

Do We Have the Knowledge We Need?

Rethinking Human-AI Decision-Making in Corporations

ANNE MARX, Department of Computer Science & ETH AI Center, ETH Zurich, Switzerland

RICARDO MAIA AVELINO, Department of Computer Science & Architecture, ETH Zurich, Switzerland

TORBjørn NETLAND, Department of Management, Technology, and Economics, ETH Zurich, Switzerland

MENNATALLAH EL-ASSADY, Department of Computer Science, ETH Zurich, Switzerland

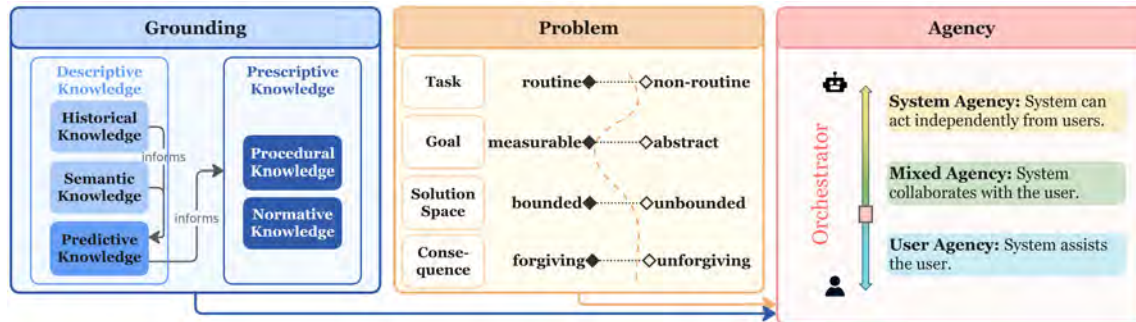


Fig. 1. Our mixed agency model for decision-making in manufacturing consists of a) grounding on available knowledge, b) problem characteristics that impact c) the level of agency.

Abstract. Organizational knowledge is fragmented across a variety of software systems, tacit expertise, and manual documents that have traditionally been designed for human consumption. As AI systems are increasingly deployed and granted decision-making roles, they require access to this knowledge. This raises two questions: how should organizations store and maintain knowledge so that it remains accessible to both humans and future AI systems, and how should agency be allocated between humans and AI across tasks with different risks and levels of uncertainty? In this position paper, we describe how organizational knowledge evolves and contribute a framework that maps task attributes and knowledge availability to recommended agency allocations and control mechanisms. We illustrate the applicability of the framework on two different manufacturing tasks: a routine operation (visual quality inspection) and a one-off strategic decision (factory location), and conclude with opportunities for future research.

CCS Concepts: • **Computing methodologies** → **Knowledge representation and reasoning**; • **Social and professional topics** → **Automation**; • **Human-centered computing** → **HCI theory, concepts and models**.

Additional Key Words and Phrases: Mixed-Initiative Systems, Human-AI Collaboration, Agency, Autonomy, AI Agents, Decision-Making, Manufacturing, Industry 5.0, Knowledge Theory, Knowledge Graph, Knowledge Management, Autonomous AI, AI

ACM Reference Format:

Anne Marx, Ricardo Maia Avelino, Torbjørn Netland, and Mennatallah El-Assady. 2026. Do We Have the Knowledge We Need? Rethinking Human-AI Decision-Making in Corporations. In *Proceedings of AutomationXP26 Workshop of the 2026 CHI Conference on Human Factors in Computing Systems, April 14, 2026, Barcelona, Spain*. ACM, New York, NY, USA, 8 pages.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2026 Copyright held by the owner/author(s).

1 Introduction

The rapid integration of artificial intelligence (AI) into companies and their increasing autonomy levels has shifted the bottleneck for safe and effective deployment from computational power to *knowledge grounding*. Current AI behavior can be unpredictable in novel out-of-distribution situations and erroneous decisions may have considerable negative physical or economic consequences. Furthermore, increased automation creates the risk of *uninvention*, which is the loss of human skill and domain knowledge when processes are automated without being fully understood or semantically represented. This reignites the long-standing challenge of selecting the level of system autonomy and automation [12], now rephrased as the distribution of agency between human and AI when deciding and acting.

