

## Essay

# What AI Cannot Learn: A Cognitive Science Perspective on Human-Centered Strategic HRM

Daniel Altieri <sup>1</sup>, Zohra Damani <sup>2</sup> and Cynthia Nebel <sup>3,\*</sup><sup>1</sup> Department of Leadership, Policy, and Organizations, Vanderbilt University, Nashville, TN 37235, USA; daniel.b.altieri@vanderbilt.edu<sup>2</sup> Independent Researcher, Seattle, WA 98101, USA<sup>3</sup> Department of Psychiatry and Behavioral Neuroscience, St. Louis University School of Medicine, St. Louis, MO 63104, USA

\* Correspondence: cynthia.nebel@health.slu.edu

## Abstract

This article addresses the growing concern that generative artificial intelligence (AI) may replace human expertise in organizations. Instead of asking whether AI should be used, it examines why human judgment rooted in experience cannot be fully replaced by current AI systems and how organizations can work with AI more effectively. Drawing on research from cognitive science, neuroscience, and organizational studies, the paper explains how people use prior experience to interpret context, notice subtle cues, and make sense of ambiguous situations—capabilities that differ fundamentally from how large language models process data. Evidence from recent studies of AI use in hiring, performance management, healthcare, and knowledge work shows recurring problems, including mistakes in unusual cases, missed context, over-reliance on AI recommendations, and reduced visibility of real skill differences among employees. In response, we propose a five-part Human–AI Collaboration Framework designed to help organizations use AI for efficiency while keeping human judgment active and accountable in key Human Resource Management decisions. The analysis shows that AI performs best in routine, data-rich situations but falls short when decisions require lived experience and contextual understanding. By framing organizations as systems built on accumulated experience, this article offers practical guidance for responsible AI integration and outlines directions for future research on human–AI collaboration.

**Keywords:** artificial intelligence; experience-based cognition; human resource management; sensemaking; organizational learning; human-AI collaboration



Received: 1 May 2026

Revised: 11 June 2026

Accepted: 8 July 2026

Published: 10 July 2026

**Copyright:** © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

## 1. Introduction

As the first computer was created, a frontier in computer science opened to try and replicate human cognition in computers (Boden, 2014; Turing, 1950). This intellectual challenge and pursuit became an important new area of study as Artificial Intelligence (AI) brought research blending computer science, linguistics, statistics, electrical and mechanical engineering, and data science with neuroscience and psychology into a new domain (Elmasry, 2024). The quest to make computers think like humans has recently moved away from science fiction and into mainstream life courtesy of key recent advances that we now refer to as Generative AI (Gen AI).

Gen AI (often simply referred to as AI) has been rapidly adopted by organizations seeking to improve efficiency and accuracy (Cohen et al., 2026). From industry consulting

leaders like McKinsey & Company (Mayer et al., 2025) to leading technology research Universities like Stanford, there is a clear recognition that a major technological shift is underway in how firms and people will use AI in many facets of life and work. The 2025 Stanford AI Index Report summarized U.S. Private GenAI investment of \$285.9 billion and a speed of adoption to reach 53% of the global population in three years: faster than other major technology shifts like personal computers or internet use (Maslej et al., 2025). The expectation that AI will create value is widely seen in the financial market valuations and in business leaders' predictions, with McKinsey & Company suggesting that over a trillion dollars in value creation could be expected as AI helps organizations leverage and create new technology (Chiu et al., 2023).

With the deployment of AI within organizations, many individuals fear that AI will quickly replace the human workforce, and researchers have documented that many worker tasks can be done with AI (Eloundou et al., 2024). While this is a valid concern, a thorough understanding of human cognition reveals that AI cannot currently replace the complex processes that develop from human experience. In what follows, we describe how prior knowledge affects the way that human experts process information qualitatively differently from AI in ways that AI can therefore not fully replace human experience-based cognition.

Existing work has largely examined automation bias, trust calibration, metacognition, and reflective practice separately. The present article contributes to this literature in three ways. First, we frame these challenges through the lens of experience-based cognition. Second, we synthesize these insights into a unified Human–AI Collaboration Framework composed of five complementary protocols. Third, we demonstrate how these principles can be operationalized across the Talent Cycle, positioning HRM as a mechanism for preserving and expanding organizational experiential capacity. Instead of arguing for or against the use of AI, we offer a framework that can be used to collaborate with AI, allowing us to harness the power of AI for efficiency while continuing to embrace the unique contribution of human experience-based cognition. This framework can be used at strategic timepoints during Human Resource Management to improve the quality of organizational processes and outputs.

