

Taş, Eylem; Memmert, Lucas; Bittner, Eva

Article — Published Version

Episodic oversight in generative AI workflows: A nine-step protocol for preserving human agency (OP-9)

Electronic Markets

Suggested Citation: Taş, Eylem; Memmert, Lucas; Bittner, Eva (2026) : Episodic oversight in generative AI workflows: A nine-step protocol for preserving human agency (OP-9), Electronic Markets, ISSN 1422-8890, Springer Berlin Heidelberg, Berlin/Heidelberg, Vol. 36, Iss. 1, <https://doi.org/10.1007/s12525-026-00915-x>

This Version is available at:

<https://hdl.handle.net/10419/341851>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>



Episodic oversight in generative AI workflows: A nine-step protocol for preserving human agency (OP-9)

Eylem Taş¹ · Lucas Memmert² · Eva Bittner²

Received: 31 August 2025 / Accepted: 4 June 2026
© The Author(s) 2026

Abstract

Generative AI is embedded in everyday knowledge work, but we know little about how workers preserve human agency when large language models (LLM) output enters real deliverables. Drawing on 15 interviews in two early-adopting German tech firms and analysis of handbooks and workflow artifacts, we examine how professionals keep control and responsibility when LLM generate parts of their work. We focus on three phases where people step in: drafting with the AI, refining its output, and reviewing the final product with others. We derive a nine-step checklist that specifies what practices to apply, and when, to keep AI-assisted text reliable, traceable, and accountable. We contribute by (1) explaining how ownership and discretion jointly ground human agency in GenAI workflows, (2) mapping when and how judgment shifts between humans and LLMs across drafting, refining, and reviewing, and (3) deriving a nine-step Oversight Protocol (OP-9) that embeds simple checks at key points in the workflow.

Keywords Human agency · Psychological ownership · Generative AI · Knowledge work · Oversight · Enterprise workflows

JEL Classification M15 · D83 · J24 · L86

Introduction

Large language model (LLM) services have rapidly diffused into knowledge-intensive work, transforming how professionals generate, draft, and refine their outputs. While these tools promise substantial efficiency gains, they also raise concerns about the erosion of human agency, understood as the capacity to shape, control, and take ownership of work products within socio-technical systems (Henkenjohann & Trenz, 2024; Zhu et al., 2025). Existing information systems (IS) research often conceptualizes oversight as a static

safeguard situated at the end of the workflow (e.g., provenance checks, disclosure statements), without accounting for the nuanced, episode-specific ways in which professionals intervene in real time (Botero Arcila, 2024; Cancela-Outeda, 2024). This end-point framing risks obscuring where agency is actually gained or lost, often earlier in the process, thereby constraining the design of interventions to episodes at which errors or biases may already be entrenched.

This study addresses this gap by examining episodic human oversight, defined as the deliberate, context-dependent adjustments that professionals make at distinct episodes of LLM-mediated, GenAI-enabled workflows. We investigate how these interventions shape the tension between automation and human agency by balancing productivity gains with the preservation of professional authorship and accountability in IS-enabled organizations, thereby extending recent calls for risk-sensitive and adaptive governance of GenAI in organizational contexts (Krakowski, 2025; Meilvang & Dahler, 2024).

We adopt a process perspective and draw on Davenport's (2005) knowledge-work archetypes to examine how human oversight unfolds in LLM-mediated knowledge work. We conceptualize oversight as unfolding across three recurring

Responsible Editor: Xusen Cheng

✉ Eylem Taş
eylem.tas@uni-hamburg.de

Lucas Memmert
lucas.memmert@uzh.ch

Eva Bittner
eva.bittner@uzh.ch

¹ Informatics, University of Hamburg, Hamburg, Germany

² Informatics, University of Zurich, Binzmühlestrasse 14, 8050 Zurich, Switzerland

episodes (drafting, refining, reviewing), inductively derived from temporal segmentation of interview narratives and triangulated with workflow artefacts (see “[Coding and analysis](#)” section/Cycle 2 for details). Reconstructing participants’ accounts end-to-end, we observed a recurring oversight flow organized around these episodes. Work typically begins with drafting, where professionals use an LLM to produce a first-pass scaffold (e.g., an outline, candidate options, or rough wording). It then moves into refining, where they actively shape and correct that output by verifying claims, adding domain-specific context, aligning tone and intent to the audience, and deciding what to keep, rewrite, or discard. Finally, the process enters reviewing, where the refined output is validated and coordinated with others (e.g., peers, managers, or legal/compliance) and prepared for handoff or release. Importantly, these episodes describe a recurring oversight cycle: depending on risk and feedback, review comments or new constraints can send work back into refining (or even back to drafting) before final release.

To explain how accountability is maintained when outputs are AI-assisted, we draw on psychological ownership (Pierce et al., 2001) and conceptualized task-related ownership as its task-level expression: the felt responsibility, control, and accountability for the work episode and its consequences. This framing helps connect governance expectations discussed in policy debates (e.g., the EU AI Act) with concrete enterprise practices and links IS concepts of accountability to everyday knowledge work (Afzal & Panagiotopoulos, 2025; Vaccaro et al., 2024). Guided by this framing, we pose the following research questions:

RQ1: How do knowledge workers experience and negotiate human agency across different episodes of LLM-mediated workflows?

RQ2: What strategies enable them to preserve authorship and accountability while maintaining the productivity gains afforded by GenAI tools?

We follow a qualitative, interpretive case-study approach across two German technology firms (Firm A $\approx$ 7000 employees; Firm B $\approx$ 2000 employees) that provided early organization-wide access to LLM services. Our dataset comprises 15 semi-structured interviews (39–95 min; mean 54), complemented by organizational artefacts: 3 internal AI-use documents (2 guidelines, 1 handbook), 17 annotated workflow screenshots, and 6 live demonstrations. We anonymized all data and obtained informed consent. Our analysis followed three cycles. First, we coded the qualitative data using established concepts (e.g., knowledge-work archetypes, psychological ownership). Second, we identified recurring oversight episodes, defined as bounded segments of work in which a person uses GenAI for a task and completes the required oversight actions before the

output is accepted or handed off, along with the risk triggers that shaped these episodes. Third, we compared cases to find cross-case patterns and mapped the underlying mechanisms.

In brief, we find that workers do not rely on a single end-point check. Instead, they intervene at three episodes in LLM-mediated workflows: when generating drafts (drafting), when revising and verifying AI text (refining), and when coordinating final checks and sign-off with others (reviewing). Across these episodes, they use a set of recurring tactics that adjust how much they rely on GenAI depending on task risk. We summarize these tactics in a nine-step Oversight Protocol (OP-9) that helps teams retain authorship and accountability while still benefiting from GenAI’s speed. OP-9 is a structured set of steps and roles that specifies when, how, and by whom GenAI outputs must be checked, documented, and escalated, so that accountability remains human and decisions remain auditable (see “[Deriving OP-9](#)” section for details).

By answering our research questions, we contribute to IS research on human–AI collaboration and socio-technical governance in three ways. First, we conceptualize episodic oversight sequencing: human interventions at drafting, refining, and reviewing stages that are calibrated to task risk. This extends work on levels of automation and human-in-the-loop oversight by showing that control is distributed across the workflow rather than located at a single release gate.

Second, we theorize risk-sensitive oversight at the task level, showing which risk cues (e.g., external audiences, legal exposure, reputational stakes) trigger stronger human intervention, where that intervention happens in the process, and why this cannot be reduced to an end-of-process audit. In other words, oversight is not just “check the final output.” It often intensifies midstream as risks become visible, which challenges compliance and provenance approaches that treat oversight mainly as an end-point verification step. Third, we derive OP-9 that shifts attention from ex post checks to in-process governance and connects policy-level debates, such as the EU AI Act, to everyday ways employees use GenAI at work.

Taken together, our findings broaden perspectives that foreground efficiency or inevitable agency erosion (Henkenjohann & Trenz, 2024; Zhu et al., 2025) by documenting conditions under which productivity gains can coexist with professional authorship and accountability, thereby locating this paper firmly within—and advancing—the ongoing discourse on human agency in LLM-mediated knowledge work.

In this paper, we proceed as follows. We first situate our work within IS research on GenAI, human agency, and governance. We then introduce our process perspective on LLM-mediated workflows and explain the three oversight episodes of drafting, refining, and reviewing. Next,

we describe our qualitative study in two organizations that have introduced GenAI tools into knowledge work. We present our findings on how workers experience and negotiate human agency across the three episodes and how collaboration and communication change when GenAI becomes part of the workflow. Finally, we discuss implications for IS research on governance and accountability and offer practical guidelines for designing oversight in enterprise GenAI adoption.

Theoretical framing

To understand how professionals sustain agency in GenAI-mediated work, we need to know: what kind of work is being done, why control and authorship matter, how workers exercise discretion over AI output, and when oversight intensity (i.e., the required strength of oversight for a task, which typically increases with risk, such as low-risk drafting versus high-stakes recommendations) changes with risk (Joshi & Vogel, 2025; Krakowski, 2025). We address these questions by combining four strands of IS and organizational theory: Davenport's classification of knowledge-intensive processes (2005), psychological ownership (Pierce et al., 2001), professional discretion (Lipsky, 1980), and risk-adaptive human–automation collaboration models (e.g., Cummings, 2014).

