Dancing with the algorithm: a framework to navigate knowledge and autonomy in AI-assisted managerial decisions

Matteo Cristofaro and Alexis J. Bañón-Gomis

Abstract

Purpose – *While artificial intelligence (AI) systems (assistive, human-in-the-loop decision support systems) increasingly participate in complex organizational judgments, their integration into knowledge processes raises fundamental challenges for autonomy, trust, and epistemic agency. This study aims to develop a dynamic, phase-specific framework that explains how autonomy evolves in relation to the data–information–knowledge–wisdom (DIKW) hierarchy and foundational knowledge management concepts during AI-assisted managerial decision-making.*

Design/methodology/approach – *The study draws on 122 in-depth interviews with senior professionals across diverse sectors, complemented by two expert focus groups. Data were analyzed using the Gioia Methodology to support inductive theory development and generate a grounded conceptual framework.*

Findings – *The authors develop the Human–AI autonomy loop (HAIL) framework, mapping decision-making to four recursive phases (frame, evaluate, commit, enact) and DIKW layers, each linked to distinct DIKW layers and autonomy configurations. Autonomy is a situated, distributed practice: managers preserve discretion through interpretive buffers, overrides and moral authorship. Trust in AI is recalibrated by phase, especially as decisions move from information to judgment. HAIL shows that autonomy is sustained through reflexive knowledge practices.*

Originality/value – *This study advances the knowledge management literature by integrating autonomy, trust and epistemic agency into a unified framework of AI-assisted decision-making. It reinterprets the DIKW model not as a linear information hierarchy, but as a socio-technical terrain where knowledge becomes actionable only when embedded in situated judgment and ethical authorship. The HAIL framework offers both theoretical insights and practical guidance for preserving human discretion and organizational wisdom in AI-assisted environments.*

Keywords *Decision-making, Autonomy, Knowledge management, Trust, Artificial intelligence, DIKW*

Paper type *Research paper*

Matteo Cristofaro is based at the Department of Management and Law, University of Rome Tor Vergata, Rome, Italy.
Alexis J. Bañón-Gomis is based at the Department of Management, Universitat Politècnica de València, Valencia, Spain.

Received 12 June 2025
Revised 3 September 2025
30 September 2025
Accepted 2 November 2025

Funding: This study was funded by the scholarship “Programa de apoyo a la carrera profesional del Profesorado” (“Program to Support Faculty Career Development”) from the Universitat Politècnica de València.

1. Introduction

Artificial intelligence (AI) systems – assistive, human-in-the-loop decision supports – are now woven into the infrastructures through which organizations generate, interpret and apply knowledge. Across marketing analytics, supply-chain planning, risk and compliance, finance, HR and customer operations, AI automates pattern recognition, classification, forecasting, retrieval and content generation. These capabilities accelerate knowledge flows, reconfigure how evidence is assembled and reshape when and why knowledge becomes actionable in managerial settings (Faraj *et al.*, 2018; Leoni *et al.*, 2024; Shrestha *et al.*, 2019). In practice, this means that what used to be slow, locally curated knowledge work is increasingly mediated by scalable pipelines – data ingestion and cleaning, model training and monitoring, promptable interfaces – and embedded in routines from dashboards and recommenders to copilot-style assistants. As these systems scale, they redistribute organizational attention, alter the thresholds for justification and recalibrate the boundaries of managerial discretion, epistemic authority and accountability.

Knowledge management (KM) research has richly catalogued technological affordances, but it has been slower to confront their epistemic implications. Much of the literature continues to treat AI as a neutral or efficiency-enhancing tool, emphasizing capabilities over the reconfiguration of judgment, legitimacy and ethical responsibility (Metaxiotis *et al.*, 2003; Rezaei *et al.*, 2024). Even critical accounts often preserve a static human–machine divide, under-specifying how AI influences what counts as relevant knowledge, whose interpretation prevails and how decisions are justified and legitimated (Jarrahi *et al.*, 2023). We use AI systems (AI hereafter) to denote socio-technical assemblies – models (e.g. machine-learning classifiers, recommenders, large language models), data pipelines, user interfaces and organizational routines – that generate assistive outputs for managerial decision processes.

A central construct in this reconfiguration is autonomy. We define autonomy as the situated capacity of decision-makers to interpret and act upon knowledge, taking responsibility for their choices and consequences, in ways that reflect their judgment, values and discretion. Autonomy is not a stable trait but a performative competence exercised under conditions of cognitive delegation to machine-generated outputs. This view aligns with a procedural understanding of autonomy as the ability to reflectively endorse and author one's actions within socially and technologically mediated contexts (Dworkin, 1988). Its salience is evident in managerial decision-making, which we understand as a structured process of framing issues, evaluating options, committing to action and implementing outcomes (Nutt, 2008).

The layered nature of knowledge further complicates autonomy in AI-mediated work. KM frameworks have long distinguished between data, information, knowledge and wisdom (DIKW) (Ackoff, 1989; Rowley, 2007). AI capabilities tend to dominate at the lower layers – where speed, scale and pattern extraction are more crucial – while offering thinner epistemic support at higher layers, where contextualization, normativity and ethical reasoning prevail. Yet, this layered configuration remains theoretically underdeveloped in KM when AI is involved. Foundational concerns – knowledge sourcing (internal vs external), exploration–exploitation tensions (March, 1991) and the creation, hiding, transfer or decoupling of knowledge (Connelly *et al.*, 2012; Sumbal *et al.*, 2017) – are rarely examined through the lens of AI mediation. We lack explanations of how AI refracts knowledge flows: what is surfaced, what remains hidden, which sources are privileged or marginalized and how recombination occurs under AI-infused workflows.

Taking stock of the state of the art reveals four interrelated gaps. First, autonomy in KM is typically theorized at a general level rather than as a practice that unfolds differently across DIKW layers; we know little about where and how autonomy is sustained or eroded as work moves from data to wisdom. Second, the mechanisms by which AI mediates knowledge sourcing and exploration–exploitation remain opaque: organizations increasingly blend internal and external data and models, but we lack accounts of how such blending reassigns interpretive authority and shapes legitimacy claims. Third, there is limited theorization of autonomy-preserving practices – the concrete routines through which decision-makers maintain authorship amid cognitive delegation (e.g. when to accept, resist or reframe AI outputs). Fourth, process models that connect autonomy and trust to the sequential phases of managerial decision-making are underspecified; most accounts remain static rather than recursive and phase-sensitive.

Guided by these gaps, we pose the following research question:

RQ1. How do human decision-makers experience and sustain autonomy across the knowledge processes involved in AI-assisted managerial decision-making?

We investigate this question through 122 interviews with senior professionals in Italian organizations using AI in managerial decisions, complemented by two expert focus groups. Using the Gioia methodology (Gioia *et al.*, 2013; Magnani and Gioia, 2023), we inductively build a grounded framework that conceptualizes decision-making as a recursive, four-phase process – frame, evaluate, commit, enact – each interacting differently with AI and mapping onto distinct layers of the DIKW hierarchy. We propose the Human–AI autonomy loop (HAIL) framework to explain how autonomy and trust are reconfigured across these phases. We suggest that AI systems are most effective at the data and information levels. In contrast, human discretion is crucial at the knowledge and wisdom levels, where contextual judgment and ethical reasoning take precedence.

This study offers three contributions to KM and adjacent debates. First, repositioning autonomy as knowledge-centric. We articulate autonomy as a layered, practice-based competence anchored in knowledge work rather than a generic individual trait, clarifying how authorship is negotiated when cognitive tasks are delegated to AI. Second, reinterpreting DIKW as a socio-technical loop. We model DIKW not as a one-way hierarchy but as a recursive circuit in which AI and humans coproduce flows across layers, with distinct autonomy and trust dynamics at each phase of managerial decision-making. Third, introducing epistemic friction as a capability. We conceptualize epistemic friction – structured resistance, challenge and recalibration of AI outputs – as a necessary organizational practice for sustaining autonomy and cultivating wisdom.

The broader significance is two-fold. For theory, centering autonomy, trust and epistemic agency advances sociotechnical perspectives in KM and connects microlevel cognitive delegation to mesolevel governance and legitimacy. For practice and society, the framework helps leaders design decision-making processes that specify nondelegable cores, calibrate human–AI trust and establish auditable routines for oversight – thereby protecting human agency and ethical accountability as AI becomes more infrastructural.

The paper proceeds as follows: Section 2 reviews relevant literature; Section 3 details methodology; Section 4 presents findings by phase; Section 5 introduces the HAIL framework; Section 6 discusses implications; Section 7 concludes.

2. Theoretical premises

2.1 Managerial decision-making, autonomy and trust in the age of artificial intelligence

Managerial decision-making is best understood as a boundedly rational, staged process rather than a single act of calculation. It unfolds through interdependent phases – preparation, evaluation, commitment and implementation – each with distinct attentional and justificatory demands (Cyert and March, 1963; Simon, 1972; Eisenhardt and Zbaracki, 1992; Gavetti, 2012; Kaplan and Orlikowski, 2013). Building on this lineage, we work with four phases: decision preparation, decision evaluation, decision-making and decision implementation (Berente *et al.*, 2021; Braun, 2025; Martin *et al.*, 2022). With this staged view in place, we can examine how contemporary AI systems reconfigure the expectations of each phase for managers.

As AI becomes embedded across these phases, managerial agency is no longer exercised in isolation but co-constituted through human–machine entanglements. In this sense, AI is not a neutral instrument but a performative agent that redistributes cognitive authority and accountability, reshaping the very architecture of decision processes (Leoni *et al.*, 2024). Crucially, these entanglements are mediated by KM practices – how knowledge is sourced, filtered, combined, hidden, transferred and legitimated (Alavi and Leidner, 2001; Jennex, 2017). Thus, before turning to each phase, we recast autonomy and trust in explicitly KM-sensitive terms.

Autonomy, in AI-infused contexts, is better understood as a situated, phase-sensitive competence rather than a stable trait. Classical accounts, which center on independent, reflective action (Dworkin, 1988; Christman, 2009), require extension when algorithms structure what counts as rational, relevant or legitimate. Following Prunkl (2024) and Catena *et al.* (2025), we distinguish autonomy-as-agency – the capacity to initiate and own decisions – and autonomy-as-authenticity – the alignment of choices with an actor's evaluative commitments. Both dimensions depend on how knowledge is accessed, interpreted and validated; autonomy is therefore inextricably linked to KM dynamics.

Trust must likewise be reframed as a performative, relational configuration shaped by transparency, fairness, and value alignment rather than by technical accuracy alone (Glikson and Woolley, 2020; Rai, 2020; Leoni *et al.*, 2024). Although "algorithm appreciation" shows managers may over-trust systems framed as objective (Logg *et al.*, 2019), such trust is often shallow and fragile, mediated by interface design, narrative framing, and institutional context (Bader and Kaiser, 2019; Brink *et al.*, 2024). With autonomy and trust re-specified, we can trace how they evolve across phases through KM mechanisms.

In the decision preparation phase, dashboards, anomaly detectors, and recommenders guide problem framing and relevancy cues, elevating some signals while obscuring others (Faraj *et al.*, 2018; Candrian and Scherer, 2022). Sourcing tilts from internal, tacit, path-dependent knowledge toward externally curated data streams and encoded heuristics (Nonaka and Takeuchi, 1995). This reweighting dampens exploration and reinforces exploitation (March, 1991). As algorithmic salience narrows the epistemic aperture, autonomy-as-agency is challenged at the very outset, shaping what will later be considered evaluable.

In the decision evaluation phase, simulations, rankings and predictive models foreground quantifiable, standardized metrics over contextual or qualitative knowledge (Rai, 2020; Gaur *et al.*, 2023). Here, algorithmic knowledge hiding can occur when disconfirming data are filtered out or when certain reasoning forms become invisible (Connelly *et al.*, 2012; Anand *et al.*, 2022). Exploitation logics intensify, crowding out morally complex considerations (Leoni *et al.*, 2024) and producing a "performativity of default" in which model outputs harden into *de facto* standards (Laitinen and Sahlgren, 2021). This evaluative narrowing sets the stage for the commitment moment.

At decision-making, converging trust and accountability pressures can induce delegated authorship – outsourcing rationales to AI to mitigate reputational risk (Leoni *et al.*, 2024; Braun, 2025). Preserving interpretive sovereignty, therefore, requires deliberate decoupling: strategically distancing human judgment from automated outputs (e.g. Chen *et al.*, 2021). As a KM move, decoupling prevents overfitting knowledge to tools and protects space for divergent reasoning. The commitment point can also be generative, integrating algorithmic outputs with narrative, contextual insight and ethical reflection (Nonaka *et al.*, 2000), thereby sustaining autonomy as authenticity.