## 2. The Cognitive Science of Experience-Based Cognition

When individuals learn something new, information must travel through several different systems. Before arriving in long-term memory, information is first processed in working memory (previously termed short-term memory). Working memory is the cognitive space in which we are actively working with information (Baddeley, 1992). This space is required for bringing together information from long-term memory with incoming information so that we can reason and problem-solve (Cowan, 2014).

However, working memory is a finite space. Miller (1956) first defined the capacity of working memory to  $7 \pm 2$  pieces of information, meaning that at any given time, we can only attend to an extremely limited amount of information. Later research indicates that the true capacity of working memory might be even smaller (Cowan, 2010) and depends critically on how information is “chunked” (Miller, 1956). That is, as we develop prior knowledge in an area, we are able to structure that knowledge in such a way that it is organized and reduces working memory demands. Take a simple example of baseball knowledge. In a passage about baseball, someone might read that “there was a force at second and the second baseman threw the ball to first to beat the runner by a step.” For a novice in baseball, this sentence requires several different pieces of information to understand and retain. What does “force at second” mean? How many bases are there? Where is the runner running? For someone with experience in baseball, they might process this simply as “a double play.” The prior knowledge about baseball allows the expert to

organize the information into one meaningful chunk, which frees up working memory to process additional information.

This effect has been demonstrated in research as well. [Simon and Chase \(1973\)](#) asked chess masters and novices to examine chess boards that represented real games of chess. The novices were able to hold approximately five chess pieces in working memory—consistent with normal capacity limitations. The experts were instead able to hold approximately sixteen pieces in mind. Using their extensive knowledge of chess plays and games, these masters were able to chunk the board using prior knowledge, strategy, and prediction. While this system often operates under conscious awareness, our ability to move through the world consistently depends on our ability to chunk incoming information in meaningful ways so that we can process more information more efficiently.

Prior experience does not only impact the amount that we can hold onto, but also how we perceive incoming information. [Chi et al. \(1981\)](#) demonstrated that individuals who are novices in physics tend to notice and organize information based on surface features of problems, while experts perceive the underlying concepts and are better able to organize problems based on functional meaning. Similarly, studies in radiology have shown that prior knowledge allows for the processing of subtle abnormalities that novices are likely to miss ([Waite et al., 2020](#)). Prior knowledge helps us to direct attention and changes what we register as meaningful in our environment, creating experience-based cognition.

Echoes of this theory can be found in [Weick's \(1995\)](#) work on sensemaking in organizations. Sensemaking describes the process through which individuals interpret ambiguous or rapidly changing situations by drawing on prior knowledge, shared experience, and existing cognitive frameworks ([Weick, 1993, 1995](#)). Just as experts in domains such as chess or medicine rely on structured knowledge to organize and interpret complex information ([Chi et al., 1981](#); [Simon & Chase, 1973](#)), individuals within organizations rely on shared schemas and past experience to interpret unfolding events ([Weick, 1995](#)).

When people encounter novel or uncertain conditions, they do not process every piece of information independently. Instead, they draw on previously developed mental models to rapidly “chunk” incoming signals into meaningful patterns ([Miller, 1956](#); [Cowan, 2010](#)). These interpretations then guide decision-making and coordinated action. In this way, sensemaking operates as a collective extension of the same cognitive mechanisms that structure individual problem solving ([Weick, 1995](#)). Prior knowledge does not simply increase the amount of information individuals can process; it also shapes how situations are framed, what signals are considered important, and which actions are perceived as viable ([Baddeley, 1992](#); [Chi et al., 1981](#)).

A striking illustration comes from the story of the Mann Gulch Fire, where a team of smokejumpers was overtaken by a fast-moving wildfire. The foreman, Wag Dodge, lit an “escape fire” to create a survivable patch of burned ground. While Dodge immediately recognized a way to survive, his crew could not interpret his action as meaningful within their shared understanding of firefighting. Their established mental models and identity as firefighters—focused on attacking the fire—did not accommodate improvisation in that moment. The mismatch between Dodge’s insight and the crew’s interpretive framework resulted in delayed comprehension and, tragically, loss of life ([Weick, 1993](#)).

This case demonstrates that sensemaking is a collective extension of individual cognitive processes. Just as chunking and prior knowledge allow an expert to efficiently process information ([Simon & Chase, 1973](#); [Chi et al., 1981](#)), shared mental models allow teams to rapidly interpret cues and coordinate action ([Weick, 1995](#)). However, reliance on existing frameworks can also introduce risk. When individuals or groups interpret situations primarily through familiar patterns, they may overlook cues that fall outside established expectations ([Weick, 1993, 1995](#)). Research on organizational sensemaking

has shown that under conditions of uncertainty, stress, or time pressure, the breakdown of shared interpretive frameworks can lead to confusion, delayed action, or misaligned responses (Weick, 1993, 1995). In these situations, effective sensemaking depends not only on individual expertise but also on communication, shared understanding, and the ability to update mental models as new information emerges (Weick, 1995).