Structuring agency through Davenport's knowledge work archetypes

We use prior work to understand who feels responsible for AI-generated text, and when they step in to check or correct it. Davenport's (2005) framework distinguishes four archetypes of knowledge-intensive processes—transaction, integration, expert, and collaboration—based on two dimensions (see Fig. 1): the degree of collaboration (individual

vs. group) and the nature of work (routine vs. interpretive/judgment).

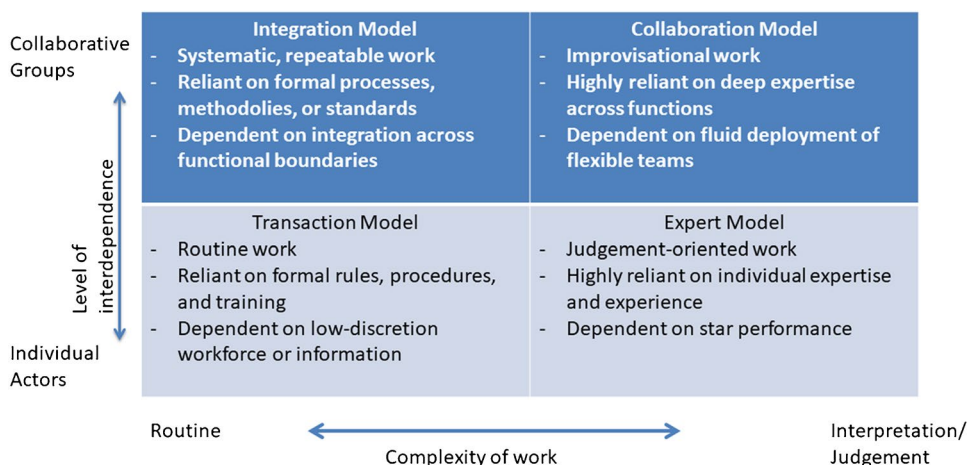
The resulting quadrants—transaction, integration, expert, and collaboration—provide a parsimonious map for locating AI–human interactions within an episodic workflow. In our context, the map clarifies where agency resides (ownership) and how it is exercised (discretion) as tasks move from routinized execution to open-ended interpretation.

With GenAI and LLMs embedded into these same processes, the structure changes. First, professionals “collaborate” with the LLM at an early drafting stage to produce a scaffold, such as an outline, draft paragraph, or list of options. Second, these AI-assisted drafts then enter shared team spaces, where collaboration focuses less on writing from scratch and more on checking, modifying, and validating AI-influenced text. This shift affects who speaks to whom, at what moment, and about which parts of the document, and it makes the distribution of responsibility more explicit, because AI contributions must be identified and discussed.

Positioning tasks within this 2×2 helps explain variation in agency practices across our cases. Specifically, Davenport's dimensions guided our attention to shifts in task-related AI ownership and discretion in AI-assisted work across tasks. For example, transaction/integration work emphasizes ex ante standardization (acceptance criteria, constrained prompting, provenance capture), whereas expert/collaboration work emphasizes in-episode discretion (delegation decisions, contrastive verification, peer review targeted at AI-exposed segments) to sustain accountable authorship under uncertainty.

Transaction (individual $\times$ routine) Highly structured, repeatable tasks performed by a single actor. Agency is narrow but well-defined; task-related AI ownership is concentrated in the individual who signs off, while discretion in AI-assisted

Fig. 1 A classification structure knowledge-intensive processes (Davenport, 2005)



work is low because procedures and acceptance criteria are pre-specified.

Integration (group $\times$ routine) Standardized processes involving several people with clear handoffs. Responsibility is shared across roles, and who is accountable for AI-related tasks is made transparent through documentation and audit trails (i.e., the minimal records needed to reconstruct what the GenAI produced, what was changed, who approved it, and why). Individual discretion in using AI is limited by shared standard operating procedures (SOPs), that is, formally documented rules and step-by-step guidelines that specify who may use GenAI, for which tasks, with what checks, and when escalation or approval is required.

Expert (individual $\times$ interpretive) Ill-structured tasks (i.e., tasks with ambiguous goals, incomplete information, and no clear “correct” procedure) require professional judgment by a single individual. In these tasks, task-related AI ownership is high and personal: the professional explicitly owns the outcome. At the same time, discretion in AI-assisted work is broad, including when and how to use the LLM and what evidence or checking is sufficient before accepting or rejecting its output.

Collaboration (group $\times$ interpretive) Co-created outputs (e.g., strategy decks, complex analyses) involving joint sensemaking. Ownership must be constructed (e.g., explicit leads, sign-off roles); discretion in AI-assisted work is negotiated across actors and episodes.

Prior work on knowledge-intensive collaboration describes pre-GenAI workflows as relatively linear: a professional drafts a document, circulates versions via shared tools or email, and colleagues’ comment on or approve these drafts in one or more review rounds (e.g., Davenport, 2005; Newell et al., 2009). Collaboration and communication focus on human-written text, and authorship is usually tied to the person who produces and coordinates the initial draft. Introducing GenAI and LLMs is therefore expected to shift collaboration toward an earlier interaction with the system to generate scaffolds and a later emphasis on jointly checking and validating AI-influenced segments, raising new questions about how responsibility for these parts is allocated and communicated.

From psychological ownership to task-related ownership

Psychological ownership describes the feeling that a target of work is “mine/ours,” typically emerging through perceived control, intimate knowledge, and self-investment (Pierce et al., 2001). In knowledge work involving GenAI, however,

the output is often co-produced and may feel only partially authored by the professional (Saldanha & Krishnan, 2012). Recent work on AI-mediated authorship shows that GenAI can weaken this ownership unless deliberate practices restore human imprint (e.g., Henkenjohann & Trenz, 2024). In Davenport’s expert and collaboration quadrants, the interpretive nature of tasks tends to foster stronger ownership than routine work. When LLMs participate in these phases, the loss or dilution of authorship is more salient, making the preservation of control central to sustaining agency.

To capture how responsibility and agency operate under these conditions, we conceptualize task-related ownership as a context-specific form of psychological ownership: the felt sense of responsibility, control, and accountability for the task episode and its consequences, regardless of whether intermediate content is AI-generated. Task-related ownership is expressed when professionals treat the task outcome as “my call” and enact ownership through scoping, verification, edits, documentation, and sign-off. Conversely, task-related ownership weakens when responsibility is displaced onto the model (“the AI wrote it”), reducing perceived control and accountability (Henkenjohann & Trenz, 2024).

Discretion in AI-assisted work

Lipsky’s professional discretion theory (1980) explains how individuals translate ownership motives into concrete oversight behaviors, accepting, modifying, or rejecting system output based on situational judgment. In AI-enabled environments, discretion is not evenly distributed—transaction work offers limited intervention points, whereas expert and collaborative work provide multiple opportunities to reassert control (Krakowski, 2025). This helps explain why oversight routines differ across task types even within the same AI system. We define discretion in AI-assisted work as the bounded latitude a worker has within a task episode to decide whether and how to use AI, how to prompt and constrain it, which quality thresholds to apply when accepting AI output, and when to escalate to human review (See Appendix A Construct Box: Ownership & Discretion in AI-assisted Task Episodes).

Toward an episodic, risk-sensitive view

Recent IS studies highlight the importance of calibrating human intervention to perceived risk (e.g., Cummings, 2014; Meilvang & Dahler, 2024). Risk-based oversight models predict that high-stakes, interpretive work prompts multi-layered checks, while low-risk, routine work invites lighter monitoring. In GenAI workflows, such escalation is often triggered by situational factors—e.g., regulatory stakes, client expectations—which alter how and when professionals intervene.

We conceptualize human agency in GenAI work not as a single post-hoc gate, but as a distributed sequence of episodic, risk-triggered interventions across the workflow. We read this through an IS contingency perspective (Tallon et al., 2019): there is no one “best” a concrete control practice embedded in the workflow that reduces error, misuse, or hidden delegation to GenAI—its effectiveness depends on fit with the situation (task risk, audience stakes, time pressure), and thus varies by episode.

Davenport’s archetypes anchor what kind of work is at stake and who holds it; task related AI ownership explains why maintaining control matters; discretion in AI-assisted work illuminates how workers act on that motivation; and risk-adaptive models clarify when intensity should change. This integrated lens allows us to theorize not only the presence of human agency in GenAI workflows, but also informs our analysis of three episodic oversight modes drafting, refining, and reviewing, (see “Coding and analysis” section/Cycle 2). By drafting, we refer to early-stage work in which professionals use an LLM to generate a first version or scaffold of a document (e.g., an outline, draft email, or list of options). Refining captures judgment-intensive revision, where workers modify, expand, or adapt AI-suggested text using their domain knowledge. Reviewing refers to multi-actor checking, coordination, and sign-off, where peers, managers, or compliance staff jointly inspect

AI-assisted content, decide what to keep or rewrite, and allocate responsibility for the final version. We use these three modes to analyze episodic human oversight in GenAI workflows, that is, how professionals intervene at different points in the process rather than only at a single end-point check.

Research design

To investigate how professionals sustain agency in GenAI-mediated workflows, we adopted a qualitative, interpretive case study design (Klein & Myers, 1999), which is well-suited for uncovering situated practices and sensemaking processes in emergent socio-technical settings. This approach (see Fig. 2) allowed us to capture both formal governance structures and lived, day-to-day oversight routines across organizational contexts.

