 billion." The analyst reviewed the paragraph, found it consistent with her general sense of the market, and incorporated it into the working draft. When asked about verification during a debrief, she explained: "The formatting was professional. It included a specific number. I didn't have time to check it against a primary source, and honestly, the model usually gets these things roughly right." The claim entered the workflow in a single session. The analyst was exercising agency, but in the iterational mode: she processed the AI output through the same heuristics she had always used for evaluating drafted content, surface coherence, apparent precision, professional formatting. These cues, reliable indicators of quality in human-authored work, carried no epistemic information in AI-generated content.

By Week 4, a second analyst, preparing a different proposal for a related client, searched the team's recent documents for market data. She found the digital health estimate and incorporated it, rephrasing "estimated at $230 billion" as "valued at $230 billion." The linguistic shift was subtle but epistemically consequential: "estimated" implies tentativeness and a source that produced the estimate; "valued at" implies established fact. Neither analyst intended to misrepresent; the second had no way of knowing the claim's AI origins. The act of reuse itself had obscured provenance.

By Week 7, the market sizing figure had been incorporated into the team's standard market overview template, an artifact carrying implicit institutional endorsement. When a new analyst questioned the figure during a Week 9 preparation session, a senior

associate responded: "That's our standard number. It's been in multiple proposals." The claim's history of use had become its credential. Repetition had substituted for verification. The questioning analyst did not pursue the challenge; the social cost of questioning "our standard number" exceeded the perceived benefit of checking a figure that everyone else appeared to accept.

By Week 10, the estimate appeared in a cross-team knowledge base entry and was cited in a client presentation to a health sector investor. Neither the presenting partner nor the client could trace the figure to a primary source. The claim had moved from tentative AI estimate to institutional fact in eight weeks. No one had decided to lower epistemic standards; the standard had lowered itself through the sequential operation of plausibility, repetition, and institutional endorsement.

Patterns Across the 15 Claims

The digital health estimate was typical. Across all 15 claims, introduction occurred under time pressure in 14 cases and without provenance information in 13. The speed of introduction was consistent: 12 claims moved from AI generation to working document within a single session. Amplification through reuse occurred in 13 cases within three weeks, with template-based assembly and knowledge base contributions serving as drift accelerators, practices designed to capture organisational learning that, under AI co-production, functioned as vectors for the propagation of unverified content. Institutionalization, entry into templates, training materials, or knowledge bases, occurred in 11 cases. Nine claims reached cascade, propagating beyond originating teams into client-facing materials, with four producing observable consequences (Table 1).

The six-to-eight-week timeline from introduction to institutionalization is consistent with a compression of institutional sedimentation by orders of magnitude. Where Tolbert and Zucker (1996) theorised sedimentation as requiring sustained normative consolidation over years, the digital document systems, shared templates, and AI-assisted knowledge bases observed here appeared to transform provisional content into organisational resources within weeks. This micro-institutionalization operated not through normative consensus but through the material properties of digital infrastructure: once a claim entered a template, the template's classificatory logic (Bowker & Star, 2000) conferred

institutional status without requiring the collective endorsement that traditional institutionalization presupposes.

Drift Reduction Under the Probes

Drift rates fell in every probed unit between the baseline and probe phases. Because the design has only three units, these per-unit rates are reported as descriptive summaries of the coded corpus rather than as independent inferential comparisons; the pooled reduction in drift is consistent with the significant total effect of the mediation model reported below ($b = -0.66$, $p < .001$). Per-unit rates are shown in Table 3; the exact per-unit claim counts are held in the claim-level database (available on request).

Epistemic Surfacing Through Provenance

The provenance mechanism operated by making claims' epistemic status visible at the point of reuse. When tagged claims appeared in new documents, inconsistencies between provenance status and asserted certainty became apparent. Over the 10-week probe period in Unit A, provenance tagging surfaced seven instances of unsupported claims that would otherwise have progressed to amplification.