In companies, we can define actions as chains of making decisions and executing them, which alter the **company world state (CWS)**: *The current snapshot of everything that the company depends on and that the company can change*, including but not limited to physical assets, digital systems, customer relations, choice of suppliers, and its people and skills. For a more granular analysis of agency distribution in companies, we describe *world state-altering actions* with the **TDE-loop**: A formalized model describing the stages of *trigger*, *decision*, *execution*, and optional *control*.

Prior work often treats autonomy and agency interchangeably [4], focuses on liability [16] or on executive power [4]. We explicitly separate task complexity from autonomy, as even very simple, well-defined tasks can be fully automated, yielding high cognitive autonomy and executive power. Building on human-AI collaboration taxonomies [7, 8], we propose a dynamic allocation of agency within the TDE-loop by a human or AI *orchestrator*, which is informed by knowledge **grounding** and the specific characteristics of the **problem** at hand (see Figure 1). Our contribution consists of (1) A dynamic model for agency: A theoretical framework that links three levels of agency (*user*, *mixed*, and *system*) to a taxonomy of corporate knowledge and problem attributes, and (2) a taxonomy of grounding: We distinguish between *descriptive knowledge* (historical, semantic, predictive) and *prescriptive knowledge* (procedural, normative), providing a roadmap for what industrial knowledge must be captured and used for increased AI autonomy and robustness.

2 Agency over the TDE-Loop for World State-Altering Actions

Previous work formalizes actions in various domains. Human-centric domains [11, 14] emphasize decision-making grounded in knowledge, skills, and context, while system-centric frameworks often aim to define an abstract, generalized procedure: The Monitor-Analyze-Plan-Execute over shared Knowledge (MAPE-K)-loop is a widely recognized engineering framework for autonomic and self-adaptive systems [10]. Reasoning and Acting (REACT) describes large language model (LLM) agents [20] by distinguishing thinking (reasoning) and acting, showing similarities to Kahneman’s [9] slow and fast thinking models. In robotics, scholars talk about the Sense-Plan-Act loop [17]. Both human-centric and system-centric frameworks lack mixed-agency scenarios. Generalizing these frameworks to human-only, mixed human-AI and AI-only decision-making for world state-altering actions, we propose the following TDE-loop with the stages: *Trigger* → *Decision*(→ *Control*) → *Execution*(→ *Control*).

Every action has a *trigger* that initiates the loop. A *decision* follows, which can itself be iterative and trigger further *decisions* or TDE-loops. Afterwards, the actor *executes*, which directly alters the CWS. Observations from the *execution* stage itself can trigger further TDE-loops. Both *decision* and *execution* can be *controlled*, establishing boundaries and safeguards, especially in high-stakes settings. Agency can be split between human or system *users* and AI *systems* across all stages. We speak of AI **system agency** when the AI primarily bears responsibility for the TDE-loop and mostly executes tasks autonomously. In **mixed agency** settings, both the user and the system collaborate, whereas in **user agency**, the user owns the TDE-loop, with the AI system at most assisting the user on demand.

3 Grounding – Descriptive and Prescriptive Knowledge

The availability of different types of knowledge shapes how we can make decisions and impacts agency. We distinguish grounding into *descriptive* (historical, semantic, predictive) and *prescriptive* (procedural, normative) knowledge, drawing on different organizational and cognitive views of knowledge, including Nonaka’s SECI Model [3], Tulving’s episodic and semantic memory [18], and Ackoff’s DIKW hierarchy [19] while adapting them to AI-enabled work (see Appdx. A).

Descriptive knowledge describes the world and can be sub-categorized as follows:

- “What happened?” *Historical knowledge* describes any historical data that was extracted from perceiving the CWS, such as human observations, data generated from machines, or general state-specific information.
- “What is?” *Semantic knowledge* describes (time- and stateless) facts about the company’s world, such as definitions, attributes (e.g., physical properties) or relations.
- “What will (most likely) happen?” *Predictive knowledge* enables the recognition of similar historic situations or logical deductions based on semantics to derive predictions of future CWSs.