In the decision implementation phase, AI assists execution, monitoring and feedback (Shrestha *et al.*, 2019). Yet implementation exposes tensions in knowledge transfer – the movement of experiential, procedural and reflective insights across actors and layers (Argote and Ingram, 2000). While AI can automate the how, it cannot ensure that the why – the causal rationale – and the what for – the purpose and intended outcomes – are communicated, retained and owned. This gap discloses a non-delegable core of managerial responsibility (Prunkl, 2024) tied to ethical framing, organizational memory and moral interpretation. Addressing it requires deliberate storytelling that conveys reasons and purposes, institutional scaffolds that embed them and a structured review that tests fidelity between actions and ends. Taken together, AI reshapes not only what managers decide but how decisions are made, justified and lived through phase-specific couplings with KM architectures. Autonomy thus becomes a knowledge-mediated, phase-contingent practice that must be actively performed and defended, while trust emerges as an institutional and interface-mediated accomplishment. This lays the groundwork for Section 2.2, which uses the DIKW schema to specify how data, information, knowledge and wisdom map onto the shifting autonomy–trust dynamics across the four phases.

2.2 Knowledge management, artificial intelligence and the architecture of managerial autonomy

KM provides a precise lens for understanding how human and algorithmic agents jointly shape managerial decision-making. It concerns how organizations generate, share, hide, recombine and apply knowledge (Nonaka and Takeuchi, 1995; Heisig, 2009; Connelly *et al.*, 2012). Building on the phased view presented in Section 2.1, we use the DIKW scaffold – data, information, knowledge, wisdom – to clarify how autonomy and trust evolve as epistemic complexity increases. In the discussion below, references to AI at the data level mainly concern perception/extraction and descriptive–predictive pipelines; at the information level, descriptive/predictive analytics and generative summarization tools dominate; at the knowledge level, prescriptive/optimization systems and generative assistants are most relevant; the wisdom level remains a domain of human phronetic judgment, with AI providing inputs but not substituting for moral purpose or accountability (Ackoff, 1989; Frické, 2009; Weinberger, 2010).

From signals to frames: the data layer. Data refers to raw, uncontextualized signals collected through sensors, transactions or digital traces. Core KM practices here include sourcing, curation, quality control, provenance and governance. AI excels in harvesting and classifying large, often external, repositories (Heisig, 2009; Malik *et al.*, 2019). Yet this external orientation can crowd out context-rich, internal, tacit knowledge embedded in routines and experience (Nonaka and Takeuchi, 1995). Autonomy implications mirror those in Section 2.1: upstream problem framing risks being deskilled, as pipelines optimized for scale privilege exploitation over exploration (March, 1991). Design choices at this layer enact epistemic compression – narrowing managerial attention toward routinized categories and away from generative ambiguity.

Making patterns legible: the information layer. Information arises when data are cleaned, structured and visualized as patterns, trends or anomalies, often via dashboards, heat maps and risk scorers (Davenport and Prusak, 1998). The step from data to information is not purely technical; it encodes assumptions about relevance and legitimacy. Algorithmic knowledge hiding can occur when interfaces or model logics obscure provenance, caveats or alternative framings (Connelly *et al.*, 2012; Anand *et al.*, 2022). This matters for autonomy-as-agency: when decision-makers cannot interrogate preselections about what counts as rational or salient, their capacity to own the framing and evaluation narrows. Trust also bifurcates: statistical accuracy can produce surface confidence, while the relational basis of trust – transparency, fairness, interpretability – remains fragile (Glikson and Woolley, 2020).

From interpretation to discretion: the knowledge layer. Knowledge entails context-sensitive understanding that connects informational patterns to situated action (Nonaka and Takeuchi, 1995). Human sensemaking becomes central at this level, even as AI systems simulate "knowledge" through scenario generation, inferential modeling or natural language outputs. Two tensions follow. First, AI lacks narrative intentionality and moral reflexivity (Floridi, 2011). Second, pressures toward conformity can produce delegated authorship or a "performativity of default," dulling discretionary judgment. Maintaining interpretive sovereignty therefore requires organizationally legitimated decoupling – selective distancing of human judgment from automated outputs to prevent overfitting knowledge to tools and to preserve space for divergent reasoning (Martinez and Cooper, 2019). Where decoupling is viable, autonomy-as-authenticity – alignment with evaluative commitments – can be articulated and defended.

Embedding purpose and accountability: the wisdom layer. Wisdom integrates judgment, values and consequence-sensitivity; it operates holistically and with temporal depth (Ackoff, 1989; Rowley, 2007). KM research frames wisdom as nonalgorithmic, emergent from reflective learning cycles, narrative reframing and experiential abstraction (Awad and Ghaziri, 2004; Bootz *et al.*, 2015). AI can assist pattern detection and feedback, but it cannot determine moral purpose or resolve institutional ambiguity on its own. At this apex, autonomy is inseparable from epistemic responsibility: to act is also to justify action in ways that are ethically defensible and institutionally coherent (Choo, 1996). Over-delegation to AI risks hollowing this responsibility; conversely, designing AI–human systems that protect narrative discretion and moral agency sustains decision quality under complexity (Leoni *et al.*, 2024; Prunkl, 2024). Knowledge creation and transfer – mobilizing heterogeneous experiences across levels – are crucial here (Argote and Ingram, 2000; Sumbal *et al.*, 2017).

Read as a scaffold rather than a staircase, DIKW clarifies how AI's strengths at lower levels (efficiency, scale, patterning) can narrow interpretive space, while higher levels demand human autonomy for meaning-making, value alignment and accountability. This layered view complements Section 2.1's phase analysis: data-centric sourcing and governance shape preparation; information filtering and visibility condition evaluation; knowledge-level discretion and decoupling enable commitment; and wisdom-oriented stewardship underwrites implementation, learning and memory.

3. Methodology

3.1 Research design and data collection

This study examines how human autonomy is enacted, diminished or safeguarded in AI-assisted managerial decision-making, using a qualitative research design grounded in interpretivist epistemology. The empirical contexts in our sample span four widely used classes of AI systems: predictive and descriptive analytics (machine learning and data mining), including classification, forecasting, anomaly detection and clustering. Prescriptive/optimization and recommender systems, such as resource allocation, scheduling, pricing and next-best-action engines. Generative AI assistants (large language model–based), used for drafting and summarization, ideation, code or copy generation and retrieval-augmented responses. Perception and extraction technologies, including natural language processing, information extraction, optical character recognition, speech-to-text and computer vision, transform unstructured inputs into structured data. Throughout, we use the term "artificial intelligence (AI)" systems to collectively refer to these assistive classes. We deliberately exclude safety-critical, fully autonomous control loops (e.g. real-time plant control without human oversight), as our focus is on AI that augments rather than replaces managerial judgment.

We adopt a Gioia-inspired grounded theory approach (Gioia *et al.*, 2013; Magnani and Gioia, 2023), which is particularly appropriate given our aim to understand how autonomy unfolds *across distinct phases of decision-making* and in tension with machine-generated outputs. Our research question – *How do human decision-makers experience and sustain autonomy across the knowledge processes involved in AI-assisted managerial decision-making?* – requires an approach capable of capturing evolving perceptions, situated behaviors and context-dependent meanings. The Gioia Methodology provides the necessary structure to move from rich, informant-centered data to theoretically informed explanations, while preserving the complexity of lived experience. It is especially well-suited for studying dynamic, emergent processes, such as autonomy, which are enacted rather than possessed. Our design combines in-depth semistructured interviews with expert focus groups to support inductive theorizing and iterative validation of findings; see Table 1. Specifically, we:

Table 1 Data collection and data analysis step

Phase	*Activity*	*Outcome*
Questionnaire development	Developed a semi-structured questionnaire grounded in a comprehensive review of the literature on decision-making, human autonomy and AI use in managerial contexts. The questionnaire was designed to capture subjective, experience-based insights from senior professionals	A final instrument composed of two sections: (1) socio-demographic information and (2) ten open-ended questions addressing human–AI autonomy dynamics
Pilot test and prevalidation	Conducted a pretest with a convenience sample to refine the questionnaire. Two academic experts in decision-making and two senior professionals with practical AI experience participated. Their feedback informed iterative revisions to ensure methodological clarity and validity	Finalized and validated version of the questionnaire, refined for conceptual clarity, relevance and ease of response
Sample identification and data collection	Identified large enterprises via the AIDA database based on European Commission and Italian criteria. Sent invitations to 1,256 top managers across sectors. Data were collected through self-administered interviews distributed via e-mail across three mailing waves	122 valid interviews collected and screened, yielding a rich data set for qualitative analysis. Nonresponse bias was tested and ruled out through statistical comparison and follow-up calls
Data analysis (interviews)	We applied the Gioia Methodology (Gioia *et al.*, 2013; Magnani and Gioia, 2023) to inductively code and analyze 1,262 meaning-rich responses. First-order concepts were derived from participants' terms, then aggregated into second-order themes and theoretical dimensions. NVivo was used to support coding, organization and transparency throughout the process	67 codes and 23 latent themes identified, organized across four recursive phases of decision-making (Frame, Evaluate, Commit, Enact) and aligned with the DIKW model
Focus group triangulation and theory building	Held two follow-ups focus groups to validate and enrich findings: one with five academic experts in AI and decision-making, and one with five senior managers from diverse industries (e.g. telco, IT, FMCG, manufacturing, services). findings were compared with interview-based themes to assess coherence, divergence and conceptual robustness	Data triangulation confirmed the four-phase model of the human–AI autonomy loop (HAIL) and supported theory refinement by integrating practitioner and academic validation

Source(s): Authors' own work

- designed a set of ten open-ended questions based on key debates in the literature on AI and human autonomy;
- pilot-tested and refined this instrument;
- recruited a diverse sample of senior professionals using AI in managerial roles;
- conducted and thematically analyzed 122 interviews following the Gioia Methodology's principles of first-order coding, second-order theorizing and aggregate dimension building; and
- conducted two focus groups to refine and test the emerging theoretical model.

This multistage design allows us to construct a grounded, explanatory process model of autonomy in AI-assist decision-making.

Importantly for traceability, we analyze experiences across a four-phase decision structure – decision preparation, decision evaluation, decision-making and decision implementation – and across DIKW levels (data, information, knowledge, wisdom). Interview prompts were neutrally worded (i.e. not phase-labeled) to avoid leading respondents. We subsequently attributed excerpts to phases and DIKW levels using explicit coding rules and an auditable mapping of protocol items to the intended phase/level.

The questionnaire was composed of two sections. Within the first, we asked respondents to report their sociodemographic info regarding: age, gender, job position, industry, and level of experience with AI.

The formulation of the open-ended questions, included in Section 2, was informed by a comprehensive review of the current literature on the interplay between human judgment and AI assistance in decision-making and autonomy. This review guided the identification of key themes relevant to the study's focus. We developed ten open-ended questions, each one explicitly linked to academic sources (see Supplementary Material SM1 for the ten questions and Supplementary Material 2 for a map of them to intended decision phases and DIKW levels). The questions were refined through an iterative process to ensure clarity, relevance and alignment with the study's objectives.

During recruitment, participants confirmed hands-on use of at least one of the AI classes already defined. In interviews, we first used the neutral term "AI system(s)" and then probed system-specific workflows (e.g. predictive dashboards vs generative copilots). In the analysis, the AI class was recorded as a case attribute and used in constant comparison to examine whether autonomy experiences differed for generative, predictive/prescriptive and perception/extraction contexts. Findings reported as generalizable to "AI systems" reflect themes consistent across classes; class-specific nuances are named explicitly. The questionnaire was validated through pilot testing with two academic experts and two experienced managers, refining questions to enhance clarity and content validity.

To construct the sample, large-sized enterprises were selected using the AIDA (*Analisi Informatizzata delle Aziende Italiane*) database. AIDA is a Bureau van Dijk/Moody's Analytics database of Italian companies compiled from official registries and statutory filings. It provides standardized multiyear financials, firmographics (economic activities/statistical classification of economic activities in the European community, location, legal form), ownership/management information and adverse events. The selection criteria were as follows: firms had to be:

- currently active;
- headquartered in Italy;
- employ more than 250 people; and either
- report annual revenues exceeding €50m or possess total assets over €43m – thresholds consistent with the European Commission's and Italian law's definition of large-sized enterprises.