Viewed together, research on working memory, expertise, and sensemaking suggests that problem solving is not solely a matter of processing more information, but of organizing experience into meaningful structures that allow individuals and groups to interpret complexity efficiently (Baddeley, 1992; Chi et al., 1981; Cowan, 2010; Miller, 1956; Simon & Chase, 1973).

### 3. Why AI Cannot Fully Replicate Experience-Based Cognition

The recent advances in AI have presented a legitimate challenge in considering whether computers can really think like people. However, significant structural differences remain between machine information processing and the experience-based cognition (EBC) that characterizes human thought. In the last five years, the world has embraced using Large Language Models (LLMs) that represent a significant convergence of machine learning, Big Data, artificial neural networks, and transformer-based attention mechanisms that produced ChatGPT (Bhattacharya et al., 2024; Cohen et al., 2026). Public response has alternated between amazement at AI's apparent insight and concern about its limits.

Current research in phenomenology, neuroscience, and cognitive science clarifies where those limits lie: experience-based cognition is a situated, embodied, and felt phenomenon that current AI architectures cannot fully replicate (Froese, 2026). Current AI cannot easily shift attention in real time to new contextual cues (Chanda & Banerjee, 2022), is restricted to digitized data (Bender et al., 2021), limited by certain accessibility and interface features (Zheng et al., 2022), and cannot understand why it is doing something, which is a dimension that is fundamental to how humans reason and what thinking enables for them (Straňák, 2026).

Embodied cognition theory, rooted in Merleau-Ponty and articulated for AI contexts by Dreyfus and Wrathall (2014), holds that human intelligence is inseparable from the physical body. Where AI “perceives” through computation of static datasets, human experience is grounded in a lived body that interacts with its environment through constant, informal feedback (Susser, 2013). Dreyfus (1996) led the argument that is still being debated, and not yet refuted, that human expertise is not based only on rule-following but on a physical being responding to a situation, drawing from their prior knowledge and evaluating current context. This relies on a vast unarticulated web of social, cultural, and physical skills whose complexity cannot be fully digitized (Su & Luvaanjalba, 2021).

We should not confuse complexity with consciousness as AI models become more complex and more capable of evaluating information. Human information processing as EBC is accompanied by subjective sensations and experiences that are an ephemeral part of thinking and making decisions within a contextual moment (Li, 2025; Ugail & Howard, 2025). Straňák (2026) argues that AI lacks capabilities to do this: it can simulate the output of an experienced person (writing a poem about grief) without undergoing the experience. AI processes the syntax of experience, or how words relate, but not the semantics: what the experience described by the words actually means in a lived context.

Learning mechanisms in humans and AI differ fundamentally. Artificial Neural Networks rely on statistical methodology like backpropagation and gradient descent to minimize error within a fixed mathematical space (Buscema, 1998; Schmidhuber, 2015). Biological cognition involves neuroplasticity, the physical rewiring of the brain in response to chemical and electrical signals (Bassett & Gazzaniga, 2011). Human neurons operate

within a neurochemical environment of neurotransmitters that regulate mood, attention, and memory, transforming raw information into meaningful experience. AI lacks this sensory and hormonal infrastructure (Chavlis & Poirazi, 2021; Kumar et al., 2026).

Experience-based cognition is deeply situated: knowledge is co-constructed between the individual and their social and physical environment. AI operates within the logically and temporally bounded frame of its training data and the context provided in a given interaction. When an exception occurs outside that distribution, AI lacks the human capacity to improvise from a lifetime of unrelated but relevant experience (Chanda & Banerjee, 2022; Dreyfus, 1996; Froese, 2026; Straňák, 2026; Susser, 2013). As we lay out in the next section, industry case studies confirm that while AI excels at high-speed pattern recognition in stable datasets, it fails predictably in situations requiring contextual synthesis, handling of atypical cases, and the kind of integrative judgment we might call intuition.

In sum, while the output from LLMs often appears to be human in nature, AI does not fully replicate human experience. AI cannot fully grasp in-the-moment context, changing conditions, or the user's prior knowledge, which means that it is restricted to the dataset and input provided. Similarly, AI does not have a subjective experience of the current situation and does not process information using the same biological mechanisms as humans that can process environmental information in addition to logically organized data. Collectively, these limitations indicate that artificial intelligence cannot fully substitute for human experience-driven cognition. Furthermore, they clarify the foreseeable challenges organizations encounter when integrating AI into actual decision-making workflows.