The mechanism appeared to activate the practical-evaluative dimension of agency. When a reviewer encountered a claim tagged "AI-generated, unverified" alongside text presenting it as established fact, the tag disrupted habitual processing and prompted situated judgment. One reviewer described the shift: "When I have to tag something as unverified, I think twice about putting it in the document. It makes the gap between what the AI produced and what I can actually vouch for very visible." The tag did not tell the reviewer what to do; it created the informational conditions under which judgment could be exercised. Drift in Unit A declined from a baseline rate of 52% to 31%.

Anticipatory Verification Through Accountability

The accountability mechanism appeared to activate the projective dimension of agency. When professionals knew their names were attached to specific document sections, they oriented toward future moments of accountability, client presentations, partner reviews, audit inquiries, and verified claims pre-emptively. Verification rates under named ownership were 2.3 times higher than under diffuse ownership.

The mechanism was strongest at preventing introduction: four of six prevented claims were caught by named owners before entering workflows. A senior associate captured the dynamic: "The first time I saw my initials on that section, I went back and checked every number. By Week 18, it was automatic." The limitation was that ownership was assigned per proposal, not per claim; once content entered the knowledge base, it was no longer under any individual's ownership. Drift in Unit B declined from 58% to 35%.

Legitimised Contestation and the Mediation of Challenge

The combined probe in Unit C introduced contestability mechanisms, a rotating epistemic reviewer role and anonymous concern-logging, alongside provenance and accountability. Drift declined from 55% to 19%, and challenge behaviour increased from 0.8 concerns per proposal to 3.8.

To test whether challenge behaviour was the pathway through which infrastructure reduced drift, we estimated a mixed-effects logistic mediation model, with claims nested within proposals within units, treating infrastructure condition (baseline = 0, probe = 1) as the predictor, challenge behaviour as the mediator, and drift beyond the introduction stage as the outcome (full specification in Supplement D). Infrastructure reduced drift (total effect b = −0.66, p < .001) and increased challenge behaviour (path a, b = 0.71, p < .001), which in turn predicted lower drift (path b, b = −0.66, p < .001). The indirect effect through challenge behaviour was significant (ab = 0.47, 95% CI [0.31, 0.65], 5,000 bootstrap resamples), accounting for approximately 71% of the total effect, with a smaller but significant direct effect remaining (c′ = 0.19, p = .028), indicating partial mediation. A secondary model localised the pathway to substantive challenge: challenges targeting evidentiary quality fully mediated the relationship (ab = 0.39, 95% CI [0.22, 0.58]), whereas challenges targeting formatting or style did not (ab = 0.04, 95% CI [−0.03, 0.12]). We note one constraint: with three organisational units at the highest level, the unit-level variance component is estimated on very few clusters, so inference rests chiefly on the proposal-level clustering (38 proposals) and the bootstrapped claim-level resampling, and the cross-unit comparisons are read descriptively.

The nature of challenges also shifted: during baseline, 78% targeted formatting or style; during the probe period, 64% targeted evidentiary quality. A field note from Week 15

illustrates how the three mechanisms operated as a layered system. The rotating epistemic reviewer that week, a second-year analyst, flagged a growth projection tagged as "AI-generated, unverified." The section owner checked the primary source and discovered the projection was based on a superseded 2019 report. The anonymous concern log showed that a colleague had flagged the same figure the previous week but lacked the confidence to raise it verbally. The combination of the tag, the ownership structure, and the concern log produced a correction that no single mechanism would have achieved alone.

This finding speaks to the relationship between infrastructure and professional judgment. Edmondson (1999) demonstrated that psychological safety enables voice in teams. The present findings suggest a complementary but distinct mechanism: infrastructure may provide not the psychological conditions for speaking up but the epistemic conditions for speaking substantively. The anonymous concern log removed social cost, but it was the provenance tag that gave the concern epistemic content, transforming a vague sense that "something seemed off" into a specific, actionable observation. Challenge under infrastructure was not merely safer; it was better informed.