Historical and semantic knowledge are directly related to Tulving’s episodic and semantic memory [18]. Once historical data and/or semantic knowledge is collected, it may be possible to predict future CWSs, for example, by training AI models on historical data to perform predictions and thus generate predictive knowledge.

Prescriptive knowledge is derived from predictive knowledge, as understanding or estimating cause-and-effect relationships leads to formulating best practices, manuals, or rules. Prescriptive knowledge is composed of:

- “What am I allowed and not allowed to do?” *Normative knowledge* describes rules and constraints for how to act, such as safety rules, company policies, usage norms, etc.
- “How should or can I do that?” *Procedural knowledge* describes the step-by-step sequence of how to act for a predefined task, often resulting in best practice guides and manuals.

4 Problem Attributes – Regularity, Consequences, Goals and Solution Space

While *grounding* provides the foundation for decision-making, we also need characteristics of the *problem* to determine how we can make the decision and who receives agency. We argue that the key problem attributes for the orchestrator to decide who should have authority over the TDE-loop include whether the situation is *routine*, the severity of the *consequences*, the clarity of the *goals*, and the boundedness of the *solution-space*.

The problem’s regularity is related to the availability of knowledge. The human-centric Skill-Rule-Knowledge (SRK) framework from Rasmussen [14] differentiates decision-making for human operators in (1) skill-based behavior for routine tasks, when not much cognitive effort is required by the operator (implicit procedural and predictive knowledge), (2) rule-based behavior for following explicit guidelines and procedures (explicit prescriptive knowledge), and (3) knowledge-based behavior for goal-controlled tasks in unfamiliar situations (normative and descriptive knowledge). Rasmussen also shows that procedural knowledge tasks lead to fewer errors than knowledge-based tasks [14]. However, procedural knowledge may be hard to obtain, e.g. if the task occurs rarely and mostly under novel circumstances. In these situations, normative knowledge and descriptive knowledge become increasingly important signals.

Perrow et al. [13] distinguish between loose coupling (forgiving) and tight coupling (rather unforgiving consequences) in systems. In manufacturing, a decision made in a tightly coupled system, such as a chemical plant, can propagate errors, potentially catastrophically. Generally, high-stakes actions in these situations follow standardized procedures

to reduce errors and are subjected to additional *controls*. The possible consequences, coupled with the novelty of the situation, are thus a crucial signal for the orchestrator to decide who should have authority over the TDE-loop.

Similarly, decision-making can follow clear, measurable goals, such as finding the cheapest vendor, or be more abstract and ambiguous, such as improving workers’ health, where human intuition can be useful for clarifying the task.

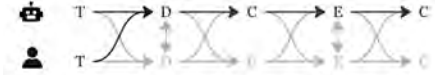
In addition, the solution space may be wider or narrower. Rittel et al. [15] describe the “Wicked Problem” as having no clear final solution and thus no stopping rule. While AI-systems can navigate wider solution spaces than humans, they struggle to act when no measurable goals have been defined. This trade-off is then critical for distributing agency as illustrated in the following case studies.

5 Case Studies

We apply our framework to two problems recently observed by the authors in manufacturing companies.

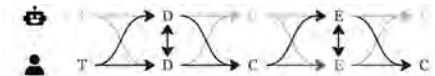
Visual Quality Control. An incoming car part (*trigger*) is tested for visual defects in the paint job (*decision*) and sorted out (*execution*) as faulty or proceed to the next manufacturing stage. **Grounding:** The company has collected thousands of historical examples and predictive annotations. The **problem** has the following characteristics: *routine* problem, clear and *measurable* goal, *bounded* solution space (accept/reject), and *forgiving* consequences.

Agency: An AI model trains on the annotated data and, in the process, develops *implicit procedural knowledge*. The orchestrator chooses system agency in the standard setting. If a novel car part is introduced, mixed agency is used, with the consequences being forgiving: System agency is used in the *decision* and *execution*, while a human *controls*. If the system proves unreliable on the novel part, human feedback can be used to update the model’s insufficient procedural knowledge. Or, explicit semantic knowledge, such as “scratches or missing paint are types of errors,” enables the use of specialized AI models that detect scratches and color changes, e.g., from other domains, or general vision language models, thus making explicit decisions for more robust system actions in these non-routine situations.