#### 4. Identifying Challenges in Working with AI

The cognitive science argument we outline is not merely theoretical: the same structural gaps that separate experience-based cognition from AI, with the lack of embodied context, dependence on digitized training distributions, and the absence of purpose-driven understanding, also surface in organizations as recurring, diagnosable failure modes (Brynjolfsson et al., 2025; Chanda & Banerjee, 2022; Steyvers & Kumar, 2024). In practice, these limitations do not appear as a general inability to be useful or process information using AI. The structural and processing differences for AI create predictable breakdowns in exactly the situations where HR and leaders most need judgment: atypical cases, ambiguous contexts, and high-stakes decisions that depend on tacit knowledge. The next section translates these foundational constraints into four concrete challenges that commonly emerge when people work with AI in organizational settings.

##### 4.1. Challenge 1: A Typical Case Failure—When Pattern Recognition Meets Its Limits

AI's most consistently documented shortcoming is handling cases outside its training distribution. Brynjolfsson et al. (2018) established that AI excels at structured, predictable tasks while faltering wherever tacit knowledge and contextual adaptability are required. A model trained on patterns achieves high average-case accuracy while remaining systematically blind to outliers. Experienced humans, drawing on embodied and situated knowledge, are often best positioned to identify, evaluate, and interpret these outliers to assess if the atypical is more relevant than the standard.

Raghavan et al. (2020) make this concrete in hiring: algorithms trained on historical success data systematically exclude candidates whose profiles differ from past high performers. The AI is functioning as designed, identifying distributional patterns. The problem is that the most relevant information about an atypical candidate's potential, a nonlinear career trajectory, unconventional experience, or background from an underrepresented group, does not appear in the data the algorithm was built to read. Without deliberate human intervention at key decision points, organizations risk systematically

filtering out the diversity of thought and experience that drives innovation and resilience (Nebel & Damani, 2024).

#### 4.2. Challenge 2: Contextual Synthesis Failure—The Limits of Data Without Lived Experience

The “frame problem” is structural: AI can only operate within the frame provided by its training data and the context supplied in any given interaction. Closely related but conceptually distinct is AI’s inability to integrate non-digitized contextual and relational information. When a situation requires cues from the physical, social, or emotional environment, information that has not, and perhaps cannot, be represented as structured data, the AI has no mechanism for accessing it.

A recently reported medical case looking at AI-assisted decisions (Dratsch et al., 2023) makes this failure concrete in a lifesaving context. A senior oncologist overrode an AI’s “benign” classification of a breast lesion not because the AI had processed imaging incorrectly, but because the oncologist integrated observations entirely outside the AI’s perceptual frame: subtle behavioral cues, a nuanced family history detail, and a tactile clinical response. The resulting biopsy revealed early-stage cancer. The AI analyzed available data correctly and still reached the wrong conclusion, because the most clinically relevant information was never in the data. As Jarrahi (2018) argues, human intuition performs precisely this integrative function under ambiguity, and that function, not routine pattern recognition, defines the dimension of human expertise.

In HR contexts, Jiang et al. (2025) document the same failure pattern in AI-assisted performance evaluation. Systems trained on quantifiable output metrics (e.g., task completion rates, response times, and error counts) often systematically underweight the contextual factors experienced managers naturally integrate. An employee navigating more challenging tasks, supporting a struggling colleague, or managing a politically complex stakeholder relationship may show temporarily reduced output while demonstrating the judgment that defines long-term high performance. An AI classifies the dip as underperformance without access to the context a human manager would consider decisive. That same research team also identifies contextual synthesis failure in compensation and promotion recommendations, where informal role expansion, cross-functional contributions, and relational labor—work disproportionately performed by certain demographic groups—often remains invisible to systems that can only measure what has been digitized.

#### 4.3. Challenge 3: Automation Bias—How AI Degrades the Judgment It Is Meant to Support

Perhaps the most urgent organizational risk is not what AI cannot do independently, but what it does to human judgment when people rely on it. Over-reliance on AI can produce cognitive atrophy, outsourcing the deep reflective thinking that defines experience-based judgment to tools that return statistical averages (Kosmyna et al., 2025). Logg et al. (2019) demonstrated that individuals frequently prefer algorithmic advice over human judgment not because the algorithm is demonstrably superior in the given case, but because a machine output feels more defensible and less personally risky. This preference becomes organizationally dangerous when the algorithm is operating outside its competence.

Dietvorst et al. (2015) added a counterintuitive finding: algorithmic trust is psychologically unstable. In this study, participants abandoned algorithmic advice after observing even minor errors, despite the algorithm’s superior average performance. Organizations therefore face a dual risk: workers who defer too readily when AI is confident, and workers who abandon AI assistance entirely after a single visible failure. Neither response constitutes calibrated, selective trust, and neither emerges naturally without deliberate organizational design.