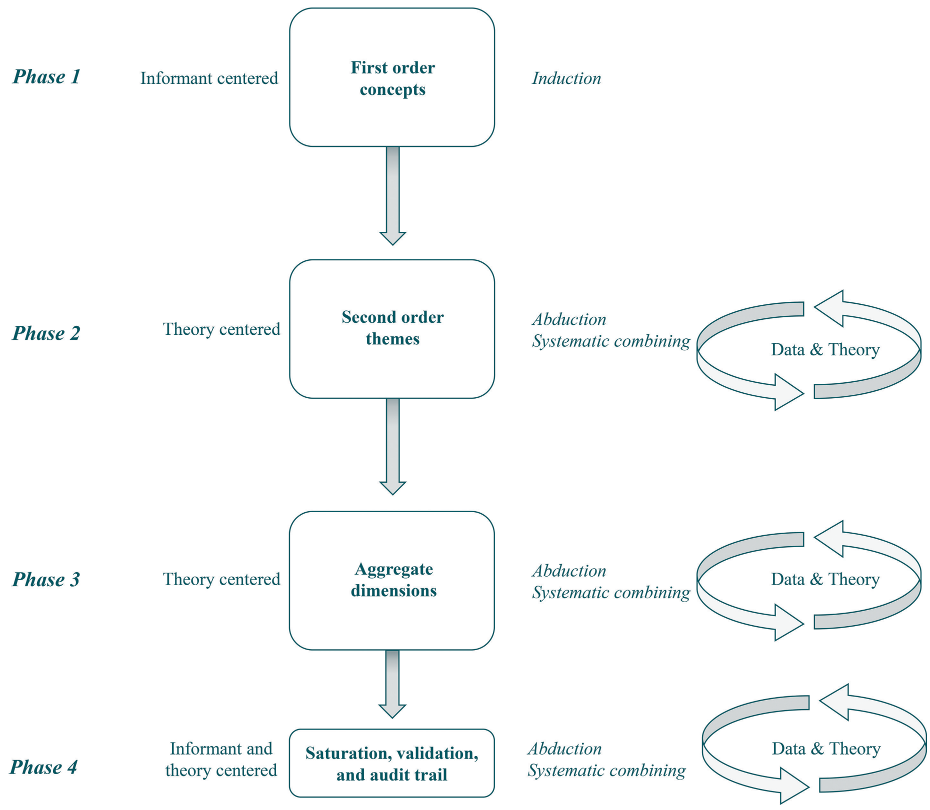
In addition, firms were included only if they provided accessible contact information for their human resources or administrative departments (i.e. a valid e-mail address).

3.2 Data analysis

Data analysis followed the principles of the Gioia methodology (Gioia *et al.*, 2013; Magnani and Gioia, 2023), which assists the development of inductive theory through a structured engagement with informant accounts and researcher-driven abstraction; see Figure 1. This approach was selected for its capacity to uncover how autonomy is experienced, constrained or preserved in AI-mediated managerial decision-making across varying levels of knowledge abstraction.

We attributed each excerpt to one of four decision phases by applying a compact set of complementary cues that specify temporal location, focal action, referent artifact and intended outcome. An excerpt was coded as *Preparation* when it concerned activities that precede option generation – such as scoping, sourcing or configuring inputs; as *Evaluation* when it focused on comparing or assessing options – interpreting, triaging or benchmarking evidence; as *Decision-Making* when it centered on commitment and its justification – choosing, authorizing or documenting rationale; and as *Implementation* when it addressed execution, monitoring and learning after commitment – enacting, delegating, auditing or retrospecting. Linguistic and material indicators assisted these judgments (for example, verbs like collect/scope/ingest for Preparation; interpret/compare/triage for Evaluation; choose/authorize/justify for Decision-Making; and enact/delegate/monitor/audit/retrospect

Figure 1 Gioia's method



Source: Authors' own work

for Implementation), as did artifact references (data sets/sensors/pipelines; dashboards/reports/scorecards; decision memos/policy toggles/approvals; standard operating procedures/runbooks/telemetry/after-action notes). Outcome orientation provided a final check: Preparation expands the option space, Evaluation narrows it, Decision-Making fixes it and Implementation operationalizes it. Operationally, we use concise definitions throughout: Preparation opens the option space by scoping, sourcing and configuring inputs; Evaluation narrows the option space by transforming inputs into comparable evidence; Decision-Making allocates authority and commits to a course of action with an articulated rationale; Implementation enacts the choice, monitors consequences and feeds organizational learning. Two coders applied these rules independently during second-cycle coding; disagreements were resolved by consensus. A targeted subset was double-coded specifically for phase attribution; agreement statistics and adjudication procedures are reported in Appendix 1.

Orthogonally to phases, each excerpt also received a DIKW label using clear inclusion/exclusion criteria. *Data* denotes raw signals and their governance – collection, curation, provenance, quality and instrumentation – typically marked by verbs such as collect, clean, label, log, validate and by artifacts such as sensors, APIs, schemas and pipelines; interpretive comparisons or policy rationales are excluded at this level. *Information* captures structured representations for comparison – summaries, metrics, rankings and visualizations – signaled by terms such as summarize, benchmark, rank, triage and visualize, as well as by dashboards, reports and scorecards; final choices or execution steps are excluded. *Knowledge* refers to integrated reasons, heuristics and justificatory logics that support commitment – verbs include infer, justify, reconcile and decide; artifacts include decision memos, playbooks and design rules – while raw ingestion and runbook execution are excluded. *Wisdom* addresses value alignment, nondelegable judgment, accountability, governance of delegation and after-action learning – verbs include evaluate ethically, steward, foresee, govern, learn; artifacts include governance policies, escalation matrices and postmortems – whereas key performance indicators tracking without normative reflection is excluded. We treat DIKW as an analytic lens rather than a strict ladder. To avoid "level leakage," each Results subsection foregrounds its designated level, while any necessary cross-level recursions (e.g. wisdom→data feedback) are explicitly signposted as DIKW Bridges. A Phase × DIKW cross-tabulation of coded segments (descriptive counts) is provided in Appendix 2, and a double-coded subset was used to assess agreement on DIKW labels.

We followed a staged Gioia procedure that makes the analyst's moves and artifacts fully transparent (first-order concepts, second-order themes, and aggregate dimensions), consistent with recent exemplars (Troise *et al.*, 2023). Phase 1 – immersion and open coding (first-order concepts): The team read all transcripts iteratively to stay close to informants' language and generate *in vivo*, first-order codes. To enhance dependability at the descriptive stage, we implemented intercoder reliability (IRR) checks on first-order coding. After jointly developing and piloting a codebook on a small set of transcripts, two coders completed a calibration exercise and then independently double-coded 30 interviews (24.6% of the corpus; 30/122) selected to cover various industries, roles, and levels of AI use intensity. IRR was computed on the presence/absence of first-order codes at the excerpt level using Cohen's kappa in NVivo 14 (QSR International). On the initial pass, the agreement was $\kappa = 0.66$, with an overall agreement of 82% across 412 double-coded excerpts. Following clarification of code definitions and boundary rules, the agreement improved to $\kappa = 0.79$, with an overall agreement of 88%. Discrepancies were reviewed in consensus meetings; unresolved cases (11 excerpts; 2.7% of the double-coded set) were adjudicated by a third researcher. We then applied the updated codebook to the entire corpus (comprising a total of 1,262 coded segments). A later stability spot-check on 12 additional interviews (9.8%) yielded $\kappa = 0.81$ and an overall agreement of 90%.

Phase 2 – second-order theorizing (theme formation): we examined patterned relationships among first-order codes to develop second-order themes that explain how autonomy is shaped in AI-mediated decision work. At this stage, we also assigned the Phase labels described and the DIKW labels to each excerpt, applying those rules independently of thematic content so that "what is said" (content) remains analytically distinct from "where it sits" (phase/level). A subset was double-coded for both labels; agreement and adjudication details are reported in Appendix 1. Here, we used established knowledge-management sensitizing concepts (e.g. internal vs external sourcing; exploration vs exploitation, knowledge hiding and knowledge transfer) strictly as lenses – not templates – to avoid forcing the data.

Phase 3 – aggregate dimensions and recursive structure: convergent themes were integrated into four aggregate dimensions that cut across cases and map to our decision phases (frame, evaluate, commit, enact) and DIKW layers (data, information, knowledge, wisdom). Labels are standardized across figures, tables and text to ensure terminological consistency.

Phase 4 – saturation, validation and audit trail: we proceeded until theoretical saturation and then validated the structure via two expert focus groups (academics; senior managers) used in a member-like fashion to probe clarity, boundary conditions and applicability; all steps were logged in a versioned audit trail (decision logs, memo archive, codebook versions).

For second-order theme development and aggregate dimensions, where interpretive judgment is central, we prioritized negotiated agreement and theory-building rigor over statistics. Three procedures supported trustworthiness: (1) constant comparison across cases and decision phases (frame, evaluate, commit, enact) and knowledge levels (data, information, knowledge, wisdom); (2) negative-case analysis and memoing to surface and bound theme exceptions; and (3) a documented audit trail (decision logs, memo archive, codebook versions) reviewed in peer debriefs. Finally, two expert focus groups (comprising five academics and five senior managers) served as member-like validation to test the clarity, boundary conditions and applicability of the emergent structure. All analyses were conducted in NVivo 14 with versioned project files.

Two focus groups – one with academic experts and one with senior managers – validated and refined emergent themes, ensuring both theoretical and practical relevance (Barbour, 2007). Focus group participants included five academic scholars – with domain expertise in human autonomy, decision-making and AI in organizational contexts – and five senior managers from industries with intensive AI adoption (e.g. finance, health-care analytics, e-commerce and manufacturing operations), where decision processes rely heavily on predictive, prescriptive, generative or perception-based AI systems. Semistructured focus groups explored core themes, with academics focusing on theoretical aspects and practitioners on practical applications. Particular attention was paid to areas of ambiguity, thematic saturation and applicability across different decision-making scenarios. The academic group focused on theoretical soundness and boundary conditions, while the practitioner group emphasized operational resonance and contextual fit. The sessions were audio-recorded and transcribed *verbatim*. Focus group transcripts were analyzed using the 18 themes, with feedback categorized as convergence, divergence or amplification. Focus group feedback was coded deductively using this framework, while also allowing for inductive coding where novel insights or critiques emerged. Instances of thematic convergence (where participant feedback supported existing interpretations), divergence (where disagreements or blind spots were revealed) and amplification (where insights elaborated or expanded on original findings) were systematically catalogued.

Triangulation enhanced credibility, transferability and framework robustness, leading to the development of a recursive model – the HAIL – that explains how autonomy evolves across decision-making phases and knowledge levels. This approach allowed us to move from rich, context-specific accounts to a theoretically grounded process model that addresses our central research question (Denzin and Lincoln, 2018; Gioia *et al.*, 2013).

4. Results

The sample was balanced by age (mainly 30–50 years), gender (50% male, 50% female) and professional experience (most with 10–29 years), as shown in Table 2. Participants were mainly top managers and executives from diverse sectors, including technology, services and the public sector. Most participants had advanced AI experience (87.7%), with none at the beginner level (see Table 3 for focus group demographics).

Participants' experiences ensure robust insights into autonomy in AI-assisted decisions. Thematic analysis shows autonomy is enacted differently across four phases: preparation, evaluation, decision and implementation – each aligned with the DIKW model (Ackoff, 1989; Rowley, 2007). Analysis produced 1,262 statements, refined into 69 codes and 17 latent themes, then grouped into four aggregate dimensions, see Figure 2 for a synthesis and Supplementary Material 3 for details. This structure highlights epistemic tensions and managerial strategies in AI-mediated knowledge work (Gioia *et al.*, 2013).