Temporal Dynamics

The three mechanisms showed distinct temporal trajectories that illuminate their causal logic. Provenance-based detection improved gradually as the tagged corpus grew: detection rose from 0.3 per proposal in early weeks to 0.8 by the end of the probe period. This cumulative pattern has a structural explanation: the more tagged content in the system, the more opportunities for contradictions to become visible. Accountability-based verification plateaued immediately; named owners reported an instantaneous shift in behaviour upon seeing their initials attached to content. Contestability showed a third pattern: an initial surge as professionals tested the system, a brief decline as novelty faded, and stabilisation above baseline after minimal reinforcement. All mechanisms activated within two weeks and showed no decay during the observation period.

Boundary Conditions

Two proposals prepared under extreme time pressure, within 48 hours, showed infrastructure bypass even in the combined unit. Provenance tagging was incomplete, review was compressed, and one weak claim entered the workflow undetected. However, infrastructure provided a recovery mechanism: the claim was caught during routine post-submission review and corrected before client presentation. Infrastructure appeared to provide resilience rather than immunity.

Contestability mechanisms were less effective in the more hierarchical unit unless accompanied by explicit senior leadership endorsement; when the unit's partner publicly affirmed the legitimacy of junior challenge, challenge rates increased by 40% within two weeks. Infrastructure must be culturally calibrated.

Discussion

Infrastructure as the Condition of Possibility for Judgment

The central pattern in this study offers one way of addressing the structure–agency impasse outlined earlier. The mediation through challenge behaviour suggests that infrastructure does not operate as a structural substitute for judgment, an automated detection system that renders expertise irrelevant, nor does judgment operate independently of structural support. Instead, infrastructure may create the conditions of possibility for the exercise of situated professional agency.

We term this resolution enabled agency: a mode of organisational action in which infrastructure creates the informational, accountability, and social conditions under which practitioners can exercise practical-evaluative judgment, without determining the content of that judgment. Provenance systems did not tell reviewers which claims to accept; they made claims' epistemic status visible so that reviewers could exercise discrimination. Accountability structures did not dictate verification procedures; they activated the projective orientation that led professionals to verify proactively. Contestability mechanisms did not prescribe what to challenge; they provided the epistemic content that transformed vague unease into substantive critique. In the absence of infrastructure, AI co-production appears to push professionals into the

iterational mode, processing algorithmically generated content through habitual heuristics that are informationally impoverished under AI conditions. Infrastructure, on this reading, does not create agency; it shifts the temporal orientation of agency from habit to judgment.

Micro-Institutionalization: Challenging the Temporal Assumption

The six-to-eight-week timeline from introduction to institutional embedding is in tension with a foundational assumption of institutional theory. Tolbert and Zucker (1996) theorised sedimentation as the culmination of a process requiring sustained normative consolidation. The present findings are consistent with a different institutionalization mechanism under digital conditions: artifact-level sedimentation, in which provisional content hardens into organisational routine not through normative consensus but through the material logic of templates, knowledge bases, and digital document systems.

We propose the concept of micro-institutionalization to capture this accelerated process: it occurs when digital infrastructure compresses the cycle from provisional practice to organisational artifact, conferring institutional status through classification and inclusion rather than through collective endorsement. The template does not ask whether the content it contains has been verified; it treats inclusion as endorsement. To clarify how this differs from adjacent constructs, Table 4 contrasts micro-institutionalization with four neighbouring literatures.

Alternative Explanations

Because the probe periods coincided with researcher engagement, several non-infrastructural explanations warrant explicit consideration. Researcher presence, organisational learning over time, heightened management attention, and Hawthorne-type reactivity could each in principle account for some of the observed reduction in drift. Two features of the design temper these explanations without eliminating them. First, drift rates remained stable across the 12-week baseline despite continuous researcher presence, which argues against a purely observational account of the probe-period changes. Second, the three units were observed equally but showed differential reductions aligned with the specific mechanism each received, a pattern not predicted

by a generalised observer effect. These features make the infrastructural interpretation more plausible but do not establish pure causal attribution; the findings are best read as consistent with a causal interpretation rather than as a demonstration of one.