In a large-scale field study of 5000 customer support agents the mechanism becomes concrete. AI assistance produced its largest productivity gains among newer, less experienced agents who used AI output to substitute for judgment they had not yet developed (Brynjolfsson et al., 2025). However, the tool that appeared to accelerate short-term performance was undermining the development of the independent judgment those workers will need when the AI is wrong. Şeker (2024) identifies trust calibration as the central organizational challenge—without a clear metacognitive framework for when deference is appropriate and when independent judgment should override the machine, automation bias is a predictable organizational outcome, not an individual character flaw.

#### 4.4. Challenge 4: Performance Compression and the Masking of Skill Gaps

While investigating business writing tasks, Noy and Zhang (2023) found that generative AI tools improved both speed and quality, with the largest gains accruing to lower-skilled workers. The result was measurable convergence in output quality, a compression that made it significantly harder to distinguish workers with deep competence from those borrowing it from the AI. As average quality improved, variation decreased, attenuating the signal organizations used to identify high performers and invest in their development. An organization operating on compressed performance signals may underinvest in training and make talent decisions based on a picture of capability that reflects AI availability rather than human skill.

Brynjolfsson et al. (2018) anticipated this dynamic: the tasks AI substitutes for most readily—retrieving information, generating standard outputs, executing well-defined procedures—are precisely the tasks through which early-career workers historically build the experiential foundation for more complex judgment. When AI handles those tasks, the developmental pipeline is disrupted. This is a governance and measurement challenge with direct equity implications: performance compression does not affect all workers equally, and evaluating workers against compressed benchmarks may simultaneously overestimate current capability and underestimate the investment required to develop it (Raghavan et al., 2020; Jiang et al., 2025).

These four challenges describe a coherent set of risks emerging at the intersection of AI capability and human judgment. Atypical case failure and contextual synthesis failure identify the boundaries of what AI can reliably process. Automation bias and performance compression identify what happens to human capability when those boundaries are not well understood or managed. None of these challenges argue against AI integration, where the productivity evidence is robust and consistent (Brynjolfsson et al., 2025; Noy & Zhang, 2023; Şeker, 2024). What they outline by design and structural limitation is that AI benefits are visibly and unevenly distributed across skill levels and case types, and the risks of over-reliance are most acute where the stakes are highest: atypical cases, ambiguous contexts, high-uncertainty decisions, and long-term development of genuine human expertise. As Jiang et al. (2025) conclude, a governance framework that keeps human judgment active and accountable at each of these pressure points is not optional—it is the precondition for responsible AI integration.

## 5. AI Collaboration Framework

Having established that AI cannot fully replicate human cognition and that over-reliance creates measurable workforce risks, we turn to the constructive challenge: how to collaborate responsibly with tools that have genuine potential to amplify human capability. The framework proposed here consists of five protocols grounded individually in cognitive science and behavioral research (see Table 1). Critically, while each protocol has an established evidence base drawn from existing research on automation bias, metacognitive

training, trust calibration, and reflective practice, the protocols have not yet been evaluated as a collaborative intervention framework. Each protocol draws on a documented mechanism; translating it into a repeatable organizational routine and balancing it with the other four allows us to propose an AI collaboration framework to mitigate key HRM challenges. To our knowledge, no published framework combines all five protocols into a collective intervention model for AI collaboration in HRM. The framework presented here is a proposed synthesis, drawing on converging lines of evidence to address the specific challenges identified above, and represents a priority area for future empirical research. After outlining the broader framework approach, in the sections that follow we apply it to critical business scenarios HR leaders are navigating now to highlight where and how it may be empirically validated.

**Table 1.** AI Collaboration Framework.

| Protocol                        | Primary Mechanism               | Activity                                                                        | Goal                                                                                |
|---------------------------------|---------------------------------|---------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Preserve-then-Compare           | Cognitive Shielding             | Draft an independent assessment before consulting AI output.                    | Prevents anchoring to AI framing; preserves independent judgment.                   |
| Adversarial AI Training         | Trust Calibration               | Critically evaluate AI in test conditions to identify errors or hallucinations. | Shifts AI from perceived authority to tool requiring active human oversight.        |
| Mandatory Meta-Commentary       | Metacognition/Cognitive Forcing | Review AI suggestion: Accept, Modify, or Reject—with written rationale.         | Forces active engagement; prevents passive acceptance of algorithmic output.        |
| Decision-Calibration Simulation | Social Contextualization        | Role-play where practitioner judgment contradicts a high-confidence AI alert.   | Builds the professional confidence to defend experience-based judgment.             |
| Gap-Analysis Reflection         | Reflective Practice             | Identify contextual cues noticed by humans that the AI model ignored.           | Reinforces the value of lived experience and sensory data that cannot be digitized. |