Opening a New Factory. In a novel geopolitical situation (*trigger*), a company must choose a new factory location (*decision*) and initiate its construction (*execution*). **Grounding:** The knowledge is dispersed into various types of descriptive knowledge, such as historical national statistics, logistics models, geopolitical developments, taxes, and expected branding impact. The **problem** has the following characteristics: *Non-routine* problem, *measurable* and clear goal, *big* solution space, and *unforgiving* consequences. The **agency** can be mixed if procedural knowledge (such as a playbook for site selection with clear criteria) or semantic knowledge (such as ontologies defining logistical resilience or political stability) is externalized and accessible to the AI. Then, the system could execute smaller tasks such as screening locations, applying externalized logic to filter location candidates, or challenging the user.

Full system agency is unlikely due to the lack of logical reasoning in AI, and *control* must remain with the user. Furthermore, managing stakeholders in this process requires applying nuanced social norms, which so far is reserved for humans.



6 Opportunities: Bridging the Agency-Grounding Gap

When available data perfectly overlaps with the problem space, AI achieves high autonomy. However, the industry currently over-optimizes for *historical* and *procedural* knowledge, neglecting the *semantic* and *normative* layers. We identify three critical frontiers for research.

Resilience-First Design. Gorman [6] warns that the failure to preserve tacit knowledge leads to “uninvention,” while Ganuthula [5] warns about “unlearning”, indicating that AI usage risks human operators losing the critical skills

required to intervene in non-routine failures. We could expect a shift from Efficiency-First to Intentional Inefficiency: systems should occasionally return agency to humans—even for fully automatable tasks—to maintain “skill resilience” and increase robustness in novel situations, and, in addition, capture semantic knowledge for on-demand access. The orchestrator must also optimize for maintaining the human-in-the-loop as a robust fallback system.

Mining the Reasoning Layer. Current documentation is often designed for human consumption, preventing AI from tracing the “why” behind a decision. This opens a vast research opportunity: How can we automatically mine semantic, normative, and procedural knowledge from unstructured legacy data, multi-modal observations, or from human interaction? Emerging work in knowledge extraction [1] demonstrates the potential of using vision language models to convert static manuals into knowledge graphs (KGs). By tracing the provenance of *procedural* knowledge back to *normative*, *semantic*, or *historic* information, we could further enable AI to act more reliably in novel situations where statistical correlations fail. We further ask: What is the optimal knowledge storage for both human and AI accessibility? How will this change AI reliability in novel situations and future human-AI collaboration?

Agency as a Dynamic Variable. One way to capture semantics is to enable future systems to move beyond static automation toward dynamic agency negotiation. The system self-assesses its grounding by checking if the current problem attributes map to its available knowledge before requesting or accepting authority. If the system detects a “grounding gap” (e.g., in our visual quality control case, a novel part with unknown semantic properties appears), it should proactively degrade its agency level and request specific human input to fill the knowledge gap.

7 Conclusion and Future Outlook

We argued that the bottleneck to autonomy and effective human-AI collaboration in companies lies in *knowledge grounding*. We introduced the *TDE-loop* to formalize actions that alter the *company world state* and provide a framework for dynamically allocating agency based on the alignment of knowledge types and problem attributes.