Research setting and case selection

Our research examines two German technology firms that serve as critical cases (Patton, 2015) for studying early organizational adoption of Generative AI (GenAI) in knowledge-intensive work; data were collected between March and June 2025. We followed Patton’s *critical case*

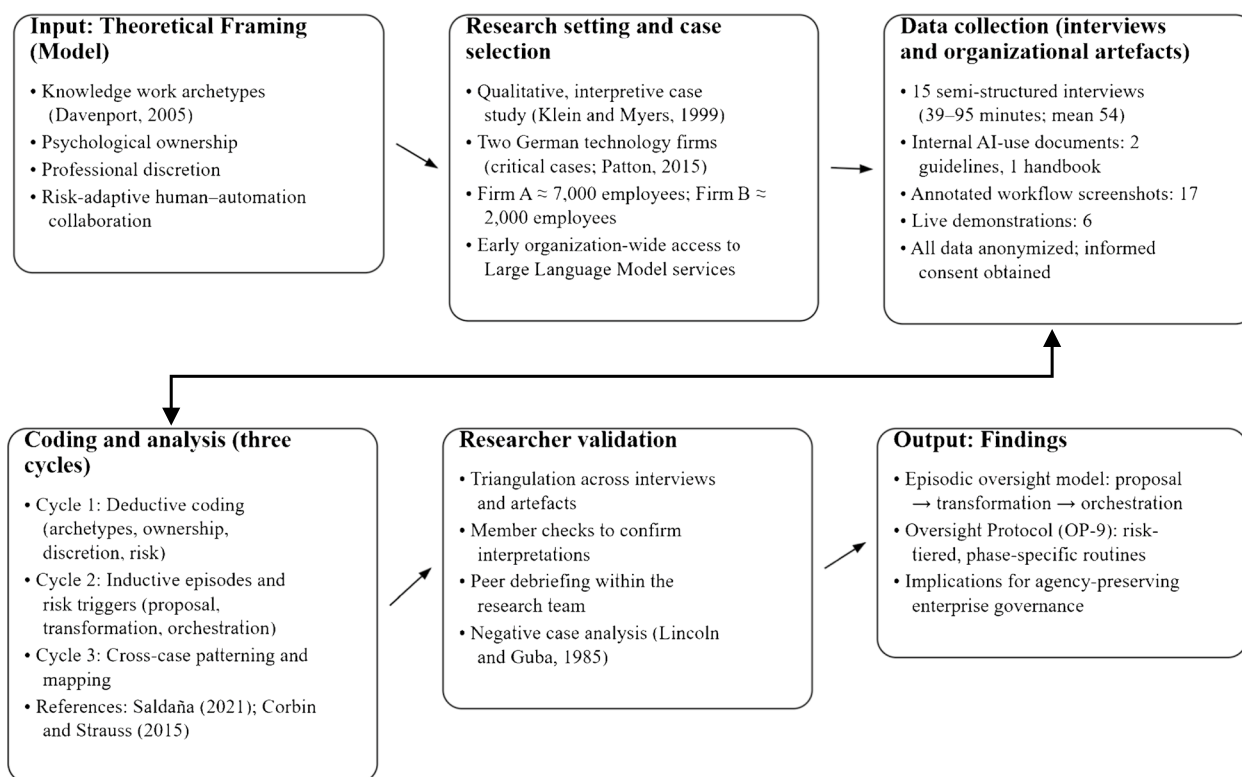


Fig. 2 Research process overview

approach, selecting sites where the phenomenon of interest is most likely to manifest in a pronounced form, on the reasoning that if ownership-sensitive human agency is observable here, under conditions of early, large-scale GenAI adoption, it is likely to be found in other knowledge-intensive settings as well. Firm A is a large IT services and consulting provider with approximately 7000 employees, while Firm B is a major application software vendor (SaaS/on-premises) employing about 2,000 staff.

We selected these organizations because they were among the first in Germany to deploy organization-wide LLM-based services into core professional workflows. These workflows span client drafting, technical documentation, and internal knowledge management, domains in which authorship, accountability, and quality assurance are highly visible. Across roles, participants described using LLMs for tasks such as drafting and revising client-facing proposals and emails, producing technical documentation (e.g., architecture notes, test cases, bug reports, release notes), and summarizing meetings or internal knowledge into handoff-ready briefs. Both firms operate under formal AI governance guidelines and professional handbooks, offering a documented and standardized view of AI-mediated practices.

This combination of early adoption, codified governance, and process transparency makes the two firms analytically

valuable for investigating how professionals negotiate human agency and ownership under real-world risk and productivity pressures—central to our research questions.

We purposefully sampled 15 professionals occupying roles that spanned the Davenport (2005) archetypes (transaction, integration, expert, collaboration), ensuring diversity in both task type and collaboration structure (Table 1).

Participants included consultants, data scientists, business managers, and software architects. Table 2 provides an overview of interviewee characteristics, including role, and tenure. This diversity enabled us to trace how agency and ownership practices varied by work type, collaboration mode, and perceived task risk.

Within each firm, we targeted roles spanning high- and low-risk deliverables (e.g., public-facing vs. internal documents) to ensure heterogeneity in risk triggers for oversight (Cummings, 2014). This theoretical heterogeneity sampling (Eisenhardt & Graebner, 2007) allowed us to examine oversight across the full spectrum of our conceptual framework.

We conducted 15 semi-structured interviews complemented by additional artefacts including internal AI-use documents (2 guidelines, 1 handbook), 17 annotated workflow screenshots, and 6 live demonstrations. Following

Table 1 Interviewee characteristics and participant codes

Firm	Interviewee Code	Sex	Position	Work experience	LLM usage frequency	Usage duration
A	I1	m	Software Developer/Consultant	Under 3 years	Daily	1–2 years
A	I2	w	Trainee Consultant	3–10 years	Monthly	Under 1 year
A	I3	m	Software Architect	more than 20 years	Weekly	1–2 years
A	I4	w	Senior Tester	10–20 years	Weekly	1–2 years
A	I5	m	Senior Manager	10–20 years	Weekly	1–2 years
A	I6	w	Business Developer	10–20 years	Weekly	1–2 years
A	I7	m	Senior Consultant	3–10 years	Daily	Under 1 year
A	I8	m	Consultant	3–10 years	Weekly	1–2 years
A	I9	m	Consultant	3–10 years	Weekly	1–2 years
A	I10	m	Software Engineer	More than 20 years	Daily	2–3 years
B	I11	w	Software Developer	3–10 years	Daily	2–3 years
B	I12	w	Business Manager	3–10 years	Weekly	1–2 years
B	I13	m	Consultant	10–20 years	Weekly	1–2 years
B	I14	m	Consultant	3–10 years	Daily	2–3 years
B	I15	m	Software Developer	3–10 years	Weekly	2–3 years

Table 2 The empirical analysis identified distinct oversight episodes that structure how participants monitored and adjusted generative AI outputs

Drafting	First LLM-enabled content scaffold; boundary ends when the worker shifts from generating to editing/verifying
Reviewing	Substantive editing + verification; boundary ends when content is routed outward for approval or integration
Refining	Peer/manager validation, tagging, sign-off, provenance capture, release coordination

established practice in qualitative IS research (Bowen, 2009; Flick, 2018; Yin, 2018), these artefacts (i.e., internal documents, screenshots, and demonstrations) served two purposes:

1. Episodic reconstruction: helping participants recall oversight moments by anchoring them in specific prompts, outputs, and edits.
2. Cross-validation: allowing us to verify reported practices against the actual outputs.

By triangulating narrative accounts with material traces, we reduced recall bias and captured the temporal ordering of drafting, refining, and reviewing activities.

Semi-structured interviews were guided by a protocol (see Appendix B) covering (i) tasks and touchpoints where LLM outputs enter deliverables, (ii) accept/modify/reject moments, (iii) experiences of responsibility/ownership, and (iv) agency/oversight routines (verification, peer review, prompt design). Interviews were held via Zoom (video), audio-recorded with consent, and lasted 39–95 min (mean = 54). Interviews were automatically transcribed and cross-checked against the audio recordings for accuracy. Prior to each interview, participants received data-protection information and provided consent for recording, transcription and anonymized use. All

identifiers (including people, teams, clients, and organizations) were removed.

Coding and analysis

Following Saldaña (2021), we analyzed our data in three iterative cycles (see Fig. 3) that combined a theory-informed deductive start with subsequent inductive elaboration. This sequencing enabled us to anchor the analysis in extant constructs (work archetypes, task related AI ownership, discretion, and perceived risk) while allowing episode structures and risk triggers to emerge from the material itself. All 15 interview transcripts and supporting artefacts (e.g., workflow screenshots, internal guidelines) were imported into MAXQDA.