### 5.1. Protocol 1: Preserve-Then-Compare (Cognitive Shielding)

The central problem that Preserve-then-Compare addresses is AI-induced anchoring bias: the well-documented tendency for an AI system's initial output to serve as a cognitive anchor that distorts subsequent human reasoning. In AI-assisted workflows, anchoring operates at the point of first exposure. Once a worker has seen an AI-generated summary, classification, or recommendation, their subsequent judgment is statistically drawn toward that output regardless of its accuracy (Agudo et al., 2024; Carter & Liu, 2025). This problem is reinforced with Gen AI, where the outputs are fluent, confident in tone, and structurally persuasive. This can amplify the anchoring effect relative to older decision-support tools.

The research basis for this protocol is among the strongest in the framework. [Steyvers and Kumar \(2024\)](#) describe the “sequential paradigm” and “on-demand paradigm” as design principles for AI-assisted decision-making in which workers form an independent judgment before accessing AI advice. The authors document that time-delayed display of AI recommendations improves decision accuracy specifically by allowing workers to reduce the anchoring effect through prior independent reflection. [Rastogi et al. \(Rastogi et al., 2022\)](#) provide direct empirical backing, demonstrating that a time-delay cognitive forcing function, which compels workers to engage with the problem before seeing the AI’s recommendation, served to reduce anchoring bias in controlled AI-assisted decision tasks. In an HR-specific context, [Carter and Liu \(2025\)](#) used two controlled experiments with 775 managers to show that performance ratings were distorted not by actual employee behavior but by the AI’s framing of an anchor. When that research team introduced a “consider-the-opposite” strategy, which functions as a cognitive analog to the Preserve-then-Compare step, they effectively debiased those ratings. [Agudo et al. \(2024\)](#), cited across multiple automation bias reviews, further confirm that when incorrect AI guidance is provided before a person has formed their own judgment, anchoring and automation bias interact to produce compounded error.

### 5.2. Protocol 2: Mandatory Meta-Commentary (Metacognition)

Mandatory Meta-Commentary addresses the issue of passive acceptance of AI output, sometimes called “autopilot reliance” in the human-AI interaction literature. Research consistently demonstrates that workers default to accepting AI recommendations without critical evaluation, not primarily because they trust the AI, but because the cognitive architecture of AI-assisted tools discourages the deliberate analytical effort required for genuine evaluation ([Bućinca et al., 2021](#)). Gen AI amplifies this dynamic because LLM outputs are linguistically polished and plausible, making errors difficult to detect without active scrutiny. The result is a pattern of “effortful appearance, effortless acceptance” that describes how workers engage with AI output by reading it, may feel they have fully evaluated it, but have not engaged the analytical reasoning required to identify errors, omissions, or contextual misalignments.

The empirical foundation for this protocol is rooted in cognitive forcing function (CFF) research. [Bućinca et al. \(2021\)](#) designed three CFF interventions that required deliberate engagement with AI explanations and found they significantly reduced overreliance compared to standard explainable AI approaches. The mechanisms are structurally equivalent to mandatory meta-commentary: they interrupt passive acceptance and require the worker to articulate an analytic response to the AI’s output. [Kupfer et al. \(2023\)](#), when studying automation bias in HR personnel preselection, introduced a “verification intensity” framework in which HR professionals are required to actively verify AI recommendations before acting. What they had found in their research is that participants scoring lower on verification intensity exhibited higher automation bias, directly supporting the value of mandatory structured engagement. [Tankelevitch et al. \(2024\)](#) applied metacognitive theory explicitly to Gen AI and human interaction to evaluate mandatory labeling of AI suggestions as metacognitive scaffolding that counteracts cognitive autopilot. [Ghosh et al. \(2026\)](#), extending this line of work into generative AI writing assistance, tested CFF conditions structurally analogous to the Accept/Modify/Reject labeling schema and found that prompting workers to state reasons for accepting or rejecting AI recommendations engaged analytical thinking in ways that unexplained AI output did not.