4.1 Decision preparation phase

The preparation phase sets the epistemic and strategic scope for decisions, with AI tools shaping managerial attention and salience at the data layer. Autonomy here is not a given – it

Table 2 Respondents demographic data

Description	*No*	*%*
Age		
51–60	2	1.6
41–50	57	46.7
30–40	63	51.6
Total	*122*	
Gender		
Men	61	50.0
Women	61	50.0
Total	*122*	
Years of experience		
+29	1	0.8
21–29	26	21.3
16–20	52	42.6
10–15	43	35.2
Total	*122*	
Job position		
Executive	37	30.3
Top manager	85	69.7
Total	*122*	
Industry		
Financial services	4	3.3
Industrial sector	4	3.3
Logistics	3	2.5
Manufacturing	3	2.5
Nonprofit organization	5	4.1
Other	43	35,2
Public sector	11	9.0
Retail	5	4.1
Services/consulting	21	17.2
Technology	23	18.9
Total	*122*	
AI experience		
Advanced	107	87.7
Intermediate	15	12.3
Total	*122*	

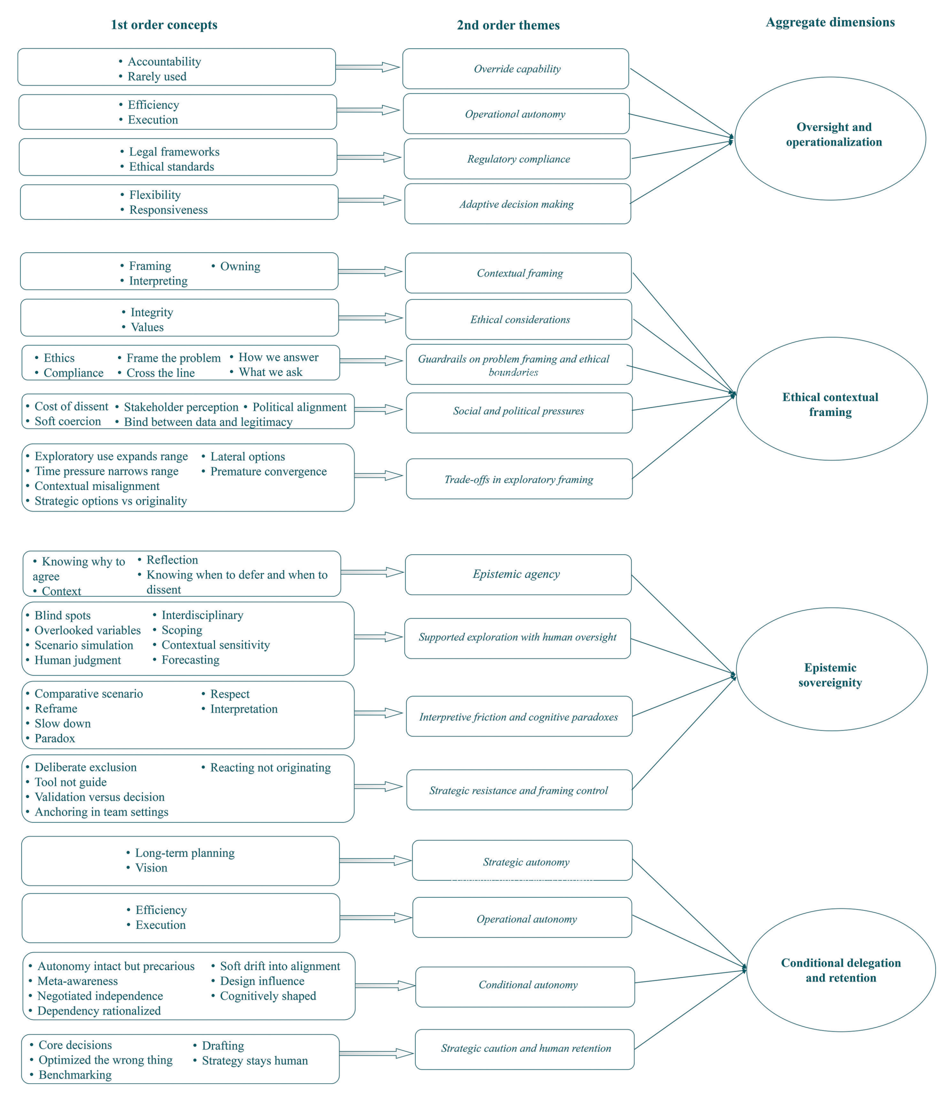
Source(s): Authors' own work

Table 3 Focus groups demographic data

Focus group	*Participant code*	*Role type*	*Sector/discipline*	*Years of experience*	*Age*
Academic	A1	Professor	Organizational behavior	15	48
Academic	A2	Associate professor	Technology and innovation management	12	44
Academic	A3	Assistant professor	Ethics and digital technology	7	38
Academic	A4	Full professor	Strategy and AI	20	54
Academic	A5	Associate professor	Human–machine interaction	10	42
Practitioner	P1	Senior manager	Telecommunications	18	50
Practitioner	P2	Executive	IT and cybersecurity	20	52
Practitioner	P3	Area director	FMCG	17	48
Practitioner	P4	VP operations	Manufacturing	22	55
Practitioner	P5	CFO	Financial services	19	51

Source(s): Authors' own work

Figure 2 Codes, themes and aggregate dimensions



Source: Authors' own work

is actively constructed through data management and governance practices that govern how data is sourced, prepared and integrated. This phase corresponds to the data layer of the DIKW hierarchy – where raw signals are collected, selected and validated for downstream use. Moreover, this is where foundational decisions about internal versus external data sources are made, shaping both the integrity and direction of AI-assisted processes. As one manager noted, "We've programmed it to stop if a human presses the override. That ability is non-negotiable" [028–02; Preparation/Data]. Data-governance–enabled autonomy emerges from custodial practices such as traceability, reversibility and the auditability of data inputs.

Crucially, design choices in this phase embed both exploration and exploitation logics. While AI often encourages pattern recognition and efficiency (exploitation), managers must preserve space for broader input exploration (e.g. refreshing sources, testing counterfactual inputs and walkthroughs of source configurations) to ensure that AI does not overly narrow the range of actionable signals. "We adapt our planning weekly. AI is a dynamic tool, not a static map" [012–02; Preparation/Data], noted one respondent, highlighting how flexibility is maintained through continual iteration and data (rather than interpretive) reconfiguration.

Data governance also operates as a filtering mechanism, guided by compliance and institutional legitimacy. Regulatory and stakeholder requirements determine the inclusion of data sources and the permissible boundaries of model use. “AI can’t be a black box when the law’s involved. Full auditability is key” [029–04; Preparation/Data], one participant emphasized. This signals how data coupling and decoupling processes are enacted: legally sensitive data is coupled to AI pipelines for transparency, while potentially bias-inducing sources are decoupled to safeguard autonomy and accountability upstream.

At the organizational level, data withholding and selective disclosure were also noted. Some participants described protective practices that limited what AI systems could access or infer. “We don’t feed it everything – we segment certain knowledge areas” [017–03; Preparation/Data], explained one senior executive, underscoring how autonomy is protected through strategic schema scoping and permissioning. Participants also discussed data handover protocols – especially in collaborative contexts involving multiple departments or partners. AI-based inputs are only admitted for downstream use when validated across functions. “It executes well, but we audit the process constantly. Execution isn’t decision-making” [021–05; Preparation/Data]. Transferability at this stage is procedural and technical; interpretive and normative filtering occurs downstream.

Autonomy threats in this phase – such as data set drift or overreliance on precoded relevance structures – are countered through data-level calibration and alignment. “Autonomy means staying loyal to our mission—even when the numbers suggest something else” [015–01; Preparation/Data], one respondent reflected. Here, data governance is both infrastructural and policy-oriented: it anchors data use in institutional memory and approved practices. Calibration techniques – like AI failure simulations at the input stage, red-teaming of sources and input reframing – function as boundary-testing exercises that maintain control over what enters the pipeline. As one manager aptly put it, “We simulate scenarios we’d never think of—but we interpret them, not the AI” [030–06; Preparation/Data].

In sum, the preparation phase is not merely a technical setup – it is a site of epistemic choreography where data management and governance processes, design safeguards and human discretion intersect. Autonomy here is realized not through isolation from AI, but through dynamic control over how data is sourced, curated and instrumented. This phase is constitutive of decision authority, as organizations use data-governance principles to ensure AI augments, rather than undermines, situated managerial judgment.

4.2 Decision evaluation phase

In the evaluation phase, autonomy is tested not by the availability of AI-generated information, but by the capacity to situate and scrutinize that information meaningfully. This phase corresponds to the Information level of the DIKW hierarchy, where raw data is organized and structured – but its relevance and legitimacy must still be interpreted through human judgment. AI provides informational breadth, but it is humans who define what is actionable. As one manager put it: “AI may offer scenarios, but we frame them. The context is ours to shape” [005-06; Evaluation/Information].

Managers act as interpretive curators, embedding AI outputs into broader socio-organizational frames. This aligns with the theme contextual framing (119 quotes; 7.94%) and reflects how autonomy is exerted through the informational decoupling of machine-generated suggestions from tacit assumptions or strategic commitments that are misaligned. In this sense, information management becomes a framing infrastructure – a repository of shared norms, interpretive routines and contextual templates that enable meaning to be co-constructed.

Crucially, decision evaluation often hinges on access to internal vs external information sources. Respondents described how internal databases, expert panels and prior cases are routinely juxtaposed with AI-generated options. “We don’t just look at what the AI says—we compare it with previous outcomes and expert judgment” [019-03; Evaluation/Information]. This interplay reflects information triangulation, an essential information-management practice that mitigates the risk of informational tunnel vision.

Guardrails and boundary conditions (149 quotes; 9.95%) play a foundational role in this phase, acting as informational filters that preclude classes of outputs from moving forward in the pipeline. “We set boundaries—some questions AI just doesn’t get to ask” [015-02; Evaluation/Information]. Operationally, these appear as explainability thresholds, relevance criteria and scope constraints encoded as information-level rules; broader ethical stewardship is treated in §4.4. As one respondent explained, “The ethics board reviews how AI is framing problems. It’s not just about solutions” [042-07; Evaluation/Information] – here, the review focuses on problem framing and representation, i.e. how information is structured.

Evaluation is also shaped by social and political pressures (60 quotes; 4.01%), which define the legitimacy of informational flows. These dynamics create tension between exploratory information use – seeking novel insights – and exploitative filtering – narrowing focus for operational feasibility. “Politics shapes what we allow AI to do—it’s not just a tech issue” [037-03; Evaluation/Information], one manager noted, revealing how institutional logics influence which AI-generated options are accepted, deprioritized or dropped at the information stage.

The trade-offs involved in exploratory framing were evident across interviews. Managers described how AI can flood them with options, many of which are technically feasible but contextually unviable. “AI gives breadth, but not always depth. We trade off novelty for feasibility” [020-04; Evaluation/Information]. This balancing act – captured in 73 quotes (4.87%) – represents a classic information-management dilemma: how to maintain informational generativity without sacrificing focus or actionability. Often, feasibility thresholds are encoded as information filters to pre-screen outputs – an instance of human-led structuring at the Information level.

Anticipatory validity checks were emphasized as preconditions for moving information downstream. Managers routinely considered whether AI-generated recommendations would be defensible and explainable, translating these concerns into information-level criteria (e.g. clarity of rationale, traceable indicators). “If I can’t explain the ethical logic of a choice the AI assists, I won’t go with it” [027-06; Evaluation/Information] – in practice, this prompts explainability requirements and relevance tests before any commitment (see §4.3).

Autonomy-preserving practices in this phase include why-not logs, override-justification templates and contextual insight maps – all of which assist interpretive traceability at the Information level. These tools allow human agents to document how they reframed or resisted AI suggestions, turning sensemaking into a deliberative audit trail. As one executive put it: “Ethics has to anchor AI use, or we’re just optimizing efficiency without accountability” [010-03; Evaluation/Information] – a reminder that information structuring is guided by clearly articulated criteria, with broader stewardship detailed in §4.4.

In summary, the evaluation phase marks a shift from extraction to interpretation. It is where information becomes meaningful – not through algorithmic structure alone, but via human acts of framing, filtering, resisting and justifying at the information level. Autonomy here concerns semantic control: the ability to shape informational representations so that what proceeds downstream is decision-relevant. Information-management systems thus function as strategic and epistemic infrastructures – transforming AI’s informational outputs into actionable comparisons while preserving human judgment.

4.3 Decision-making phase

The decision-making phase marks the transition from interpreted information to committed action – where deliberations culminate in a formal choice. This phase corresponds to the knowledge level of the DIKW hierarchy: data has been structured, information contextualized and now insight must be synthesized into justified courses of action. While AI assists this process through ranking, nudging and predictive analytics, autonomy hinges on the ability to navigate, repurpose or resist these cues, ensuring that decisions remain contextually valid and defensible.

A central theme emerging here is epistemic sovereignty (101 quotes; 6.74%) – the capacity to interrogate, not merely accept or reject, AI recommendations. “Autonomy, in my view, is less about rejecting AI output and more about understanding why you agree. That’s epistemic power” [003–01; Decision/Knowledge]. Decision authority, in this context, becomes a knowledge performance grounded in interpretive ownership. Structured reflection rituals – such as “AI rationale unpacking” sessions – were used to assess alignment between system logic and organizational judgment. “It’s essential to know not just what AI recommends but why – that’s where human autonomy lives” [017-04; Decision/Knowledge].

This links closely to the theme of Supported Exploration with Human Oversight (207 quotes; 13.82%). AI broadens the solution space, offering permutations that may be novel or counterintuitive. Yet human judgment acts as the selective filter – choosing what to retain, pursue or discard. “AI expands our option set, but we’re the ones choosing which doors to open” [009-01; Decision/Knowledge]. Teams often engage in comparative simulations or “dry-run decisions” to explore consequences across options, thereby coupling exploratory potential with justificatory selectivity. This cognitive coupling process keeps machine logic grounded in organizational sense-making at the Knowledge level.

Friction between AI and human reasoning frequently emerged as a productive tension. The theme Interpretive Friction and Cognitive Paradoxes (141 quotes; 9.41%) captures how clashes between human intuition and AI outputs spur critical reflection. “Sometimes AI shows you a contradiction you hadn’t seen—makes you think harder” [026-03; Decision/Knowledge]. Far from being barriers, these moments catalyze deeper insight, serving as autonomy checkpoints where meaning is re-evaluated. “It’s when the logic doesn’t sit right that we get our best insights—that friction matters” [038-02; Decision/Knowledge]. Such paradoxes trigger knowledge recursivity – an iterative integration dynamic in which established beliefs are confronted and reworked in light of emergent cues.