Cycle 1—Deductive structuring Guided by our conceptual framing, we developed an a priori code family that captured: (a) Davenport’s knowledge-work archetypes (transaction, integration, expert, collaboration), (b) task related AI ownership levels, (c) discretionary actions over AI output (accept, modify-lite, modify-substantial, reject), and (d) perceived task risk (high/low). Using this codebook (see Appendix C), we applied first-cycle coding to segment interviews into meaning units and tag each excerpt with exactly one Ownership code and one Discretion code, alongside the relevant

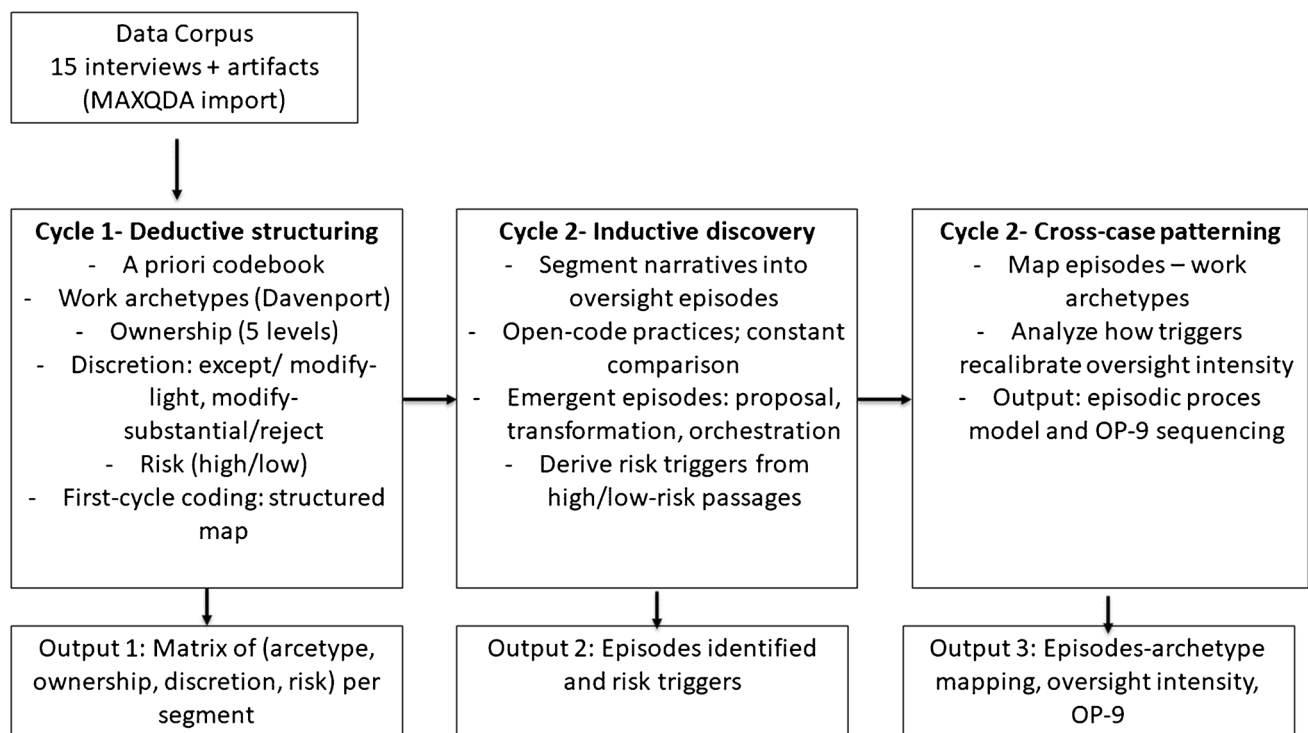


Fig. 3 Coding and analysis: three iterative cycles (Saldaña, 2021)

archetype and risk level. For example, “I pasted the paragraph exactly as ChatGPT wrote it—only fixed two typos” was coded as Ownership: Low/Delegated and Discretion: Accept, whereas “I rewrote every sentence so it sounds like me” was Ownership: Shared and Discretion: Modify–Substantial. This deductive pass provided a structured map of where (archetype), how (discretion), and under what stakes (risk) professionals intervened in GenAI-mediated work, and it created the foundation for tracing control patterns across tasks.

We operationalize ownership in five levels (no, low/delegated, partial/co-authored, shared/joint, and high/human) (Pierce et al., 2001) which map to distinct accountability claims and governance expectations. By contrast, discretion denotes the decision rights exercised over AI outputs; here the theoretically complete editorial choice set collapses into four mutually exclusive actions: accept, modify (light), modify (substantial), or reject (Lipsky, 1980).

To establish inter-coder reliability, two coders independently coded a 15% pilot subset; given the high agreement, the remaining corpus was coded by a single coder, with ambiguities resolved through consensus.

Cycle 2—Inductive discovery of episodes and risk triggers After first-cycle deductive coding (work archetype $\times$ ownership $\times$ discretion $\times$ risk), we segmented transcripts into oversight episodes using temporal cues in the narratives (e.g., “when I first generated the draft,” “after I started revising”) and corroborated episode boundaries with artefacts (e.g., workflow screenshots). We then inductively open-coded the practices within each episode (constant comparison) and conducted axial grouping by the episode’s position and function in the workflow. Across cases, these groupings consistently clustered into three coherent episodes: (1) early scaffold generation and boundary-setting (drafting), (2) judgment-intensive editing and verification (refining), and (3) multi-actor review, traceability, and release coordination

(reviewing) (see Fig. 4). We use ‘episode’ to signal an analytic unit that may repeat, overlap, or be compressed, rather than a fixed sequential stage model.

Figure 4 visualizes how mapping tasks to Davenport’s archetypes helped us distinguish the dominant oversight episode in each task context, which in turn guided the crystallization of drafting, refining, and reviewing as recurring process episodes.

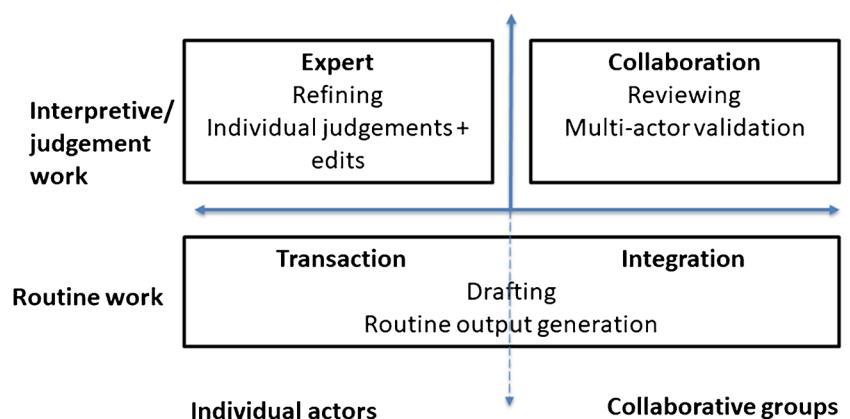
Cycle 3—Cross-case patterning and episode–archetype mapping

Following grounded-theorizing principles (Corbin & Strauss, 2015), we examined cross-case regularities in ownership and discretion by episode and archetype. Axial grouping showed that drafting episodes predominantly accompanied transaction/integration work (routine scaffolding and boundary setting), refining episodes concentrated around expert work (judgment-intensive revision and verification), and reviewing episodes aligned with collaboration work (multi-actor coordination, traceability, release). We then analysed how the inductively derived risk triggers reconfigured oversight intensity within and across episodes—for example, how client-facing audiences or legal implications amplified discretionary intervention even in otherwise routine tasks.

Deriving OP-9

After Cycle-2 open coding surfaced episode-specific practices, we conducted axial grouping to consolidate recurring moves into nine functionally distinct oversight tactics. For each tactic, we required: (i) ≥ 4 independent first-order code clusters across both firms, (ii) corroboration in at least one material artefact (guideline, screenshot, demo), and (iii) a decision rule specifying its risk trigger and boundary with adjacent tactics. We then cross-tabulated presence by episode and work archetype (Cycle-3) and ran a negative-case pass; tactics that failed these tests were either merged or dropped. The final nine tactics—OP-9—represent the stable set that met all criteria.

Fig. 4 Linking Davenport’s knowledge work archetypes to three oversight episodes (drafting, refining, reviewing)



We monitored concept emergence after each interview. We initially surfaced 14 candidate tactics; after negative-case analysis and boundary testing, we merged near-duplicates and dropped low-coverage items (<3 clusters), yielding the final nine. No new oversight tactics appeared after interview I13; interviews I14–I15 consolidated boundaries and supplied additional negative cases (e.g., situations where a move was not chosen despite high risk). We therefore treat OP-9 as the stable set of episode-specific discretion moves in our context, while noting that future settings may add or re-weight tactics.

OP-9 is our workflow-level oversight protocol that keeps GenAI use reviewable and accountable. It specifies (1) task scoping and risk classification, (2) input constraints (what data may or may not be shared), (3) output verification requirements, (4) documentation standards for traceability, and (5) escalation rules when uncertainty or risk exceeds predefined thresholds.

Researcher and validation

The research team comprised external IS scholars with no prior relationship to the firms. Our outsider status enhanced candor in interviews but required deliberate member checks to validate interpretations. Given the episodic nature of oversight, member checks were particularly valuable for confirming timing and triggering conditions. Credibility was further strengthened by the following steps (Lincoln & Guba, 1985):

- Triangulation across interviews, artefacts, and guidelines
- Peer debriefing within the research team to challenge initial categorizations
- Negative case analysis: we actively sought misfitting excerpts, discussed why they conflicted, tightened episode boundary rules (intent, actors, timing).

Results

We present how knowledge workers experience and negotiate human agency as LLM text enters their deliverables (RQ1) and which strategies sustain agency without eroding productivity (RQ2). We structure the Results around the three empirically derived episodes (drafting, refining, reviewing) and show how risk triggers reconfigure task-related AI ownership and discretion within each. We first show how ownership and discretion look in each episode, then we integrate them into a risk-tiered OP-9.