Achieving reliable mixed agency requires a fundamental shift in how we view corporate knowledge. We conclude with four provocations for the future of human-AI collaboration in companies:

- **Explicit Knowledge as the New Digital Gold:** When high-quality training data is scarce, and models rely on context, the competitive advantage will shift to proprietary *reasoning layers*. Specialized and industry-specific knowledge storages with high accessibility may become as valuable as the AI models themselves, leading to a marketplace for “corporate wisdom” rather than just pretrained transformers.
- **Closing the Gap to Reality:** Externalized knowledge can vastly differ from values, processes, and actions taken in reality. These discrepancies will need to be iteratively checked and overcome through feedback loops, e.g., through provenance modeling or interactive sessions. Resolving conflicting realities and norms can open up further research opportunities.
- **Graph-Optimized AI Architectures:** Backed by the resurgence of neuro-symbolic AI [2], we foresee the rise of architectures co-designed with a graph-based knowledge storage to provide native reasoning. These systems will not only retrieve context but also follow logic more reliably, enabling them to navigate irregular problems.
- **The Orchestrator as a Mediator:** The orchestrator of the future will serve as a knowledge farmer and bi-directional translator, such as converting human *normative* intent into explicit machine-executable constraints, and translating machine *predictive* confidence into human-understandable agency requests.

By treating knowledge as a dynamic infrastructure accessible to both humans and AI, we can transition from rigid automation to resilient, mixed agency environments that navigate the unpredictable complexities of the modern world.

Acknowledgments

This research was primarily supported by the ETH AI Center through an ETH AI Center doctoral fellowship to Anne Marx. Additional support was provided by Design++ and Halter Group AG through a postdoctoral fellowship to Ricardo Maia Avelino, and by the Swiss National Science Foundation (SNSF) under grant number 10003068.

References

- [1] Guillermo Gil de Avalor, Laura Maruster, and Christos Emmanouilidis. 2026. Procedural knowledge extraction from industrial troubleshooting guides using vision language models. *arXiv preprint arXiv:2601.22754* (2026).
- [2] Lauren Nicole DeLong, Ramon Fernández Mir, and Jacques D. Fleuriot. 2024. Neurosymbolic AI for reasoning over knowledge graphs: A survey. *IEEE Trans. on Neural Networks and Learning Systems* (2024). doi:10.1109/tnnls.2024.3420218
- [3] Maria Luisa Farnese, Barbara Barbieri, Antonio Chirumbolo, and Gerardo Patriotta. 2019. Managing knowledge in organizations: A Nonaka’s SECI model operationalization. *Frontiers in Psychology* 10 (2019), 2730. doi:10.3389/fpsyg.2019.02730
- [4] K.c.J. Feng, David W. McDonald, and Amy X. Zhang. 2025. Levels of autonomy for AI agents. *arXiv preprint arXiv:2506.12469* (2025).
- [5] Venkat Ram Reddy Ganuthula. 2024. The paradox of augmentation: A theoretical model of AI-induced skill atrophy. *Available at SSRN 4974044* (2024). doi:10.2139/ssrn.4974044
- [6] Michael E. Gorman. 2002. Types of knowledge and their roles in technology transfer. *J. Technology Transfer* 27, 3 (2002), 219–231. doi:10.1023/a:1015672119590
- [7] Steffen Holter and Mennatallah El-Assady. 2024. Deconstructing human-AI collaboration: Agency, interaction, and adaptation. *Computer Graphics Forum* 43 (2024), e15107. doi:10.1111/cgf.15107
- [8] Steffen Holter, Caterina Moruzzi, and Mennatallah El-Assady. 2026. Toward agency in human-AI collaboration. *IEEE Computer Graphics and Applications* 46 (2026), 13–25. doi:10.1109/mcg.2025.3623892
- [9] Daniel Kahneman. 2011. Thinking, fast and slow. *Farrar, Straus and Giroux* (2011).
- [10] Jeffrey O. Kephart and David M. Chess. 2003. The vision of autonomic computing. *Computer* 36, 1 (2003), 41–50. doi:10.1109/mc.2003.1160055
- [11] John E Laird. 2019. *The Soar cognitive architecture*. MIT press.
- [12] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. 2000. A model for types and levels of human interaction with automation. *IEEE Trans. on Systems, Man, and Cybernetics* (2000), 286–297.
- [13] Charles Perrow. 2011. *Normal accidents: Living with high risk technologies*. Princeton University Press.
- [14] Jens Rasmussen. 2012. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Trans. Systems, Man, and Cybernetics* 3 (2012), 257–266. doi:10.1109/tsmc.1983.6313160
- [15] Horst W. J. Rittel and Melvin M. Webber. 1973. Dilemmas in a general theory of planning. *Policy sciences* 4, 2 (1973), 155–169.
- [16] Lisa Soder, Julia Smakman, Connor Dunlop, Weiwei Pan, Siddharth Swaroop, and Noam Kolt. 2024. Levels of autonomy: Liability in the age of AI agents. In *Workshop on Socially Responsible Language Modelling Research*.
- [17] Ankit Srivastava. 2019. Sense-plan-act in robotic applications. In *Intelligent Robotics Seminar*. 1–8.
- [18] Endel Tulving et al. 1972. Episodic and semantic memory. *Organization of Memory* 1 (1972), 381–403.
- [19] Ilkka Tuomi. 1999. Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. *J. of Management Information Systems* 16, 3 (1999), 103–117. doi:10.1080/07421222.1999.11518258
- [20] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R. Narasimhan, and Yuan Cao. 2022. React: Synergizing reasoning and acting in language models. In *Proc. Int. Conf. on Learning Representations*.