### 5.3. Protocol 3: Adversarial AI Training (Trust Calibration)

Miscalibrated trust can be mitigated through Adversarial AI Training to counteract the tendency for workers to develop over-reliance on AI recommendations through unguided experience. Deliberately developing a more accurate and stable trust assessment that can reflect the AI system's actual reliability profile is a critical step to reinforce during adoption of Gen AI technology. While organizational AI deployment research consistently finds that governance structures capable of providing workers with structured exposure to AI errors remain rare in practice (Jiang et al., 2025; Seker, 2024). In the absence of such mechanisms, workers encounter AI errors only through unguided operational experience, and this is a context in which research suggests miscalibration compounds over time rather than correcting (Glickman & Sharot, 2024; Kücking et al., 2024). In more simple terms, AI deployment efforts offer no mechanism for workers to encounter AI errors in a controlled, low-stakes context where the error is knowable. As a result, workers who have never experienced consequential AI failure can treat high-confidence AI output as authoritative, while workers who have experienced a single dramatic failure may overcorrect into algorithm aversion (Dietvorst et al., 2015). Neither trajectory produces calibrated trust.

The research basis draws on converging findings across neuropsychology, empirical human-factors research, and human-AI interaction studies. Glickman and Sharot (2024) provide neuropsychological grounding by demonstrating that repeated human-AI feedback loops actively reshape human perception, emotion, and social judgment over time. Their research found that using Gen AI was not merely reflecting prior biases but generating new ones. This establishes that without intentional exposure to AI errors in a controlled training context, workers develop reinforced over-reliance that may be neurologically resistant to correction through passive experience. Kücking et al. (2024), in an empirical study of automation bias in clinical AI decision support, find that non-specialists are the most susceptible to automation bias precisely because they have the least independent domain knowledge against which to check AI output. Vasconcelos et al. (2023) demonstrate that overreliance is not an inevitability of human cognition but a strategic decision responsive to perceived costs and benefits. They found that workers placed in contexts where the cost of uncritical acceptance was clear after showcasing structured exposure to AI errors, a shift into more analytical evaluation modes took place in consideration of the change in information exposure.

### 5.4. Protocol 4: Decision-Calibration Simulation (Social Contextualization)

Decision-reliance is a distinct problem from cognitive over-reliance when we look at how people and organizations handle information generated by technology-based decision support systems. Logg et al. (2019) establish the baseline dynamic showing individuals systematically prefer algorithmic advice over human judgment, even in conditions where human judgment can be shown to be of superior quality. Their research highlighted that there is a risk created in an organizational climate where overriding an AI recommendation requires active social justification. They found there can be social and organizational pressure to defer to an expert system like AI, even when a practitioner's experiential judgment contradicts a high-confidence algorithmic recommendation. Empirically, a professional may recognize that an AI recommendation is wrong and yet lack the depth of expertise or confidence, organizational standing, or communicative tools to defend a divergent judgment to a supervisor, a committee, or a client.

The research basis for this protocol also draws on trust calibration literature in human-AI interaction and on the broader human factors literature on appropriate reliance. Wischniewski et al. (2023), in a systematic survey of over 1000 studies in the field, confirm that trust calibration is among the most actively researched problems in human-automation

interaction. Their research team identified that dynamic, adaptive calibration, of the kind that role-play simulation is designed to develop, remains an identified gap in the empirically validated toolkit for AI training. Şeker (2024), synthesizing LLM integration research in business contexts, identifies trust miscalibration as a priority governance concern and explicitly notes the absence of structured organizational protocols to address it. Logg et al. (2019) provide direct empirical support for the social dimension of the problem: because workers consistently exhibit algorithmic preference, the skill of publicly defending an experiential judgment that contradicts a high-confidence AI output requires explicit development, not merely cognitive recognition of the problem.

#### 5.5. Protocol 5: Gap-Analysis Reflection (Reflective Practice)

Unlike the other four protocols, which primarily target the moment of AI output consumption, Gap-Analysis Reflection operates prospectively by building a structured practice of naming and reinforcing what human expertise contributes that AI cannot to address the problem of cognitive atrophy from AI over-reliance. Glickman and Sharot (2024) provide the most pressing empirical rationale. Their neuropsychological research demonstrates that, without structured reflection, human-AI feedback loops gradually alter human perception, emotion, and social judgment in the direction of the AI's systematic biases. They found that the gap between human contextual knowledge and AI pattern-matching tends to narrow not by improving the AI capabilities but by degrading the human practices. Gap-Analysis Reflection is designed as a direct structural counterweight to this process.