Some managers reported strategic resistance and framing control (53 quotes; 3.54%) – the deliberate decoupling of decisions from AI-generated frames. This was not necessarily a rejection of accuracy, but of relevance. “We limit AI’s framing ability. It assists, but doesn’t define, the issue” [011-02; Decision/Knowledge]. Others described how AI was excluded from final deliberations to safeguard epistemic legitimacy and retain strategic coupling with institutional memory. “Sometimes we ignore its suggestions—not because they’re wrong, but because they’re misaligned” [035-05; Decision/Knowledge]. These practices can involve selective exclusion of AI inputs deemed inappropriate for the context or culture at the point of commitment.

Knowledge practices were critical in navigating these dynamics. Calibration tools such as autonomy rehearsal rituals, design audits and justification templates enabled decision-makers to assess when to trust, when to adapt and when to override. These tools assist knowledge transfer across levels and actors, ensuring that final choices reflect not just system outputs but collective insight and articulated rationale. “It’s not just about taking the best option—it’s about owning the path we choose” [024–01; Decision/Knowledge].

In summary, the decision phase is where knowledge becomes consequential – transforming potential insight into enacted choice. Autonomy here is a situated, performative competence – enacted through synthesis, justified dissent and strategic restraint – and it is inseparable from responsibility, understood as the capacity to explain one's decision and accept its consequences. AI systems may act as coactors, but final authority and accountability rest with human decision-makers who can narrate both the why (reason) and the what for (purpose) of the choice. Through interpretive sovereignty, knowledge-level oversight and clear justification practices, the decision phase affirms that autonomy is not about choosing alone, but about choosing wisely and answerably under technological augmentation.

4.4 Decision implementation phase

The implementation phase – corresponding to the Wisdom tier of the DIKW model – shifts the focus from deciding to enacting, embedding decisions into real-world consequences. Here, autonomy is no longer about access or interpretation but about ethical stewardship, foresight and narrative accountability. Managers are not just executors of decisions, but custodians of organizational intent, responsible for ensuring that AI-enabled actions remain traceable, defensible and value-aligned.

A key mechanism enabling autonomy in this phase is Conditional Autonomy (72 quotes; 4.81%). Rather than blanket automation, organizations define delegation architectures – circumscribed zones where AI can act independently, framed by reversibility and situational awareness. As one executive stated, "Autonomy here is conditional—we grant it when the situation's predictable" [043-05; Implementation/Wisdom]. Another echoed: "AI takes the wheel sometimes, but only within a sandbox we control" [008-06; Implementation/Wisdom]. These forms of governed independence reflect a wisdom-oriented governance logic that embeds execution rules and exception protocols, ensuring that AI operates within human-defined epistemic and ethical boundaries.

Beyond the operational layer, managers insisted on maintaining Strategic Autonomy (48 quotes; 3.20%). This involves retaining control over vision, values and trajectory of implementation – elements that AI cannot anticipate or ethically anchor. "Autonomy is being able to see past the numbers to the long-term vision" [022-01; Implementation/Wisdom], explained one executive. Another clarified, "Strategic autonomy is strongest when the decision structure remains with people, not algorithms" [012-08; Implementation/Wisdom]. In stewardship terms, this entails governance for foresight: holding onto value-driven logics and institutional memory when guiding downstream actions.

A third theme – Operational Autonomy (54 quotes; 3.60%) – describes the delegation of routine tasks to AI, with human oversight retained at key checkpoints. "We delegate operations to AI but not the logic behind them" [018-03; Implementation/Wisdom], emphasized one manager. "It executes well, but we audit the process constantly. Execution isn't decision-making" [021-05; Implementation/Wisdom]. This reflects boundary maintenance at the wisdom level: a clear separation between automated doing and responsible knowing, ensuring the *locus* of final authority remains human.

Still, many respondents expressed concern over automation creep in high-stakes domains. The theme Strategic Caution and Human Retention (115 quotes; 7.68%) highlights the deliberate refusal to automate core strategic or ethically sensitive decisions. "We've kept key decisions human-led. AI advises, we decide" [036-04; Implementation/Wisdom]. Another affirmed: "There's a line we don't cross—AI doesn't replace core strategic judgment" [014-01; Implementation/Wisdom]. This reflects governance safeguards in which ethical irreducibility and institutional checks prevent value drift or deferral of responsibility to machines.

Crucially, governance and learning systems in this phase become vehicles of institutionalized wisdom – not simply codifying best practices, but preserving non-delegable roles. Organizations deploy "non-delegable core maps," boundary markers, and post hoc audits to ensure that AI's role remains visible, bounded and correctable. Visibility is further managed through decision storytelling, with many respondents emphasizing the communicative and legitimizing power of narrating AI's role: "We tell the story of how we used AI—because that's part of the decision's credibility" [025-07; Implementation/Wisdom].

In this way, autonomy becomes a narrative and governance outcome rather than a mere moment of choice. Decision implementation is the arena where human authorship is defended and made durable – through foresight, boundary design and institutional stewardship that safeguard judgment. Here, responsibility means the capacity to explain the why of a decision and justify the what for (its purpose and intended outcomes) – and to accept the consequences of both. Accordingly, wisdom is not only the selective application of knowledge, but also knowing when not to delegate, how to narrate delegation, and why restraint can be a moral act in AI-augmented execution.

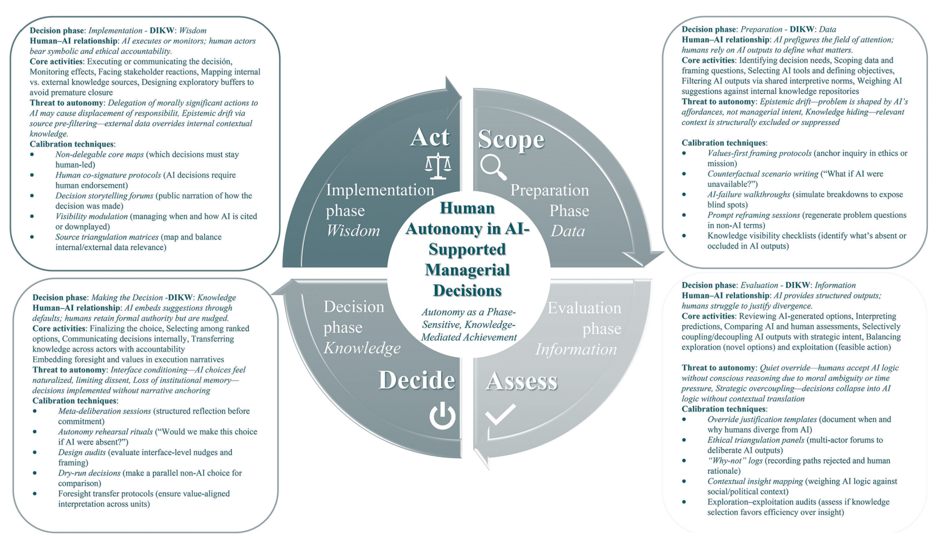
5. Discussion

Drawing on our findings, we propose the Human-AI autonomy Loop framework that redefines autonomy not as a static right or trait, but as an evolving capacity to shape, resist, and reframe knowledge under conditions of algorithmic coagency (Figure 3). The HAIL framework is circular in a concrete way: what is learned at the wisdom level informs the next round of data work. In practice, lessons learned guide what to collect, how to label it, and which traces to keep. This closes the loop between wisdom and data and aligns with Jennex's (2017) revised knowledge pyramid, where higher-level understanding shapes data capture and metadata (Ackoff, 1989; Frické, 2009; Weinberger, 2010). Appendix 3 maps four AI system classes – Predictive/Descriptive analytics, Prescriptive/Optimization and recommender systems, Generative AI assistants and Perception & Extraction tools – onto DIKW layers and the four decision phases (Frame, Evaluate, Commit, Enact), indicating typical contributions, autonomy risks and safeguards.

The HAIL framwork advances both DIKW and foundational KM by:

- rejecting linearity in favor of recursive coupling between phases and layers;
- dislocating the locus of interpretation from a purely human center toward distributed, contested knowledge construction in human–AI systems – extending knowledge-creation

Figure 3 Human–AI autonomy loop (HAIL) framework



Source: Authors' own work

perspectives (Nonaka and Takeuchi, 1995; Heisig, 2009) and aligning with socio-technical accounts (Introna, 2015); and

- redefining autonomy from a trait-like property to a phase-contingent practice of epistemic authorship, shaped by trust configurations, interface affordances and narrative legitimacy (Kellogg *et al.*, 2020).

This contrasts with trait-centric autonomy and accuracy-centric trust accounts by specifying how autonomy and trust vary across decision phases and DIKW levels. In addition, the HAIL framework clarifies that wisdom feeds back into data: organizations use what they learn to set measurement plans, adjust schemas and refine labels for the next cycle.

Decision preparation (Frame): framing sovereignty vs sourcing efficiency. Traditional KM emphasizes knowledge sourcing (internal vs external) to widen inputs (Grant, 1996; Alavi and Leidner, 2001). Our evidence shows that AI preselects and prestructures relevance – rebalancing toward external feeds and encoded heuristics – and risks epistemic drift when problem frames are silently inscribed in pipelines. We therefore introduce framing sovereignty: the human capacity to define relevance, granularity and scope before algorithmic defaults harden. This extends sourcing logics by shifting emphasis from volume and breadth to early-phase authorship over what enters the decision space. Practically, framing choices are written into data work: teams decide what to log, which sources to exclude and what context to preserve (e.g. timestamps, provenance).

Decision evaluation (Evaluate): Interpretive autonomy under algorithmic filtering and hiding. Evaluation is often cast as knowledge filtering – prioritizing "actionable" information (Tiwana, 2002). Our findings show that AI tools foreground quantifiable comparators while systemic/algorithmic knowledge hiding can occur when interfaces or model logics obscure provenance, caveats or alternatives (Connelly *et al.*, 2012; Anand *et al.*, 2022). This extends knowledge-hiding research from interpersonal intent to pipeline- and interface-level effects with direct implications for interpretive autonomy. We specify ethical triangulation – combining algorithmic outputs with narrative sensemaking and peer consultation – to surface what AI omits. Thus, transparency is necessary but insufficient; interpretive capacity must be institutionally protected. In practice, evaluative judgments become structured data: accept/reject reasons, audit flags and counterexamples are recorded as labels, notes and training artifacts for the next evaluation round.

Decision-making (Commit): Encoding rationale into structure. Classic work highlights exploration–exploitation tensions (March, 1991). In practice, evaluation pipelines tilt commitment toward exploitation, risking symbolic autonomy – retained in principle, constrained in practice. To counter this, the HAIL framwork defines a legitimacy bandwidth: organizational permission to choose a path that may be locally "less optimal" yet responsibly justified by purpose. Concretely, each commitment must include a documented why (causal rationale) and a what for (intended outcomes, stakeholder value, ethical and strategic aims). Encoding both into decision records and governance artifacts anchors accountability for consequences and safeguards future-oriented choices. This repoliticizes knowledge use (Alvesson and Kärreman, 2001) and complements "algorithm appreciation" (Logg *et al.*, 2019) by explaining why surface trust in accuracy coexists with fragile, layered trust dependent on transparency, fairness, and interpretability (Bader and Kaiser, 2019; Brink *et al.*, 2024). Our contribution is to tie these legitimacy levers to concrete safeguards (e.g., decoupling and narrative justification) that preserve interpretive sovereignty at the moment of commitment. Operationally, commitments are logged with short, structured rationales and policy toggles; these entries shape future prompts, constraints, and optimization targets.

Decision implementation (Enact): Governed delegation and knowledge coupling/decoupling. Implementation is often framed as knowledge transfer and best-practice diffusion (Argote, 2011). We extend this by theorizing governed delegation: human actors retain narrative and strategic authorship while AI executes operational logic. We further distinguish knowledge coupling (tight data-to-action linkages) from decoupling (strategic distancing in ambiguous or high-stakes contexts) to institutionalize reversibility and protect a nondelegable core of ethical authorship (Prunkl, 2024). This contrasts with techno-rationalist diffusion views by centering discretion, reversibility, and moral agency as first-class implementation concerns. In DIKW terms, Enact corresponds to the Wisdom level: situated judgment applied in context. In human–AI decision-making, that wisdom does not stop at execution; it shapes what the system and organization learn next. To make the loop explicit, implementation produces two kinds of record:

1. raw traces (telemetry, timestamps, human overrides, exception paths); and
2. curated lessons (after-action notes that state what worked, what failed and why).