Acronyms

AI artificial intelligence. 2–5, 7

CWS company world state. 2, 3

KG knowledge graph. 5, 7

LLM large language model. 2

MAPE-K Monitor-Analyze-Plan-Execute over shared Knowledge. 2

REACT Reasoning and Acting. 2

SRK Skill-Rule-Knowledge. 3

A Theoretical Foundations of Grounding

Our taxonomy of knowledge grounding synthesizes concepts from cognitive science and organizational theory, adapted for human-AI collaboration. Nonaka’s famous SECI model [3] focuses on the cycle of converting tacit to explicit (codified) knowledge in organizations. Parts of our descriptive knowledge mirror Tulving’s cognitive distinction between episodic memory (time-stamped events) and semantic memory (general facts) [18]. Both works do not account for AI systems and how actionable knowledge develops. Finally, Ackoff’s influential DIKW hierarchy focuses on the distinction of data, information, knowledge, and wisdom, which has in the past been criticized, as it is disputable whether knowledge or data comes first [19]. Although we agree on the hypothesis that causal relationships enable forecasting, our framework does not distinguish among mere data, information, and knowledge.

B Comparison of Action Loops

The TDE-loop formalizes world state-altering actions in a way that allows for dynamic agency allocation, differing from existing system-centric loops:

- **vs. MAPE-K [10]:** While MAPE-K is an engineering loop for self-adaptive systems, it misses to integrate humans. The TDE further explicitly introduces an *orchestrator* to negotiate authority between human and AI at each stage.
- **vs. Sense-Plan-Act [17]:** TDE accounts for “triggers” external to the system’s sensors and inserts a “control” stage, which is essential for the *unforgiving consequences* common in manufacturing.
- **vs. ReAct [20]:** TDE also separates the cognitive *decision* from the physical/digital *execution* while further allowing to verify reasoning before the world state is altered.

C The Orchestrator Logic for Agency Allocation

The following logic describes how agency can be assigned by the orchestrator:

- **System Agency:** Appropriate when grounding is high (historical + predictive), the task is routine, and consequences are forgiving.
- **Mixed Agency:** Required for non-routine tasks with possibly high consequences in collaborative settings. Here, the system may suggest *decisions* but the user usually directly retains authority over *execution*, or indirectly through *control* agency. In other settings, the system can also *control* user actions for increased efficiency or safety.
- **User Agency:** Mandatory for abstract goals and novel situations where procedural, normative, or semantic logic is not machine-accessible.

D Knowledge Graphs as Semantic Infrastructure

We advocate for KGs as a reasoning layer.

- (1) **Generalization:** KGs allow AI to reason about novel parts by following semantic links (e.g., “Part X is a subset of Component Y”).

- (2) **Skill Retention:** By externalizing expert logic into graph nodes and relations, organizations mitigate the risk of “uninvention” as domain experts retire and automation levels increase. Knowledge can then be dynamically accessed on-demand through a human-interpretable interface.
- (3) **Explainability:** In mixed-agency settings where AI uses KGs to base it’s decisions, the human can trace the arguments and reasoning back to the KG.