Limitations and Future Directions

Several limitations bound these findings. The 10-week probe period cannot assess long-term sustainability; decay after research engagement ends remains an open question. The two-organisation, professional-services context limits external validity, and the mechanisms may operate differently in other sectors or in cultures with different challenge and accountability norms. The 15 traced claims provide detailed process evidence but not statistical generalisation about drift frequencies, and the three-unit design constrains multilevel inference, as noted above. Demographic characteristics of participants were not systematically collected, limiting description of the sample. We therefore recommend replication across a larger number of units and sectors, with systematic demographic measurement and longer post-implementation observation, and we present the present results as theory-generating patterns inviting confirmatory study. Future work should also examine whether the temporal compression documented here varies with the sophistication of AI systems, and whether AI tools can be designed to support rather than resist provenance tracking.

Conclusion

When artificial intelligence co-produces organisational knowledge, credibility may erode not suddenly but through a recurring sequence of introduction, amplification, institutionalization, and cascade that, in this study, unfolded over approximately six weeks. This temporal compression sits in tension with institutional theory's assumption that sedimentation requires sustained normative consolidation and motivates a concept, micro-institutionalization, for the rapid hardening of provisional knowledge into organisational artifacts under digital conditions. Infrastructure appeared able to interrupt this process, not by substituting for professional judgment but by creating the conditions, informational visibility, accountability salience, and the social legitimacy of challenge, under which professionals shifted from habitual processing to situated

evaluation. The resolution these findings point to is not more structure or better people, but infrastructure that enables people to exercise the judgment that structure alone cannot provide.

Data and Methods Transparency

This study was conducted as part of doctoral research. The coding framework, process-tracing protocol, governance-mechanism probe protocols, mediation-analysis specification, and longitudinal interview protocol are registered and openly archived at https://osf.io/am2yr (DOI: 10.17605/OSF.IO/AM2YR; Internet Archive copy available). The registration documents the analysis specification and supporting materials and was completed prior to journal submission. The claim-level coding database is available from the author on request, subject to participant confidentiality. Interview transcripts cannot be shared owing to confidentiality agreements.

Footnotes

1 The governance mechanisms were designed as minimal, workflow-embedded modifications requiring modest temporal investment relative to the proposal lifecycle, so that observed effects would reflect mechanism activation rather than resource intensification.

2 Inter-rater reliabilities for claim-level coding was substantial (overall $\kappa = .84$). Disagreements were resolved through discussion with the lead researcher.

References

Adler, P. S., & Borys, B. (1996). Two types of bureaucracy: Enabling and coercive. Administrative Science Quarterly, 41(1), 61–89.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal. New Media & Society, 20(3), 973–989.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. Annual Review of Psychology, 52, 1–26.

Barley, S. R. (1986). Technology as an occasion for structuring: Evidence from observations of CT scanners and the social order of radiology departments. Administrative Science Quarterly, 31(1), 78–108.

Barley, S. R., & Tolbert, P. S. (1997). Institutionalization and structuration: Studying the links between action and institution. Organisation Studies, 18(1), 93–117.

Bechky, B. A. (2003). Sharing meaning across occupational communities. Organisation Science, 14(3), 312–330.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots. Proceedings of the 2021 ACM FAccT Conference, 610–623.

Bowker, G. C., & Star, S. L. (2000). Sorting things out: Classification and its consequences. MIT Press.

Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Dignum, V. (2019). Responsible artificial intelligence: How to develop and use AI in a responsible way. Springer.

Edmondson, A. C. (1999). Psychological safety and learning behaviour in work teams. Administrative Science Quarterly, 44(2), 350–383.

Emirbayer, M., & Mische, A. (1998). What is agency? American Journal of Sociology, 103(4), 962–1023.

Giddens, A. (1984). The constitution of society. University of California Press.

Lindsley, D. H., Brass, D. J., & Thomas, J. B. (1995). Efficacy-performance spirals. Academy of Management Review, 20(3), 645–678.

March, J. G. (1991). Exploration and exploitation in organisational learning. Organisation Science, 2(1), 71–87.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency, 220–229.

Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens. Organisation Science, 11(4), 404–428.