Task-related AI ownership

Our data analysis indicate that employees anchor their sense of authorship in the amount of post-generation editing required by the task's risk level. In the refining episode, across roles, workers reclaim agency through deliberate editing. When the model merges external sources into a cited paragraph, workers mark those lines and double-check provenance: "I flag the AI bits so the next editor sees exactly where the machine spoke." (Consultant, I8). Here, ownership feels partial and conditional on citation checks rather than on wholesale rewriting.

By contrast, text for external audiences receives intensive revision. A senior tester describes a three-step pass: "If it's going to a customer, I read it aloud, ditch the buzzwords, and double-check the numbers—otherwise it's the bot talking, not me" (I4). After these edits, employees consistently claim ownership of the prose: "It's my content that's been revised and optimized. It's not content that the tool created in terms of substance... it's an edited version" (I6, Business Developer).

After committing to revision, some workers (e.g., I2, I4–9) make a swift manual tone pass, adjusting terminology and cadence until the prose matches both their personal voice and corporate standards. They repeatedly use stylistic fit as a signal of ownership they owned the text: "It doesn't match my style... I'd phrase it differently, and I usually do. It's important that it's attributable to me if it's from me" (Senior Tester, I4). Under severe time pressure or for low-stakes internal notes, some cede credit to the system altogether: "That mock-up note is really the tech's success; I just hit send" (Consultant, I9).

Our analysis of organizational documents reveals the same risk-tiered approach to AI output seen in interview data. Those documents explicitly advise workers to subject high-risk deliverables (e.g., customer-facing copy) to thorough revision and provenance checks.

Not all ownership judgements are settled. Several interviewees (e.g., I1, I3, I7) describe a grey zone in which the system's contribution remains salient even after editing, making ownership feel partial: "I don't know if this can ever be 100% my result if the tool is involved." (Senior Consultant, I7).

This ambivalence is most pronounced when large sections of text are suggested by the model and only lightly adapted. A minority (I2, I9) externalize ownership to the technology when its imprint dominates the deliverable, typically in low-stakes contexts or when time pressure narrows the scope for revision: "It's more the technology's success; it doesn't have much to do with me." (Consultant, I9).

Importantly, several interviewees contrasted these experiences with pre-GenAI collaboration routines, in which ownership was implicitly tied to whoever authored the initial draft and coordinated subsequent rounds of feedback. Under GenAI, ownership and communication become more negotiated: workers must explicitly signal where LLMs contributed, invite targeted peer review on those segments, and articulate to colleagues which parts of the document they are prepared to “own.” Collaboration thus includes communicative work of boundary drawing around AI contributions, rather than assuming a uniform human authorship across the entire artifact.

Discretionary control: Layered quality assurance as the default

Our data analysis uncovers a *graduated quality-assurance* routine that expands only when task risk warrants additional scrutiny. A prevalent move is to walk the draft sentence by sentence, rewriting, pruning, or re-prompting as soon as something smells off. As one put it: “Really go through sentence by sentence... analyze, ‘How would I have done it if I had more time?’, and adjust from there” (Consultant, I8). Participants describe an active acceptance routine that combines *self-review*, *source checking*, and *peer verification* in proportion to risk.

First, eight of fifteen interviewees initiate rolling self-critique loops: they prompt the model to surface its weak spots and narrow their view to the flagged lines (e.g., I1, I3–4). An interviewee notes, “GPT highlights the rough edges; I zoom straight in” (Software Architect, I3).

Second, eleven of fifteen participants (e.g., I2–4, I7–11) run a source-credibility check: “Links ending in ‘.blog’ raise a red flag—I hunt for the original study and swap it in” (Senior Tester, I4). “If something reads like hard facts... I first check the links—are these reputable institutes or publishers? Then I adapt the text because it often sounds too ‘AI-ish’” (Consultant, I9).

Third, when audience risk is higher (e.g., customer-facing, leadership briefings), several participants (e.g., I1, I4) add a peer-review pass and apply AI-assisted margin tags to mark model-influenced lines, a procedure explicitly recommended in both companies’ AI guidelines: “before I pass this on, please check it... it feels much better” (Consultant, I9); “Tagging those lines is just courtesy—my teammate sees exactly where the machine spoke” (Consultant, I8). In practice, they treat the draft like a pull request: each tagged line functions as an inline comment, and at least two colleagues must “approve” before the text is finalized. Some participants (e.g., I1–3, I6, I9) explicitly point to these established review norms (e.g., structured diffs, inline feedback, and explicit sign-offs) as the template for their LLM-assisted

deliverables. As this layered process unfolds, ownership shifts from low/delegated toward shared/joint.

Notably, neither organization’s AI guidelines nor its employee handbooks prescribe a formal AI–human collaboration workflow. They simply emphasize that users bear ultimate responsibility for the accuracy and quality of any AI-generated content. For example, one guideline states: “All AI-generated text must be reviewed for factual accuracy and adherence to company style before publication” (Firm A).

Reviewing and coordinating LLMs in the workflow

Multi-step prompting remains the default for generic tasks such as summaries, outlines, and template text. However, for domain-critical passages (e.g., legal disclaimers, KPI commentary, and detailed architecture notes) participants abandon delegation and write the section themselves, judging that crafting an airtight prompt would take longer than simply drafting it, especially when time pressure was low (e.g., I6, I9).

Some participants (e.g., I2–3, I7) described that, as release nears, they append a *prompt-provenance stamp*, a practice that does not appear in any official handbook or AI-use policy. The final prompt and model version are stored invisibly in metadata to guarantee traceability. When deadlines are imminent, they adopt a micro-team split so one member refines content while another performs compliance checks in parallel; a delivery lead noted the team. An interviewee added that this split “turned last-minute stress into a checklist exercise” (Software Developer, I11).

To blunt the “fluency illusion” (the tendency to accept polished-sounding prose without scrutiny, linked to lower neural engagement and poorer recall), several participants (e.g., I3–4, I6) impose a one-minute pause when handling high-risk tasks—such as regulatory reports or client-facing documents—before deciding whether to keep the model’s output. This aligns with risk-based human–automation oversight principles, which recommend additional cognitive checkpoints under heightened task stakes (e.g., Sterz et al., 2024). The brief pause breaks the spell of first impressions and gives workers space to test factual soundness and stylistic fit.

A typical micro-workflow unfolds as follows: upload or paste source → request central points → apply the one-minute pause → verify alignment with intent → rewrite for the audience. Four participants add a “three-key-messages” pass as a further check before redrafting in their own words. Articulating the three takeaways the target audience should retain and auditing the draft so every paragraph advances at least one of them. This step recentres intent, curbs drift, and guides revision: “If I can’t clearly state the three key

messages for this audience, the draft isn't ready. I write them down, then rewrite so those points are unmistakable" (Business Manager, I12).

For public-facing deliverables, several participants at Firm B escalated oversight to the reviewing stage—instituting named human sign-off and explicit AI-use disclosures—consistent with the EU AI Act's expectations for human oversight and transparency (cf. Arts. 14 and 50).

Compared with earlier, non-AI workflows in these firms, where collaboration typically began only once a human-authored draft was ready to circulate, GenAI introduces an additional layer. Individuals first work with the LLM to generate and refine drafts, and only then invite colleagues into the process. When these AI-assisted drafts enter shared tools, collaboration and communication change their focus: instead of jointly producing every line of text, teams coordinate who will check which AI-influenced segments, where to apply stricter review, and who will sign off. Practices such as AI-margin tags, micro-team splits for last-minute review, and prompt-provenance stamps make these responsibilities visible and support joint decision making about whether the text is "ready" to leave the team. In this way, GenAI shifts collaboration from co-writing to co-governing AI-generated content.

Oversight protocol (OP-9): Strategies that sustain critical oversight

Building on our episode-level analysis, we identify nine discrete strategies (OP) that, in concert, maintain rigorous human review without sacrificing GenAI's speed advantages. We define an OP as a reusable oversight move (a discrete tactic) that can be applied within an episode; an episode may involve multiple OPs. Table 3 summarizes the OPs by mapping each practice to its typical risk trigger and its protective function for preserving agency and ensuring output quality.

Table 4 provides plain-language definitions of these terms to ensure consistent interpretation.

In the drafting phase, workers establish generation boundaries via risk-tiered delegation and no-delegate zones, forwarding only low-risk "standardized text" to the model (OP1–3 in Table 3). During refining, they apply rolling self-critique loops, a one-minute pause, tone-match skims, and a three-key-messages pass to reshape drafts so that form, fact, and focus all reflect human intent (OP4–7). Finally, in reviewing, teams layer in peer-review passes with AI-margin tags, embed an invisible prompt-provenance stamp, and—under tight deadlines—adopt a micro-team split to parallelize compliance checks (OP8–9).

OP-9 is modular, so teams apply only the practices required for a given AI use case and select a subset of the nine practices based on identified risk triggers.

Risk-tiered model of oversight

Across accounts, risk tier served as the trigger that activated these oversight practices. Risk triggers were inductively derived from interview accounts describing moments when participants intensified oversight, restricted delegation, or involved additional reviewers. Where the stakes were low, responsibility could be partially delegated and checks remained light. Where the stakes were high, responsibility re-centralized with the professional, supported by formal review and documentation steps. This pattern illustrates how oversight intensity was consistently scaled according to contextual risk, rather than applied uniformly across all tasks.