We then turn wisdom into next-cycle data through a simple routine aligned with DIKW: capture (collect traces and notes), curate (distill lessons with clear rationales), translate (convert rationales into structured fields – labels, flags, counterexamples) and instrument (update logging, labeling and required context so the next Frame/Data phase starts with improved inputs). For example, if a team overrides a recommender due to fairness concerns, that rationale becomes a structured flag in the log, a training example in the data set and a guardrail in the policy table – so the next Data layer already encodes today's Wisdom. In short, implementation does not only consume data; it writes tomorrow's data from today's wisdom, connecting Wisdom (Enact) back to Data (Frame) in a way that is simple, auditable and consistent with revised DIKW logic (Jennex, 2017).

Taken together, the HAIL framwork reconceptualizes autonomy as embedded within KM, challenging paradigms that privilege volume, speed and codification. Trust is reframed from an accuracy proxy to a practice of structuring dissent, surfacing invisible knowledge and protecting reversibility. Knowledge becomes a negotiated accomplishment between human interpretive traditions and algorithmic representations. In doing so, we integrate and extend DIKW (by operationalizing recursion), knowledge creation/transfer (by relocating interpretive labor across human–AI assemblages), exploration–exploitation (by specifying legitimacy bandwidth), and knowledge hiding (by theorizing algorithmic forms and practical safeguards). The contribution is both theoretical and practical: the HAIL framwork offers a mid-range, empirically grounded explanation of when and how AI reshapes autonomy across phases, and delineates governance levers – framing sovereignty, ethical triangulation, decoupling and governed delegation – to sustain managerial authorship in AI-mediated decisions. In brief, the HAIL framework is featured as a loop because what organizations learn (wisdom) is used to decide what they measure and store next (data), making the cycle explicit and auditable (Jennex, 2017).

6. Implications

6.1 Implications for theory

This study advances theory in three ways:

1. KM and AI scholarship on decision-making, in dialogue with Leoni *et al.* (2024) and Rezaei *et al.* (2024);
2. autonomy theory in socio-technical systems; and
3. decision-making foundations under technological augmentation.

Advancing AI–KM decision research: from linear pipelines to socio-epistemic loops. We challenge the dominant view that treats decision-making as a linear flow of data aggregation, knowledge sharing and application (e.g. Leoni *et al.*, 2024; Nguyen and Malik, 2022). We show that managerial autonomy is not merely a function of informational access or algorithmic explainability (Rezaei *et al.*, 2024), but a recursive, knowledge-centric capacity to author, resist and reshape meaning across decision phases. Unlike Metaxiotis *et al.* (2003), who emphasize efficiency and alignment, we highlight how AI can constrain epistemic sovereignty by standardizing relevance and compressing ambiguity (Catena *et al.*, 2025). Where prior studies present AI as neutral infrastructure, our findings indicate that it actively structures the who, what and how of knowledge engagement – risking displacement of human authorship unless countervailing practices are enacted.

We introduce framing sovereignty to denote the human capacity to define not only the relevance and structure of decision problems before AI scripts them, but also their intended ends. Accordingly, we reconceive DIKW as a reflexive loop, where each level is an arena of negotiation between algorithmic framing and human interpretive agency about both the why (causal rationale) and the what for (purpose and intended outcomes). We operationalize responsibility as the capacity to explain the why and justify the what, and to accept the consequences of both. Embedding these practices into decision governance introduces moral agency into AI–KM research, shifting theory from "pipelines of codified knowledge" to contested terrains of performative and moral agency, where trust, resistance and narrative alignment determine what counts as valid knowledge and which purposes are legitimately pursued.

Recasting autonomy as situated, performative and epistemically distributed. Existing autonomy research often conceives autonomy as a static trait (Christman, 2009), a psychological need (Ryan and Deci, 2017) or a task-design feature (Binns and Veale, 2021). Our framework advances a performative and epistemically distributed understanding of autonomy – one enacted through micropractices of reframing, dissent, alignment or override across decision phases. This view draws on relational autonomy (Laitinen and Sahlgren, 2021; Prunkl, 2024), but it moves beyond this concept by emphasizing temporal–epistemic variation. Autonomy is not simply "relational," but phase-specific: anticipatory in preparation, interpretive in evaluation, strategic in commitment and ethical in implementation. We contribute three novel insights:

1. we link autonomy directly to epistemic labor – the situated work of making knowledge meaningful (what for) and/or contestable (why);
2. we theorize autonomy as variably enacted across decision phases, shaped by AI affordances and trust configurations; and
3. we introduce a new vocabulary – framing sovereignty, symbolic autonomy, governed delegation – to explain how autonomy is won, eroded or reasserted in AI-mediated settings.

This extends the temporal-relational accounts of Faraj *et al.* (2018) and Kaplan and Orlikowski (2013), offering a more granular theorization of autonomy's evolving nature in algorithmic contexts.

Reconstructing decision-making theory under AI augmentation. Classical decision theory assumes that more data yields better outcomes, especially with AI-enabled pattern recognition and predictive analytics (Metaxiotis *et al.*, 2003). Our findings show that autonomy and decision quality do not scale linearly with data abundance. Progressing up the DIKW hierarchy toward wisdom, decision quality hinges on phronesis – practical judgment that integrates moral purpose, prudence and stakeholder-oriented reasoning to orient choices toward the common good in the face of uncertainty. Left uninterrogated, AI outputs risk value-insensitive optimization and may degrade decision quality (Floridi, 2011; Bao *et al.*, 2023). Building on the turn to wisdom management, we treat wisdom as normatively loaded rather than epiphenomenal: a level where ethical discernment, accountability and legitimacy govern how knowledge is applied – and when it should not be used (Jakubik, 2023; Jakubik and Müürsepp, 2022). In this reconstruction, responsibility is not exhausted by the capacity to explain the decision's why (causal rationale); it centrally requires justifying the what for (purpose and intended outcomes) and accepting the consequences of that justification. Accordingly, governance should mandate explicit articulation of both rationale and purpose before, during and after algorithmic assistance, and empower refusal when purposes cannot be responsibly justified.

In this framing, autonomy becomes moral–epistemic stewardship. Decision-makers curate not only knowledge boundaries but also value boundaries, applying three guardrails before endorsing algorithmic recommendations:

1. a moral-purpose test (is the recommended end coherent with the organization's stated values and the common good?);
2. a dignity and harm test (does the action respect stakeholder dignity and minimize foreseeable harm?); and
3. an intertemporal proportionality test (are short-term gains outweighed by longer-term social or ecological costs?).

When these guardrails fail, phronetic overrides (reasoned refusals, redesigns or delays) take precedence over algorithmic optima.

Crucially, wisdom is actionable and measurable. Recent KM research operationalizes practical wisdom as a multidimensional construct – including moral purpose in decision-making, subject-matter expertise, self-reflection, external reflection, integrative thinking and "exceeding the bounds of rationality" (judicious use of intuition) – and demonstrates that higher practical wisdom suppresses knowledge sabotage/hiding and promotes knowledge sharing (Serenko, 2024). This evidences a behavioral pathway through which wisdom safeguards autonomy: by cultivating climates where voice, candor, and sharing are normatively expected, the organization counters behaviors (e.g. evasive hiding, playing dumb) that otherwise erode human control over AI-mediated decisions.

Accordingly, the HAIL framework contributes to decision theory by:

- recasting the decision-maker from rational calculator to curator of knowledge and value boundaries;
- showing that autonomy is shaped not only by AI inputs and phase-contingent access to diverse knowledge modalities but also by explicit ethical commitments at the wisdom level; and
- locating KM within a normative architecture – aligned with wisdom management – through which organizations govern agency, voice, and interpretive leeway (including when to override AI).

6.2 Implications for practice

The HAIL framework provides a new foundation for designing AI-integrated decision systems that retain meaningful human autonomy while leveraging the efficiencies of automation. By understanding autonomy as a phase-contingent and knowledge-mediated achievement, managers can take concrete steps to calibrate AI use across decision stages rather than applying uniform solutions. This requires organizations to actively structure autonomy into workflows, tools and accountability systems.

A first implication concerns the decision preparation phase – associated with the data level of DIKW – where autonomy is shaped by the ability to frame problems before AI preprocesses them. Here, organizations must protect what HAIL calls framing autonomy by introducing practices that separate initial human sensemaking from AI-driven salience detection. This can include structured human-only scoping workshops, the use of AI-silent zones for early ideation and cross-functional framing rituals that delay algorithmic input. The goal is to ensure that decision problems emerge from human relevance structures, not machine-encoded priors. Practically, this may require reconfiguring the workflow so that managers explicitly articulate strategic objectives before consulting AI dashboards or forecasts. These interventions prevent the premature closure of problem space – a core risk in AI-assisted preparation – thereby preserving human interpretive authorship at the outset.

A second implication arises in the decision evaluation and decision-making phases, where AI primarily operates at the information and knowledge levels. The HAIL framework shows that this is where interpretive autonomy is most at risk – not because of a lack of control, but because of subtle dependencies that form when AI advice appears statistically confident or institutionally legitimate. To manage this, organizations should embed interpretive friction protocols within evaluation routines. For instance, managers can be required to document divergent interpretations, justify AI overrides, or engage in structured disagreement before proceeding to final choices. One effective tactic is the implementation of "shadow interpretation teams" – parallel groups who evaluate the same AI outputs independently and report on potential biases, missing contexts, or ethical blind spots. During the choice phase, where the HAIL framework identifies "meta-conditioned autonomy," human actors should have tools to own their decisions narratively. This includes internal platforms to explain rationales for dissenting from AI recommendations, digital traceability tools for value-based justification and peer review boards for ethically sensitive decisions. These practices uphold autonomy not by rejecting AI, but by ensuring that human evaluators remain epistemically active and morally visible.

A third implication concerns the decision implementation phase, mapped to the wisdom tier of the DIKW model. Although AI systems may execute and monitor independently, the HAIL framework shows that autonomy does not end at choice; it is retrospectively constructed through how responsibility is attributed, narrated, and owned. Answering the why (causal rationale) is necessary but not sufficient; responsibility also requires justifying the what for (purpose and intended outcomes) and accepting the consequences of that justification. Firms should adopt conditional delegation: AI may act independently only when both the "why" and the "what for" are articulated ex ante, recorded, and validated within a defined legitimacy bandwidth. Where either element is unclear – or the stakes are ethical, strategic, or irreversible – human oversight is mandatory. Implement nondelegable core maps to demarcate decisions that require human validation and cosignature mechanisms, so that high-stakes AI-triggered executions receive human affirmation. Post-implementation, create narrative accountability structures (e.g. storytelling reviews, decision logs) that document and communicate rationale, purpose, and consequences, attributing credit and blame appropriately. These practices institutionalize wisdom-level autonomy: implementation is designed not as blind execution but as narratively governed responsibility, where decision-makers must explain the why, justify the what for, and own the outcomes.

Taken together, these implications urge a shift from technical integration to epistemic design: the deliberate structuring of workflows, roles, and protocols to sustain human autonomy in a looped, AI-assisted decision environment. Instead of relying on static governance models or post hoc audits, the HAIL framework encourages organizations to shape autonomy at every phase – as framing, interpreting, choosing, and owning. This is not simply about "keeping humans in the loop," but about redesigning the loop itself to ensure that AI enhances, rather than erodes, the human capacity to judge, justify, and lead.

6.3 Limitations and future research

While this study advances a novel framework for understanding how managerial autonomy is enacted in AI-assisted decision-making processes, it is important to acknowledge its conceptual and methodological limitations. Rather than framing these as deficiencies to be corrected, we position them as productive tensions that open fertile avenues for future research.

First, the reliance on self-reported, retrospective accounts – whether through open-ended surveys or focus group dialogue – limits our access to real-time, embodied experiences of decision-making. Although our interpretivist design privileges subjective meaning-making, it cannot capture the fine-grained temporal dynamics of how autonomy shifts moment by moment within actual human–AI interactions. Future research could build on this by incorporating ethnographic observation, think-aloud protocols or even experimental simulations that capture how autonomy is negotiated in the flow of live decisions. Such approaches would allow researchers to explore the micromovements of deference, resistance and reassertion that remain under-theorized in our current model.

Second, the HAIL framework is based on a culturally and institutionally bounded sample: senior Italian professionals operating within large organizations. While this sample offered rich insights into how autonomy is exercised in a high-context, moderately hierarchical business culture, it may not generalize across other national, institutional or organizational settings. Autonomy may be framed differently in flatter Scandinavian firms, Silicon Valley startups or public-sector bureaucracies. Future research should therefore investigate how sociocultural variables – such as power distance, uncertainty avoidance or legal-institutional trust – mediate the enactment of human autonomy in algorithmically assisted decisions.