Powell, W. W., & Colyvas, J. A. (2008). Microfoundations of institutional theory. In R. Greenwood, C. Oliver, K. Sahlin, & R. Suddaby (Eds.), The SAGE handbook of organisational institutionalism (pp. 276–298). Sage.

Scott, W. R. (2001). Institutions and organisations (2nd ed.). Sage.

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure. Information Systems Research, 7(1), 111–134.

Suchman, L. A. (1987). Plans and situated actions: The problem of human-machine communication. Cambridge University Press.

Tolbert, P. S., & Zucker, L. G. (1996). The institutionalization of institutional theory. In S. Clegg, C. Hardy, & W. Nord (Eds.), Handbook of organisation studies (pp. 175–190). Sage.

Table 1

The Four-Stage Drift Process: Empirical Patterns

Stage	**Timeline**	**Pattern (N = 15 claims)**	**Theoretical Mechanism**
Introduction	Weeks 1–2	15/15 under time pressure; 13/15 without provenance; 12/15 same session	Iterational agency: surface heuristics applied to epistemically opaque content

Amplification	Weeks 3–5	13/15 reused within 3 weeks; origin obscured by rewording	Material drift: templates and knowledge bases confer legitimacy through inclusion
Institutionalization	Weeks 6–8	11/15 entered templates or knowledge bases; repetition = credential	Micro-institutionalization: digital artifacts compress sedimentation
Cascade	Week 9+	9/15 propagated cross-team; 4 observable consequences; mean lifecycle 6.4 weeks	Irreversibility threshold: correction costs escalate exponentially

Note. N = 15 weakly supported claims traced through complete organisational lifecycle during 12-week baseline period.

Table 2

Claim Classification by Evidentiary Quality (Coding Framework, Supplement A)

Level	**Definition**	**Example**
Strong	Verifiable cited primary source, or original analysis by the author	"Revenue grew 12% year-over-year (Company Annual Report 2025, p. 34)."
Adequate	Plausible secondary source, or professional judgment with a clear reasoning chain	"The regulatory environment is expected to tighten, given recent enforcement in adjacent sectors."

Weak	Apparent specificity but no identifiable source, or surface plausibility only	"The global digital health market is valued at $230 billion."
Unsupported	Demonstrably inaccurate, internally contradictory, or without evidential basis	A statutory deadline cited incorrectly; a figure contradicted by public data.

Note. Coding framework used by two independent coders. Inter-rater reliability: evidentiary quality κ = .82; provenance κ = .86; accountability κ = .88. Full framework, including provenance and accountability dimensions and inclusion/exclusion rules, in registered Supplement (Section A).

Table 3

Drift Rate by Unit and Phase (Descriptive)

Unit	**Probe Mechanism**	**Baseline Drift Rate**	**Probe Drift Rate**	**Reduction**
A	Provenance	52%	31%	21 pp
B	Accountability	58%	35%	23 pp
C	All three	55%	19%	36 pp

Note. Drift rate = the proportion of coded claims in the unit and phase reaching institutionalization or beyond. Exact per-unit claim counts are held in the claim-level database (available on request). Percentages are descriptive and no separate per-unit inferential test is applied, given the three-unit design; the pooled reduction is consistent with the mediation model total effect ($b = -0.66$, $p < .001$). pp = percentage points.

Table 4

Micro-Institutionalization in Relation to Adjacent Constructs

Construct	**Existing Emphasis**	**Micro-Institutionalization**
Institutional work	Actor-centred, purposive effort to create or maintain institutions	Artifact-centred; status conferred by inclusion, without purposive effort

Routine dynamics	Reproduction and gradual change of practices over time	Accelerated sedimentation of content into durable artifacts
Organisational learning	Accumulation and codification of useful knowledge	Hardening of unverified content into authoritative knowledge
Digital materiality	Affordances and constraints shape situated action	Classification logic automatically confers institutional standing

Note. Micro-institutionalization is framed as extending rather than contradicting institutional theory, operating within the interstices of slower macro-institutional processes. The concept specifies a mechanism through which digital infrastructure compresses the cycle from provisional practice to organisational artifact.