Across cases, we identified a set of recurring risk triggers that prompted participants to intensify oversight. These triggers included task and impact-related cues (e.g., critical or high-stakes outputs, external audiences), content sensitivity cues (e.g., legal or compliance-relevant text), model-related cues (e.g., uncertainty or fluent but suspicious outputs), and process-related cues (e.g., stylistic mismatch, loss of coherence, or time pressure). When such triggers were present, participants shifted from baseline practices toward more restrictive or collaborative forms of oversight, which informed the risk-tiered sequencing captured in OP-9.

Figures 1, 2, 3, and 4 establish the task context (archetypes) and the episode structure; Table 3 defines OP-9; Fig. 5 then integrates these elements into a risk-tiered operating flow. Figure 5 visualizes the decision logic through which professionals calibrate their oversight of GenAI outputs according to risk tier.

Rather than depicting oversight as a static "human-in-the-loop" checkpoint, the figure demonstrates a branching structure where the intensity of human intervention scales with the contextual stakes of the task. Low-risk cases—such as routine, reversible, or purely internal outputs—move through a lightweight path characterized by quick plausibility checks, stylistic skims, or forwarding standardized text with minimal friction. By contrast, high-risk cases—such as client-facing deliverables, legally sensitive material, or outputs with reputational consequences—trigger a cascade of stronger interventions: peer review with AI margin tagging, provenance stamping, parallelized team checks, and mandatory sign-offs. In this way, risk tier acts as the organizing principle that activates a layered set of safeguards proportionate to potential impact.

Although our core pattern is organized by task risk and workflow episode, the interviews indicate that these practices may also be shaped by person-related differences. Senior participants and those with stronger domain responsibility appeared more likely to impose stricter review routines and clearer no-delegate zones, especially for client-facing or high-stakes content. By contrast, frequent GenAI

Table 3 Oversight Protocol (OP-9)

<i>OP</i>	<i>Episode</i>	<i>Practice</i>	<i>Risk trigger</i>	<i>Ownership level</i>	<i>Protective mechanism (with example)</i>
<i>OP1</i>	Drafting	Risk-tiered delegation	Low vs. high task risk	No or Low ownership	Allocates human effort to high-risk content (e.g., draft customer emails yourself; forward slide headers)
<i>OP2</i>	Drafting	No-delegate zones	Critical or sensitive content	High ownership	Ensures human authorship for sensitive material (e.g., hand-write legal disclaimers rather than prompt the LLM)
<i>OP3</i>	Drafting	Forward standardized text	Routine, low-impact text	No or Low ownership	Maximizes speed where risk is minimal (e.g., send glossary entries unchanged)
<i>OP4</i>	Refining	Rolling self-critique loops	Model uncertainty	Shared → High	Targets review to AI-flagged weak spots (e.g., “GPT, highlight any rough phrases in this draft”)
<i>OP5</i>	Refining	One-minute pause	Fluency illusion	Low → Partial	Interrupts automatic acceptance (e.g., wait one minute before pasting to catch hidden errors)
<i>OP6</i>	Refining	Tone-match skim	Stylistic mismatch	Low → Partial	Reclaims authorial voice (e.g., adjust “Our solution...” to “I recommend...”)
<i>OP7</i>	Refining	Three-key-messages pass	Coherence or focus risk	Partial	Aligns structure to core takeaways (e.g., list three key points and ensure each paragraph maps to one)
<i>OP8</i>	Reviewing	Peer review + AI-margin tags	High audience or stakeholder risk	Partial → Shared	Enables shared validation and accountability (e.g., tag AI-generated lines and request colleague sign-off)
<i>OP9</i>	Reviewing	Prompt-provenance stamp & micro-team split	Imminent deadlines	High ownership	Preserves traceability and parallelizes checks (e.g., embed prompt/version in metadata; split tasks between members)

users did not necessarily delegate more; instead, they often seemed more fluent in using the tool while still varying in how much authorship they reclaimed. Because our study was not designed for systematic subgroup comparison, we treat these as indicative patterns that future research should examine more directly.

Discussion

We asked how professionals preserve human agency when LLM-generated text enters their deliverables. Our findings show that they do so by intervening at three points in the workflow: when deciding what the LLM may draft, when revising and checking its output, and when coordinating final review and sign-off with others. Across these episodes, agency is expressed through how workers claim ownership of AI-assisted text and how much discretion they exercise in accepting, modifying, or rejecting it. Importantly, the three workflow episodes are empirical abstractions that emerged

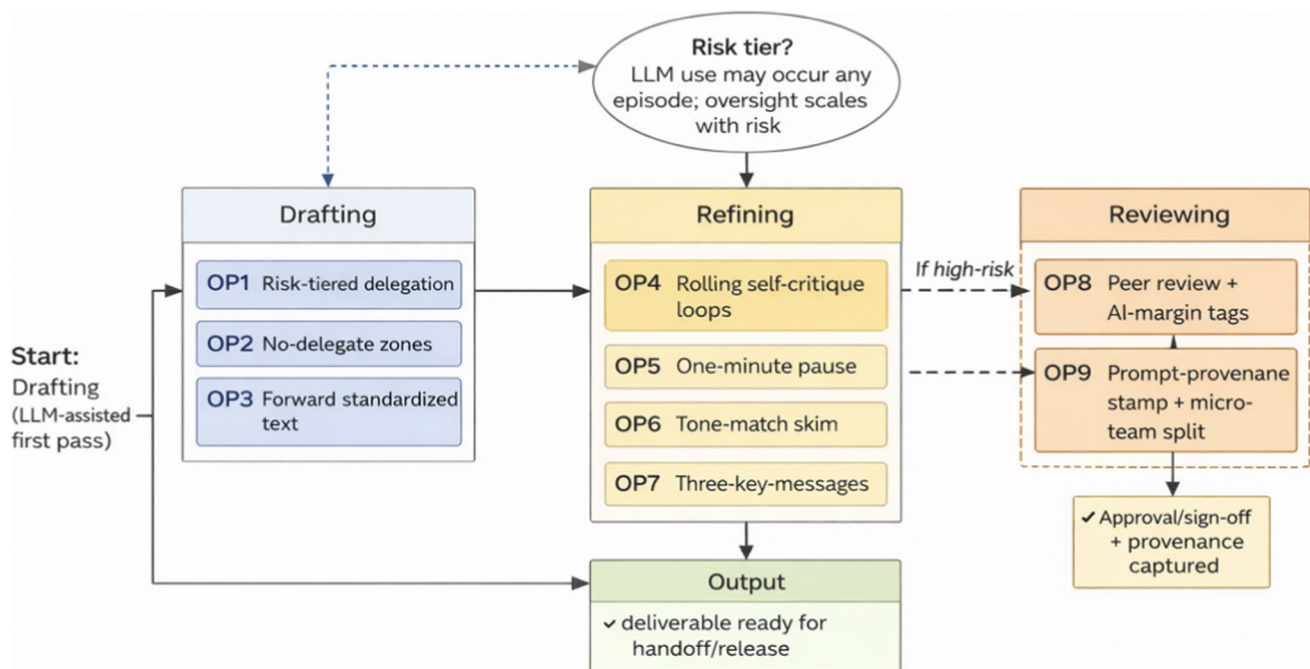
from work archetypes clustering rather than being pre-specified checkpoints. At drafting, workers draw clear boundaries. They decide up-front which content the AI may draft and which remains off-limits, a practice reminiscent of contingent automation in decision support (Ranerup & Svensson, 2023). This initial “gatekeeping” not only aligns with algorithmic discretion theory (Meilvang & Dahler, 2024) but also sets the stage for feeling responsible: when you choose what the machine handles, you claim what remains yours.

During refining, our participants turn into thoughtful editors. Self-critique loops, one-minute pause, tone-match skims, and three-message checks form a suite of cognitive scaffolds, echoing work on cognitive load management (Maleck & Gross, 2023). These lightweight rituals combat the “fluency illusion” and recalibrate attention, demonstrating that workers actively co-construct authorship by selectively accepting or reshaping model suggestions.

In the reviewing phase, accountability becomes collective. Teams borrow code-review norms (e.g., inline tagging, provenance stamps, parallel review splits) to forge

Table 4 Definitions of technical terms used in OP-9

OP1 Risk-tiered delegation	A rule that scales LLM permission and required oversight with task risk (more freedom in low-risk drafting, tighter limits and stronger review in high-stakes work)
OP2 No-delegate zones	Designated parts of a task where AI assistance is not permitted, ensuring direct human authorship for sensitive material (e.g., hand-writing legal or compliance disclaimers, final recommendations, or client-facing claims rather than prompting the LLM)
OP3 Forward standardized text	Keeping pre-approved, low-risk elements unchanged or inserting them from standardized templates (e.g., boilerplate, headers, disclaimers, glossary terms), rather than letting the LLM rewrite them
OP4 Rolling self-critique prompt (loops)	A follow-up prompt that asks the model to list weaknesses, uncertainties, or checks needed (used as a support tool, not as proof of correctness)
OP5 One-minute pause	A short, intentional pause (about one minute) before copying AI text into a deliverable, to reduce “automatic acceptance” and trigger a quick credibility and tone check
OP6 Tone-match skim	A fast-editing pass that aligns AI phrasing with the author’s voice and the organization’s style (e.g., replace generic sales language, remove buzzwords, adjust stance and claims)
OP7 Three-key-messages pass	Writing the three main takeaways, then edit the AI text so it supports them and remove anything off-message or overstated
OP8 AI-margin tag	Marking AI-generated or AI-edited passages (e.g., comments, highlights) so reviewers can see “where the model spoke.”
OP9 Prompt-provenance stamping	Saving the final prompt and the model/version (and, where possible, key settings) alongside the output (e.g., in document metadata, a log, or a ticket) so the AI contribution can be traced later
OP9 Micro-team split	Under tight deadlines, the team splits into two people working in parallel, where one refines the content while the other runs compliance/quality checks (e.g., policy, legal, client-risk). The point is to keep speed without skipping oversight

**Fig. 5** Risk-adaptive oversight

a shared audit trail. This improvised governance bridges text work and software development, extending digital traceability research (Elish & boyd, 2018) into the realm of prose.