Third, we focus on managerial perspectives, largely ignoring the design side of AI systems. This asymmetry risks reproducing a user-centric view that treats AI as an exogenous actor, rather than a sociotechnical system coshaped by human choices, data politics and organizational priorities. A promising direction for future inquiry would be to "follow the AI" across its design, deployment and use contexts, thereby illuminating how autonomy is embedded – or foreclosed – through specific technical affordances, interface design or data training choices. Such research could bridge critical algorithm studies with decision sciences, addressing the growing call for socio-algorithmic accountability in managerial contexts.

Fourth, our phase-based structure – while analytically generative – risks imposing a false linearity on decision processes that are often recursive, overlapping and messily constructed. Although we explicitly theorized the HAIL as cyclical, the framework remains vulnerable to misinterpretation as a prescriptive sequence. Future work could complicate this structure by applying processual or practice-theoretic lenses (e.g. Feldman and Orlikowski, 2011) to show how autonomy is not only enacted in distinct phases, but also emergently co-constructed through iterations, interruptions and improvisations.

Fifth, we foreground autonomy but give less analytical space to power, emotion and identity. Our findings suggest that decisions involving AI are not only cognitive acts but also moral and political ones, entangled with professional self-conceptions and organizational legitimacy. Yet these dimensions are not fully theorized in the present study. Future research might adopt a more critical orientation to explore how perceived autonomy intersects with emotional labor, identity work or organizational politics. For example, does algorithmic decision assist pose a greater threat to certain professional identities than others? How do emotions like shame, pride or anxiety shape the uptake or rejection of AI advice? These questions are vital for understanding the lived experience of autonomy beyond abstract functionality.

Sixth, we call for deeper interdisciplinarity in future research. The phenomena we study – autonomy, AI and decision-making – span disciplinary borders but are often siloed in empirical and theoretical treatments. We encourage future scholars to draw from human–computer interaction, philosophy of mind, moral psychology and critical algorithm studies to construct richer, more nuanced accounts of how humans cocreate decisions with machines. In particular, the concept of "autonomy" requires ongoing philosophical interrogation: is it simply decision latitude, or does it also imply reflexivity, moral accountability or the ability to dissent? Each conceptualization carries normative and managerial implications that deserve further scrutiny.

Finally, we urge future work to confront a foundational paradox: as AI systems become more “intelligent,” human autonomy may paradoxically become more performative – a signaling act rather than a substantive mode of control. The more efficient and predictive these systems become, the more managers may feel compelled to defend their autonomy even as their actual influence wanes. This emergent condition of “symbolic autonomy” warrants serious investigation, particularly in high-stakes domains such as finance, health care and public policy, where decision legitimacy is closely tied to perceptions of human oversight.

In sum, rather than closing the autonomy debate with a definitive model, our study opens a new frontier for research – one that treats autonomy not as a static attribute, but as a socially constructed, emotionally charged and contextually fluid practice. We hope future scholars will take up this invitation, extending and challenging the HAIL framework across settings, systems and theoretical traditions.

7. Conclusion

This study reconceptualizes managerial autonomy in AI-assisted decision-making as a dynamic, phase-contingent process rooted in knowledge practices. Drawing on 122 interviews and expert focus groups, we introduce the HAIL framework that traces how autonomy is enacted, challenged and recalibrated across four decision phases – Preparation, Evaluation, Decision and Implementation – each mapped onto a corresponding layer of the DIKW hierarchy. Rather than viewing autonomy as a stable trait, we show it emerges through situated practices of framing, interpreting, justifying and executing decisions within AI-assisted knowledge environments. Our findings extend KM research by shifting focus from AI’s role in optimizing knowledge flows to its impact on epistemic agency. We argue that autonomy must be designed into AI-assisted KM systems – not assumed – through structures that preserve interpretive discretion, enable contestation and maintain narrative accountability. The HAIL framework reveals that sustaining autonomy is not about resisting AI but about embedding mechanisms for reflexive human engagement throughout the decision process. Ultimately, the future of knowledge-intensive decision-making will hinge not only on AI capabilities but on our ability to govern how knowledge and agency coevolve – ensuring that wisdom remains a distinctly human achievement within AI-rich organizational contexts.

Acknowledgements

The authors acknowledge the support of the Universitat Politècnica de València through a research grant and the hospitality of the University of Rome Tor Vergata during the visiting research period.

References

Ackoff, R.L. (1989), “From data to wisdom”, *Journal of Applied Systems Analysis*, Vol. 16, pp. 3-9.

Alavi, M. and Leidner, D.E. (2001), “Review: knowledge management and knowledge management systems: conceptual foundations and research issues”, *MIS Quarterly*, Vol. 25 No. 1, pp. 107-136.

Alvesson, M. and Kärreman, D. (2001), “Odd couple: making sense of the curious concept of knowledge management”, *Journal of Management Studies*, Vol. 38 No. 7, pp. 995-1018.

Anand, A., Offergelt, F. and Anand, P. (2022), “Knowledge hiding–a systematic review and research agenda”, *Journal of Knowledge Management*, Vol. 26 No. 6, pp. 1438-1457.

Argote, L. (2011), “Organizational learning research: past, present and future”, *Management Learning*, Vol. 42 No. 4, pp. 439-446.

Argote, L. and Ingram, P. (2000), “Knowledge transfer: a basis for competitive advantage in firms”, *Organizational Behavior and Human Decision Processes*, Vol. 82 No. 1, pp. 150-169.

Awad, E.M. and Ghaziri, H. (2004), *Knowledge Management*, Prentice Hall, Upper Saddle River, N.J.

Bader, V. and Kaiser, S. (2019), "Algorithmic decision-making? The user interface and its role for human involvement in decisions supported by artificial intelligence", *Organization*, Vol. 26 No. 5, pp. 655-672.

Bao, Y., Gong, W. and Yang, K. (2023), "A literature review of human–AI synergy in decision making: from the perspective of affordance actualization theory", *Systems*, Vol. 11 No. 9, p. 442.

Barbour, R. (2007), *Doing Focus Groups*, Sage Publications, London.

Berente, N., Gu, B., Recker, J. and Santhanam, R. (2021), "Managing artificial intelligence", *MIS Quarterly*, Vol. 45 No. 3, pp. 1433-1450.

Binns, R. and Veale, M. (2021), "Is that your final decision? On the role of explanations in automated decision-making", *AI and Law*, Vol. 29 No. 2, pp. 79-103.

Bootz, J.-P., Lievre, P. and Schenk, E. (2015), "Solicitation of experts in an undetermined environment: the case of a polar exploration", *Journal of Knowledge Management*, Vol. 19 No. 5, pp. 900-911.

Braun, T. (2025), "Liability for artificial intelligence reasoning technologies – a cognitive autonomy that does not help", *Corporate Governance: The International Journal of Business in Society*, doi: 10.1108/CG-09-2024-0471.

Brink, A., Benyayer, L.D. and Kupp, M. (2024), "Decision-making in organizations: should managers use AI?", *Journal of Business Strategy*, Vol. 45 No. 4, pp. 267-274.

Candrian, C. and Scherer, A. (2022), "Rise of the machines: delegating decisions to autonomous AI", *Computers in Human Behavior*, Vol. 134, p. 107308, doi: 10.1016/j.chb.2022.107308.

Catena, E., Tummolini, L. and Santucci, V.G. (2025), "Human autonomy with AI in the loop", *Philosophical Psychology*, pp. 1-28.

Chen, H., Yao, Y. and Zhou, H. (2021), "How does knowledge coupling affect exploratory and exploitative innovation? The chained mediation role of organisational memory and knowledge creation", *Technology Analysis & Strategic Management*, Vol. 33 No. 6, pp. 713-727.

Choo, C.W. (1996), "The knowing organization: how organizations use information to construct meaning, create knowledge and make decisions", *International Journal of Information Management*, Vol. 16 No. 5, pp. 329-340, doi: 10.1016/0268-4012(96)00020-5.

Christman, J. (2009), "Autonomy, recognition, and social dislocation", *Analyse & Kritik*, Vol. 31 No. 2, pp. 275-290, doi: 10.1515/auk-2009-0205.

Connelly, C.E., Zweig, D., Webster, J. and Trougakos, J.P. (2012), "Knowledge hiding in organizations", *Journal of Organizational Behavior*, Vol. 33 No. 1, pp. 64-88.

Cyert, R.M. and March, J.G. (1963), *A Behavioral Theory of the Firm*, Prentice-Hall, Englewood Cliffs, NJ.

Davenport, T.H. and Prusak, L. (1998), *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press.

Denzin, N. K. and Lincoln, Y. S. (Eds), (2018), *The SAGE Handbook of Qualitative Research*, 5th ed., Sage.

Dworkin, G. (1988), *The Theory and Practice of Autonomy. Cambridge Studies in Philosophy*, Cambridge University Press, Vol. 1988, pp. i-vi.

Eisenhardt, K.M. and Zbaracki, M.J. (1992), "Strategic decision making", *Strategic Management Journal*, Vol. 13 No. S2, pp. 17-37.

Faraj, S., Pachidi, S. and Sayegh, K. (2018), "Working and organizing in the age of the learning algorithm", *Information and Organization*, Vol. 28 No. 1, pp. 62-70.

Feldman, M.S. and Orlikowski, W.J. (2011), "Theorizing practice and practicing theory", *Organization Science*, Vol. 22 No. 5, pp. 1240-1253.

Floridi, L. (2011), *The Philosophy of Information*, Oxford University Press.

Frické, M. (2009), "The knowledge pyramid: a critique of the DIKW hierarchy", *Journal of Information Science*, Vol. 35 No. 2, pp. 131-142, doi: 10.1177/0165551508094050.

Gaur, L., Afaq, A., Arora, G.K. and Khan, N. (2023), "Artificial intelligence for carbon emissions using system of systems theory", *Ecological Informatics*, Vol. 76, p. 102165, doi: 10.1016/j.ecoinf.2023.102165.

Gavetti, G. (2012), "Toward a behavioral theory of strategy", *Organization Science*, Vol. 23 No. 1, pp. 267-285.

Gioia, D.A., Corley, K.G. and Hamilton, A.L. (2013), "Seeking qualitative rigor in inductive research", *Organizational Research Methods*, Vol. 16 No. 1, pp. 15-31.

Glikson, E. and Woolley, A.W. (2020), "Human trust in artificial intelligence: review of empirical research", *Academy of Management Annals*, Vol. 14 No. 2, pp. 627-660.

Grant, R.M. (1996), "Toward a knowledge-based theory of the firm", *Strategic Management Journal*, Vol. 17 No. S2, pp. 109-122.

Heisig, P. (2009), "Harmonisation of knowledge management – comparing 160 KM frameworks around the globe", *Journal of Knowledge Management*, Vol. 13 No. 4, pp. 4-31, doi: 10.1108/13673270910971798.

Introna, L.D. (2015), "Algorithms, governance, and governmentality: on governing academic writing", *Science, Technology, & Human Values*, Vol. 41 No. 1, pp. 17-49, doi: 10.1177/0162243915587360. (Original work published 2016).

Jakubik, M. (2023), "Evolution of knowledge management towards wisdom management", *Journal of Information & Knowledge Management*, Vol. 22 No. 06, p. 2350051.

Jakubik, M. and Müürsepp, P. (2022), "From knowledge to wisdom: will wisdom management replace knowledge management?", *European Journal of Management and Business Economics*, Vol. 31 No. 3, pp. 367-389.

Jarrahi, M.H., Askay, D., Eshraghi, A. and Smith, P. (2023), "Artificial intelligence and knowledge management: a partnership between human and AI", *Business Horizons*, Vol. 66 No. 1, pp. 87-99.

Jennex, M.E. (2017), "Big data, the internet of things, and the revised knowledge pyramid", *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, Vol. 48 No. 4, pp. 69-79.

Kaplan, S. and Orlikowski, W.J. (2013), "Temporal work in strategy making", *Organization Science*, Vol. 24 No. 4, pp. 965-995.

Kellogg, K.C., Valentine, M. and Christin, A. (2020), "Algorithms at work: the new contested terrain of control", *Academy of Management Annals*, Vol. 14 No. 1, pp. 366-410, doi: 10.5465/annals.2018.0174.

Laitinen, A. and Sahlgren, O. (2021), "AI systems and respect for human autonomy", *Frontiers in Artificial Intelligence*, Vol. 4, p. 705164.

Leoni, L., Gueli, G., Ardolino, M., Panizzon, M. and Gupta, S. (2024), "AI-empowered KM processes for decision-making: empirical evidence from worldwide organisations", *Journal of Knowledge Management*, Vol. 28 No. 11, pp. 320-347.

Logg, J.M., Minson, J.A. and Moore, D.A. (2019), "Algorithm appreciation: people prefer algorithmic judgment to human judgment", *Journal of Experimental Psychology: General*, Vol. 148 No. 10, pp. 1787-1802.

Magnani, G. and Gioia, D. (2023), "Using the Gioia methodology in international business and entrepreneurship research", *International Business Review*, Vol. 32 No. 2, p. 102097.