Our findings also clarify how collaboration and communication differ between traditional and GenAI-mediated workflows. In traditional settings, collaboration was organized around exchanging human-written drafts and comments, and

communication focused on improving content that everyone assumed to be authored by humans. In the GenAI setting, we observe a two-step pattern: professionals first collaborate with the LLM to create a draft, and only then begin human-to-human collaboration around that AI-assisted text. As a result, communication within teams shifts from “What should we write?” to “Which parts did the LLM shape, what do we keep or change, and who is prepared to stand behind the final version?” GenAI thus reconfigures, rather than simply increases or decreases, collaboration by changing what must be jointly discussed: authorship boundaries, oversight duties, and acceptable reliance on AI.

While we anticipated that workers would seek to retain authorship in all circumstances, our findings indicate that in contexts characterized by stringent time constraints or low task criticality, some participants consciously relinquished full ownership—delegating entire sections or notes to the AI with minimal deliberation. This behavior runs counter to classic psychological-ownership models (Pierce et al., 2001), which emphasize a near-universal drive to claim stewardship. We theorize that, in these moments, throughput imperatives trump ownership, revealing a boundary condition: when speed becomes paramount, agency can be voluntarily relinquished. Unpacking this tension opens new lines of inquiry into how professional identity adapts when AI “does the heavy lifting.”

Our episodic process model dovetails with everyday work routines (Vaccaro et al., 2024) by showing how material-technical affordances (prompts, tags, metadata) and human intentions are woven into daily practice. It also enriches Tallon et al.’s contingency theory in IS (2019), underscoring that no single oversight approach fits all contexts and practices must flex in response to task risk, audience stakes, and time pressure. Finally, by spotlighting improvised traceability tactics, we signal a need for future platforms to natively embed these accountability rituals, bridging the gap between policy pronouncements and on-the-ground routines.

These findings suggest fertile ground for future work: How might AI interfaces better automate the one-minute pause or three-message check without burdening users? Will habitual delegation under time pressure erode professional craftsmanship over time? And how do cultural norms shape the threshold at which workers switch from human drafting to AI delegation? By posing these questions, we chart a path forward for research that ensures GenAI remains a partner in human creativity rather than a silent usurper of agency.

Theoretical contributions

Our main contribution is to show how GenAI can change, rather than simply erode, human agency by allowing workers to shape ownership and discretion at different episodes of the workflow. Our OP-9 demonstrates that effective oversight is

phase-specific and risk-sensitive, preserving accountability, complementing governance work that stresses human oversight and operational guardrails under the EU AI Act and related regimes (Holzinger et al., 2025).

Recent IS/governance accounts commonly discuss human oversight as a discrete checkpoint, typically at release, disclosure, or provenance documentation, without detailing how oversight unfolds inside day-to-day work (e.g., Botero Arcila, 2024; Cancela-Outeda, 2024). While these approaches stress the importance of human agency, they rarely specify when and where interventions should occur within everyday work. By contrast, we explain how episodic oversight sequencing: deliberate human interventions at drafting → refining → reviewing, calibrated to task risk, and we show how this sequencing changes collaboration and communication patterns in GenAI workflows compared with traditional, non-AI settings. This extends governance-oriented discussions by specifying the temporal structure of oversight within knowledge-intensive workflows and aligns with human-AI interaction work calling for designed feedback/oversight loops rather than one-off checks.

Prior IS research in automation-rich settings highlights shifting discretion and accountability, but tends to treat agency in broad terms (e.g., failure typologies and institutional/governance responses) (Krakowski, 2025). We refine this by introducing ownership calibration: worker-driven practices (e.g., no-delegate zones for sensitive text, micro-team splits near deadlines, prompt-provenance stamping) that keep authorship and decision rights with humans while automation drafts.

By positioning oversight steps that depend on who is responsible and when they step in, our study shifts the question from whether AI displaces human agency to how professionals actively negotiate and embed it. This offers a theoretically grounded basis for designing socio-technical systems that integrate automation without eroding professional authorship, in line with current IS/governance work emphasizing sustained human oversight under real organizational constraints (Holzinger et al., 2025).

Practical implications

Our findings carry several implications for organizations seeking to integrate GenAI into knowledge-intensive workflows without eroding professional authorship or decision rights.

Based on our findings, we recommend that rather than imposing generic or end-point review mandates, managers can embed *episodic oversight sequencing* into GenAI workflows. This means specifying when and how human interventions should occur across drafting, refining, and reviewing phases, with review intensity calibrated to task

risk. Tool designers can support this by providing built-in “intervention markers” and prompts at risk-critical junctures.

Our OP-9 highlights worker-driven mechanisms, such as no-delegate zones, micro-team splits under deadline pressure, and prompt-provenance stamping, that keep decision rights with humans while automation drafts content. Leaders should codify these routines into team charters and standard operating procedures, making them part of the organization’s governance toolkit rather than ad-hoc coping strategies. This also helps distribute authorship responsibility transparently in multi-contributor settings.

Many current AI use policies focus on compliance checklists (e.g., disclosure, bias mitigation) but fail to reflect the real decision points in day-to-day work. A concrete illustration arises in the EU AI Act’s generic compliance duties. Article 50 imposes uniform transparency obligations—requiring that users be informed when interacting with AI systems and that synthetic audio, image, video, or text be marked as artificially generated—irrespective of workflow stage. Article 10 prescribes dataset governance and bias management (e.g., examination for potential biases and measures to detect, prevent, and mitigate them) as baseline requirements for high-risk systems. These provisions articulate *what* must be disclosed or audited, but not *where* adaptive oversight should occur across drafting, internal review, and release, thereby exemplifying a checklist-oriented, one-size-fits-all approach. By grounding governance in actual workflows (as surfaced by our empirical episode analysis) organizations can bridge the gap between policy intent and worker experience, improving adherence and reducing “shadow practices” that circumvent oversight.

Our evidence suggests that dynamic, phase-sensitive oversight better matches the realities of fast-cycle knowledge work. Organizations can use these insights to shape industry standards and compliance frameworks that safeguard agency without imposing productivity-draining bottlenecks.

Taken together, practitioner and policy maker should view GenAI integration not as a binary choice between automation and human control, but as an opportunity to operationalize ownership-sensitive oversight. By doing so, they can leverage GenAI’s generative capacity while reinforcing professional accountability, a balance that is increasingly critical in the evolving future of work.

Implications for digital platforms and governance

Our study reveals how oversight routines that change with task risk can inform—and be embedded within—digital

platform design and governance, ultimately strengthening accountability, trust, and operational integrity.

1. Embedding context-sensitive oversight into platform infrastructure

OP-9 demonstrates that oversight should be both granular and dynamic, adapting to project phases, sensitivity of content, and contextual risk. Platforms should embed these routines as built-in affordances: for instance, configurable checkpoints that trigger mandatory human review before AI-generated content advances from drafting to client deliverables, or phased prompts that guide users through validation steps aligned with the episodes of drafting, refining, and reviewing.

2. Enabling graduated governance at the ecosystem level

Instead of applying the same rules to every task, platforms can use OP-9 to set different levels of review and checking based on how risky the task is. In line with risk-based governance, the amount of review is tailored to the context, the type of content, and what is at stake in the final output. Such differentiation empowers platforms to self-regulate with flexibility, ensuring more intense oversight where stakes are highest while preserving efficiency elsewhere.

3. Designing governance mechanisms for transparency, accountability, and collective responsibility

Drawing from OP-9, platforms can enable three core governance enhancements:

Traceability and auditability: Capture metadata that links AI-generated outputs to prompts, prior human edits, and review histories. This ensures visibility into who did what, when, and why.

Risk-tier dashboards: Provide administrators with configurable tools to set oversight thresholds dynamically—by document type, client risk level, or regulatory exposure—enabling real-time governance tuning.

Institutionalized collaborative oversight: Incorporate features like peer-review workflows (similar to software pull requests) for AI-generated content. This fosters shared accountability, distributed responsibility, and enhanced trust across teams.

4. Bridging micro routines and platform-level governance

By transposing individual oversight patterns into platform capabilities, OP-9 helps bridge micro-level practices with platform design. These routines serve as underlying practices for platform governance—comparable to how robust Electronic Markets research translates individual or organizational practices into broader governance models (Degen & Gleiss, 2025; Springer et al., 2025). In doing so, our findings lay the groundwork for digital platforms that operate with both pro-

ductivity and integrity, reinforcing stakeholder trust and regulatory compliance without sacrificing agility.

Limitations and future research

Our study offers a situated account of how knowledge workers preserve agency when integrating GenAI outputs into professional deliverables, but several limitations shape the scope and generalizability of our findings. First, our empirical base consists of two German technology firms that were early adopters of enterprise-wide LLM access. While these “critical cases” (Patton, 2015) enabled us to observe mature usage patterns, they also reflect organizational cultures with above-average digital literacy and governance maturity. In addition, because both cases are located in Germany, our findings may also be shaped by the cultural context in which risk is interpreted and managed (Van Everdingen & Waarts, 2003). This matters because our model is explicitly risk-sensitive: the thresholds at which workers intensify oversight, restrict delegation, or reclaim authorship may partly reflect culturally shaped preferences for dealing with uncertainty and ambiguity. Accordingly, the risk triggers and oversight routines we observed should not be assumed to travel unchanged across national settings. Future research should therefore examine whether our episodic oversight sequencing and ownership calibration patterns emerge similarly, or need adaptation, in less digitally mature, more resource-constrained, and especially more culturally diverse contexts.

Second, our data rely on retrospective interviews and artifact elicitation, which illuminate perceived ownership and discretion but cannot capture all tacit cognitive and affective dynamics. Controlled field experiments and fine-grained interaction logging could complement our qualitative insights by tracing how oversight decisions unfold in real time, potentially revealing micro-temporal patterns that retrospective accounts overlook.

Third, we focused on text-based GenAI workflows. Other modalities such as code generation, visual design, or multimodal reasoning may involve distinct ownership cues and oversight triggers. Extending our model to these domains would test its cross-modal robustness.

Fourth, while our OP-9 reflects worker-driven adaptations, its long-term impact on productivity, quality, and professional identity remains untested. Longitudinal research could explore whether habitual delegation under time pressure erodes ownership, or whether embedded oversight rituals become institutionalized as sustainable governance mechanisms. We note that OP-9 reflects a consolidated set of protocols observed in our cases, while additional protocols may emerge in other organizational contexts.

Finally, our lens privileges the perspective of human users and their immediate teams. Future work should integrate multi-stakeholder perspectives including clients, compliance officers, and system designers to understand how ownership and discretion are negotiated across organizational boundaries and how these negotiations shape trust, accountability, and regulation compliance. Relatedly, our data suggest that episodic oversight may also vary with person-related characteristics such as role seniority, domain experience, and familiarity with GenAI. Because our study was designed to theorize workflow episodes rather than compare worker profiles systematically, these patterns remain indicative. Future research could examine more directly how individual differences shape delegation thresholds, authorship claims, and oversight intensity across GenAI-supported work.

By addressing these limitations, future studies can refine our theorization of ownership-sensitive oversight and inform the design of socio-technical systems that scale GenAI’s benefits without undermining professional authorship and agency.

Conclusion

This study shows how GenAI can reconfigure, rather than erode human agency by enabling workers to calibrate ownership across the workflow. Using case evidence from two German technology firms, we develop the OP-9, a set of nine phase-specific, risk-sensitive practices that preserve speed while safeguarding authorship and accountability. By theorizing episodic oversight sequencing and ownership calibration, we extend governance discussions (Botero Arcila, 2024; Cancela-Outeda, 2024; Holzinger et al., 2025) beyond static end-point checks to the temporal and contextual structuring of human intervention in GenAI-mediated work. Our findings offer a process model for researchers studying socio-technical integration of automation and practical guidance for organizations seeking to embed GenAI without compromising professional control.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12525-026-00915-x>.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The data supporting the findings of this study are not publicly available due to confidentiality considerations but are available from the corresponding author upon reasonable request.

Declarations

Competing interests The authors declare that they have no known competing financial or non-financial interests that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Afzal, M., & Panagiotopoulos, P. (2025). How public officials perceive algorithmic discretion: A study of status quo bias in policing. *Public Administration Review*. <https://doi.org/10.1111/puar.13957>
- Botero Arcila, B. (2024). AI liability in Europe: How does it complement risk regulation and deal with the problem of human oversight? *Computer Law & Security Review*, 54, Article 106012. <https://doi.org/10.1016/j.clsr.2024.106012>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Cancela-Outeda, C. (2024). The EU's AI act: A framework for collaborative governance. *Internet of Things*, 27, Article 101291. <https://doi.org/10.1016/j.iot.2024.101291>
- Corbin, J., & Strauss, A. (2015). Basics of qualitative research: Techniques and procedures for developing grounded theory (4th ed.). SAGE Publications.
- Cummings, M. M. (2014). Man versus machine or man + machine? *IEEE Intelligent Systems*, 29(5), 62–69. <https://doi.org/10.1109/MIS.2014.87>
- Davenport, T. H. (2005). *Thinking for a living: How to get better performances and results from knowledge workers*. Harvard Business Press.
- Degen, K., & Gleiss, A. (2025). Time to break up? The case for tailor-made digital platform regulation based on platform-governance standard types. *Electronic Markets*, 35, Article 5. <https://doi.org/10.1007/s12525-024-00747-7>
- Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*. <https://doi.org/10.5465/amj.2007.24160888>
- Elish, M. C., & boyd, d. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs*, 85(1), 57–80. <https://doi.org/10.1080/03637751.2017.1375130>
- Flick, U. (2018). *An introduction to qualitative research* (6th ed.).
- Henkenjohann, R., & Trenz, M. (2024). Challenges in collaboration with generative AI: Interaction patterns, outcome quality and perceived responsibility. In *Proceedings of the 32nd European Conference on Information Systems (ECIS2024)*. Association for Information Systems. https://aisel.aisnet.org/ecis2024/track05_fow/track05_fow/3
- Holzinger, A., Zatloukal, K., & Müller, H. (2025). Is human oversight to AI systems still possible? *New Biotechnology*, 85, 59–62. <https://doi.org/10.1016/j.nbt.2024.12.003>
- Joshi, N., & Vogel, D. (2025). Writing with AI lowers psychological ownership, but longer prompts can help. *Proceedings of the 7th ACM Conference on Conversational User Interfaces*, 1–17. <https://doi.org/10.1145/3719160.3736608>
- Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, 23(1), 67–93. <https://doi.org/10.2307/249410>
- Krakowski, S. (2025). Human-AI agency in the age of generative AI. *Information and Organization*, 35(1), Article 100560. <https://doi.org/10.1016/j.infoandorg.2025.100560>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. CA: SAGE Publications.
- Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York, NY: Russell Sage Foundation.
- Maleck, M., & Gross, T. (2023). AnswerTruthDetector: A combined cognitive load approach for separating truthful from deceptive answers in computer-administered questionnaires. *I-Com*, 22(3), 241–251. <https://doi.org/10.1515/icom-2023-0023>
- Meilvang, M. L., & Dahler, A. M. (2024). Decision support and algorithmic support: The construction of algorithms and professional discretion in social work: Beslutningsstøtte og algoritmisk støtte: Konstruktionen af algoritmer og det professionelle skøn i socialt arbejde. *European Journal of Social Work*, 27(1), 30–42. <https://doi.org/10.1080/13691457.2022.2063806>
- Newell, S., Robertson, M., Scarbrough, H., & Swan, J. (2009). *Managing knowledge work and innovation* (2nd ed.). Palgrave Macmillan.
- Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). SAGE Publications.
- Pierce, J. L., Kostova, T., & Dirks, K. T. (2001). Toward a theory of psychological ownership in organizations. *Academy of Management Review*, 26(2), 298–310. <https://doi.org/10.2307/259124>
- Ranerup, A., & Svensson, L. (2023). Automated decision-making, discretion and public values: A case study of two municipalities and their case management of social assistance. *European Journal of Social Work*, 26(5), 948–962. <https://doi.org/10.1080/13691457.2023.2185875>
- Saldanha, T. J. V., & Krishnan, M. S. (2012). Organizational adoption of web 2.0 technologies: An empirical analysis. *Journal of Organizational Computing and Electronic Commerce*, 22(4), 301–333. <https://doi.org/10.1080/10919392.2012.723585>
- Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). SAGE Publications.
- Springer, V., Randhawa, K., Jovanović, M., Ritala, P., & Piller, F. T. (2025). Platform design and governance in industrial markets: Charting the meta-organizational logic. *Research Policy*, 54(6), Article 105236. <https://doi.org/10.1016/j.respol.2025.105236>
- Sterz, S., Baum, K., Biewer, S., Hermanns, H., Lauber-Rönsberg, A., Meinel, P., & Langer, M. (2024). On the quest for effectiveness in human oversight: Interdisciplinary perspectives. *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 2495–2507. <https://doi.org/10.1145/3630106.3659051>
- Tallon, P. P., Queiroz, M., Coltman, T., & Sharma, R. (2019). Information technology and the search for organizational agility: A systematic review with future research possibilities. *Journal of Strategic Information Systems*, 28(2), 218–237. <https://doi.org/10.1016/j.jsis.2018.12.002>

- Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, 8(12), 2293–2303. <https://doi.org/10.1038/s41562-024-02024-1>
- Van Everdingen, Y. M., & Waarts, E. (2003). The effect of national culture on the adoption of innovations. *Marketing Letters*, 14(3), 217–232. <https://doi.org/10.1023/a:1027452919403>
- Yin, R. K. (2018). *Case Study research and applications: Design and methods* (6th ed.). Publisher page (SAGE).
- Zhu, J., Xiao, M., Wang, Y., Zhai, F., Zhou, Y., & Zong, C. (2025). TROVE: A challenge for fine-grained text provenance via source sentence tracing and relationship classification. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 11755–11771). <https://doi.org/10.48550/arXiv.2503.15289>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.