Malik, Z., Hashmi, K., Najmi, E. and Rezgui, A. (2019), "Wisdom extraction in knowledge-based information systems", *Journal of Knowledge Management*, Vol. 23 No. 1, pp. 23-45.

March, J.G. (1991), "Exploration and exploitation in organizational learning", *Organization Science*, Vol. 2 No. 1, pp. 71-87.

Martinez, D.E. and Cooper, D.J. (2019), "Assembling performance measurement through engagement", *Accounting, Organizations and Society*, Vol. 78, p. 101052.

Martin, K.E., Shilton, K. and Smith, J. (2022), "Human values and the governance of AI", *Journal of Business Ethics*, Vol. 180, pp. 531-549.

Metaxiotis, K., Ergazakis, K., Samouilidis, E. and Psarras, J. (2003), "Decision support through knowledge management: the role of the artificial intelligence", *Information Management & Computer Security*, Vol. 11 No. 5, pp. 216-221.

Nguyen, T.M. and Malik, A. (2022), "Impact of knowledge sharing on employees' service quality: the moderating role of artificial intelligence", *International Marketing Review*, Vol. 39 No. 3, pp. 482-508.

Nonaka, I. and Takeuchi, H. (1995), *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*, Oxford University Press.

Nonaka, I., Toyama, R. and Konno, N. (2000), " "SECI, Ba and leadership: a unified model of dynamic knowledge creation", *Long Range Planning*, Vol. 33 No. 1, pp. 5-34.

Nutt, P.C. (2008), "Investigating the success of decision making processes", *Journal of Management Studies*, Vol. 45 No. 2, pp. 425-455.

Prunkl, C. (2024), "Human autonomy at risk? An analysis of the challenges from AI", *Minds and Machines*, Vol. 34 No. 3, p. 26, doi: 10.1007/s11023-024-09665-1.

Rai, A. (2020), "Explainable AI: from black box to glass box", *Journal of the Academy of Marketing Science*, Vol. 48 No. 1, pp. 137-141.

Rezaei, M., Pironti, M. and Quaglia, R. (2024), "AI in knowledge sharing, which ethical challenges are raised in decision-making processes for organisations?", *Management Decision*, Vol. 63 No. 10, doi: 10.1108/MD-10-2023-2023.

Rowley, J. (2007), "The wisdom hierarchy: representations of the DIKW hierarchy", *Journal of Information Science*, Vol. 33 No. 2, pp. 163-180, doi: 10.1177/0165551506070706.

Ryan, R.M. and Deci, E.L. (2017), *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*, Guilford publications.

Serenko, A. (2024), "Practical wisdom in the workplace: conceptualization, instrument development, and predictive power", *Journal of Knowledge Management*, Vol. 28 No. 7, pp. 2092-2119.

Shrestha, Y.R., Ben-Menahem, S.M. and Von Krogh, G. (2019), "Organizational decision-making structures in the age of artificial intelligence", *California Management Review*, Vol. 61 No. 4, pp. 66-83.

Simon, H.A. (1972), "Theories of bounded rationality", in McGuire, C.B. and Radner, R. (Eds), *Decision and Organization*, North-Holland, Amsterdam, pp. 161-176.

Sumbal, M.S., Tsui, E. and See-To, E.W.K. (2017), "Interrelationship between big data and knowledge management: an exploratory study in the oil and gas sector", *Journal of Knowledge Management*, Vol. 21 No. 1, pp. 180-196.

Tiwana, A. (2002), *The Knowledge Management Toolkit: Orchestrating IT, Strategy, and Knowledge Platforms*, Prentice Hall.

Troise, C., Battisti, E., Christofi, M., van Vulpen, N.J. and Tarba, S. (2023), "How can SMEs use crowdfunding platforms to internationalize? The role of equity and reward crowdfunding", *Management International Review*, Vol. 63 No. 1, pp. 117-159.

Weinberger, D. (2010), *Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room*, Basic Books.

Further reading

Adadi, A. and Berrada, M. (2018), "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)", *IEEE Access*, Vol. 6, pp. 52138-52160.

Ardito, L., Cerchione, R., Mazzola, E. and Raguseo, E. (2022), "Industry 4.0 transition: a systematic literature review combining the absorptive capacity theory and the data–information–knowledge hierarchy", *Journal of Knowledge Management*, Vol. 26 No. 9, pp. 2222-2254.

Baldassarre, M. (2016), "Think big: learning contexts, algorithms and data science", *Research on Education and Media*, Vol. 8 No. 2, doi: 10.1515/rem-2016-0020.

Benjamins, R., Barbado, R. and Sierra, D. (2019), "Responsible AI by design: explanation and accountability as essential requirements", *AI and Society*, Vol. 34 No. 4, pp. 789-799.

Binns, R. (2018), "Algorithmic accountability and public reason", *Philosophy & Technology*, Vol. 31 No. 4, pp. 543-556.

Bitektine, A. and Haack, P. (2015), "The 'macro' and the 'micro' of legitimacy: toward a multilevel theory of the legitimacy process", *Academy of Management Review*, Vol. 40 No. 1, pp. 49-75.

Bosancic, B. (2020), "Information, data, and knowledge in the cognitive system of the observer", *Journal of Documentation*, Vol. 76 No. 4, pp. 893-908.

Burr, C., Cristianini, N. and Ladyman, J. (2020), "An analysis of the interaction between human and artificial intelligence in decision-making", *Minds and Machines*, Vol. 30 No. 2, pp. 187-209.

Cristofaro, M., Giardino, P.L., Camilli, R. and Hristov, I. (2024), "Unlocking the sustainability of medium enterprises: a framework for reducing cognitive biases in sustainable performance management", *Journal of Management & Organization*, Vol. 30 No. 3, pp. 490-520.

Danaher, J. (2020), *Automation and Utopia: Human Flourishing in a World without Work*, Harvard University Press.

Deepu, T.S. and Ravi, V. (2021), “Exploring critical success factors influencing adoption of digital twin and physical internet in electronics industry using grey-DEMATEL approach”, *Digital Business*, Vol. 1 No. 2, doi: 10.1016/j.digbus.2021.100009.

Doshi-Velez, F. and Kim, B. (2017), “Towards a rigorous science of interpretable machine learning”, doi: 10.48550/arXiv.1702.08608.

Eppler, M. and Mengis, J. (2004), “The concept of information overload: a review of literature from organization science”, *The Information Society*, Vol. 20 No. 5, pp. 325-344.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. and Vayena, E. (2020), “AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations”, *Minds and Machines*, Vol. 30 No. 1, pp. 77-97, doi: 10.1007/s11023-019-09585-8.

Gonçalves, A.R., Pinto, D.C., Shuqair, S., Dalmoro, M. and Mattila, A.S. (2024), “Artificial intelligence vs. autonomous decision-making in streaming platforms: a mixed-method approach”, *International Journal of Information Management*, Vol. 76, p. 102748.

Grieves, M. (2024), “DIKW as a general and digital twin action framework: data, information, knowledge, and wisdom”, *Knowledge*, Vol. 4 No. 2, pp. 120-140, doi: 10.3390/knowledge4020007.

Grote, G. (2021), “Trust in human-AI interaction: models, implications, and governance”, *AI and Society*, Vol. 36 No. 1, pp. 129-144.

Jarrahi, M.H. (2018), “Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making”, *Business Horizons*, Vol. 61 No. 4, pp. 577-586.

Kaplan, S. (2008), “Framing contests: strategy making under uncertainty”, *Organization Science*, Vol. 19 No. 5, pp. 729-752.

Lee, J.D. (2018), “Trust in automation: designing for appropriate reliance”, *Human Factors*, Vol. 60 No. 1, pp. 57-66.

Methnani, L., Aler Tubella, A., Dignum, V. and Theodorou, A. (2021), “Let me take over: variable autonomy for meaningful human control”, *Frontiers in Artificial Intelligence*, Vol. 4, p. 737072.

Omotayo, F. (2015), “ Knowledge management as an important tool in organisational management: a review of literature”, *Library Philosophy and Practice (e-Journal)*,

Pekkanen, J., Kianto, A. and Vanhala, M. (2023), “Autonomy and knowledge processes: a systematic literature review”, *European Conference on Knowledge Management*, *Academic Conferences International Limited*, Vol. 2, p. 1025-R31.

Power, D.J. (2002), “Decision support systems: concepts and resources for managers”, *Westport: Quorum Books*, Vol. 13, pp. 1-20.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.F., Breazeal, C., . . . and Wellman, M. (2019), “Machine behaviour”, *Nature*, Vol. 568 No. 7753, pp. 477-486, doi: 10.1038/s41586-019-1138-y.

Raisch, S. and Fomina, K. (2025), “Combining human and artificial intelligence: hybrid problem-solving in organizations”, *Academy of Management Review*, Vol. 50 No. 2, pp. 441-464.

Willmott, M. (2003), “Citizen brands: corporate citizenship, trust and branding”, *Journal of Brand Management*, Vol. 10 No. 4, pp. 362-369.

Wright, P. and Schultz, P. (2018), “Decision-making in the era of AI: ethical considerations and risk mitigation”, *AI and Ethics*, Vol. 2 No. 3, pp. 245-261.

Xiong, W., Fan, H., Ma, L. and Wang, C. (2022), “Challenges of human-machine collaboration in risky decision-making”, *Frontiers of Engineering Management*, Vol. 9 No. 1, pp. 89-103, doi: 10.1007/s42524-021-0182-0.

Supplementary material

The supplementary material for this article can be found online.

Appendix 1

Table A1 Summary of coding reliability and trustworthiness procedures

Stage	*Unit of analysis*	*% double-coded/n*	*Metric*	*Result*	*Resolution*
First-order (descriptive)	Excerpts within transcript	24.6% (30/122); 412 excerpts	Cohen's κ; overall agreement	Initial: $\kappa = 0.66$; 82% → Postcalibration: $\kappa = 0.79$; 88%	Consensus meetings; third-coder adjudication (11 excerpts; 2.7%)
Stability spot-check	Excerpts within transcript	9.8% (12/122)	Cohen's κ; overall agreement	$\kappa = 0.81$; 90%	No adjudication required
Second-order (interpretive)	Themes across cases	–	Negotiated agreement; audit	Codebook v1.0–1.3; decision logs	Team reviews; negative-case analysis
Aggregate dimensions	Cross-theme patterns	–	Peer debrief; focus groups	Convergence/divergence/ amplification tracked	Revisions to theme boundaries

Source(s): Authors' own work

Appendix 2

Table A2 Cross-tabulation (counts)

Decision making phases	*Data*	*Information*	*Knowledge*	*Wisdom*	*Row total*
Preparation	230	60	24	8	322
Evaluation	40	252	24	20	336
Decision-making	32	12	220	40	304
Implementation	8	8	40	244	300
Total	310	332	308	312	1,262

Source(s): Authors' own work

Table A3 Percentage view (row-wise)

Decision making phases	*Data (%)*	*Information (%)*	*Knowledge (%)*	*Wisdom (%)*
Preparation	71.4	18.6	7.5	2.5
Evaluation	11.9	75.0	7.1	6.0
Decision-making	10.5	3.9	72.4	13.2
Implementation	2.7	2.7	13.3	81.3

Note(s): If an excerpt initially matched multiple DIKW levels, coders resolved to a single primary label. Phase labels follow the temporal/action/artifact/outcome cues exposed within methodology. Inter-coder agreement and adjudication are summarized in Appendix 1
Source(s): Authors' own work

Appendix 3

Table A4 AI system classes mapped to DIKW layers and decision phases, with typical autonomy risks and safeguards

AI class	*Primary DIKW contribution*	*Decision phases most affected (frame/evaluate/ commit/ enact)*	*Typical autonomy risks*	*Autonomy-preserving safeguards*
Predictive/ descriptive (ML and data mining)	Data → information (aggregation, patterning)	Frame, evaluate	Over-weighting historical patterns; metric myopia; false objectivity	Data governance and provenance; counter-metric reviews; alternative framing prompts; sandbox scenario tests
Prescriptive/ optimization and recommenders	Information → knowledge (action suggestions)	Evaluate, commit	Optimization without values; goal mis-specification; automation bias	Value/constraint reviews; multi-objective criteria; opt-out and override protocols; red-team "what-ifs"
Generative AI assistants (LLMs)	Information ↔ knowledge (summarize, ideate, articulate)	Frame, evaluate, enact	Fluent but unfounded rationales; leakage of internal knowledge; prompt steering	Source citation/RAG; confidentiality controls; traceable prompts; justification checklists
Perception and extraction (NLP/ CV/OCR)	Data formation (from unstructured inputs)	Frame, enact	Garbage-in; biased extraction; loss of context	Human validation on samples; bias audits; retention of raw context; escalation rules

Source(s): Authors' own work

Corresponding author

Matteo Cristofaro can be contacted at: matteo.cristofaro@uniroma2.it