The research basis combines foundational theoretical frameworks with emerging empirical work on generative AI and tacit knowledge. Schön's (1983) reflective practice framework and Lave and Wenger's (1991) situated learning theory both establish that structured reflection on the relationship between observed practice and underlying knowledge frameworks is essential to the development and preservation of expert judgment. In the AI context, Tankelevitch et al. (2024) extend this logic by framing structured gap-logging as a form of metacognitive audit that makes the invisible visible. They showed that requiring workers to document what their embodied, contextual knowledge contributed contextually served to help define structure to knowledge that the AI's model could not access and counteract passivity of engagement. Guo and Hu (2025), investigating generative AI's role in apprenticeship systems, find that while AI tools can support the articulation of tacit knowledge through facilitated interaction and annotation, capturing tacit knowledge remains a significant challenge, and that structured reflective processes are required to make it visible. This research observation confirmed from the opposite direction that the gap between human tacit knowledge and AI data-driven inference is real, durable, and requires active management. In pre-print, the research team of Fischer et al. (2025) developed a taxonomy of critical reflection prompts for AI-assisted decision-making, including gap-logging prompts to uncover implicit assumptions and factors not considered by the AI model. This is the most recent work directly transferable to workplace Gen AI contexts.

Each protocol thus addresses a distinct failure mode identified in the literature: anchoring and passive acceptance (Protocols 1 and 3), miscalibrated trust and algorithmic decision-bias (Protocols 2 and 4), and erosion of embodied expertise (Protocol 5). As we propose the integrated framework, the protocols are designed to be mutually reinforcing. Adversarial AI Training shapes the trust baseline that Meta-Commentary and Decision-Calibration Simulation then activate in practice, while Gap-Analysis Reflection sustains the experiential foundation that makes Preserve-then-Compare meaningful over time. However, we do not envision a need to rigidly adhere to applying all protocols in all instances to see value in leveraging the AI Collaboration Framework. We envision value in

further research into the interaction effects between protocols and empirical tradeoffs for the suggested cognitive friction balanced against organizational efficiency.

## 6. What Humans (and Leaders) Contribute to HRM in the Age of AI

If AI systems are fundamentally constrained by their reliance on historical data patterns, then the role of HRM is not simply to implement these systems, but to cultivate the range of experiences, perspectives, and meaning-making processes available within an organization.

Rather than thinking of talent as a static resource, The Talent Cycle (Nebel & Damani, 2024) provides mechanisms through which experience is accumulated, structured, and transferred. From a learning science perspective, this reframes organizations not solely as repositories of information, but as distributed systems of lived, embodied, and socially constructed knowledge. In contrast to large language models (LLMs), which generate outputs based on statistical patterns in data, organizations operate through what can be conceptualized as a “large experience model”. This distinction becomes clearer when mapped directly to the Talent Cycle (see Table 2).

**Table 2.** Human and AI contributions at each phase of the Talent Cycle.

| Talent Cycle Phase | AI Contribution                            | Human Contribution                                     |
|--------------------|--------------------------------------------|--------------------------------------------------------|
| Talent Acquisition | Pattern recognition, screening efficiency  | Contextual judgment, identifying potential, future fit |
| Talent Management  | Content delivery, performance tracking     | Meaning-making, experiential learning, development     |
| Talent Retention   | Attrition prediction, engagement analytics | Relationship-building, motivation, cultural continuity |
| Talent Exits       | Knowledge capture systems                  | Knowledge transfer, narrative, network extension       |

This comparison highlights a fundamental distinction: AI operates on historical patterns, while human systems generate value through the creation, integration, and reinterpretation of experience over time. For HR leaders, this reframes the objective from process optimization to the intentional design of experiential capacity. The critical question is not simply who is present in the organization, but what range of experiences, interpretations, and mental models are available when decisions are made. Teams with broader experiential range are better equipped to challenge assumptions, detect weak signals, and generate adaptive responses under conditions of uncertainty (Weick, 1995; Wang et al., 2019). At the same time, reliance on existing frameworks can introduce risk when situations fall outside established expectations (Weick, 1993, 1995), reinforcing the need to continually update and reinterpret experiences. HRM plays a central role in enabling this process by designing systems that promote learning, reflection, and exposure to varied contexts across the Talent Cycle.

If the strategic value of human systems lies in their capacity for experience-based cognition, then HRM must also shape how individuals interact with AI. The challenge is not simply adopting AI tools but embedding them within systems that preserve and strengthen human judgment. Without intentional design, individuals may default to passive reliance on AI outputs, leading to reduced critical thinking and diminished learn-

ing. Learning science suggests that effective human–AI collaboration requires structured cognitive engagement. The practices described in Table 3 translate these principles into HR systems.

**Table 3.** Human-AI Collaboration Framework mapped onto the Talent Cycle.

| Human–AI Design Principle       | Talent Cycle Phase | Organizational Application                                  | Cognitive Outcome                                   |
|---------------------------------|--------------------|-------------------------------------------------------------|-----------------------------------------------------|
| Preserve-then-Compare           | Talent Acquisition | Independent evaluation before AI-assisted screening         | Reduces anchoring bias, preserves original judgment |
| Adversarial AI Training         | Talent Management  | Train employees to identify AI errors and limitations       | Builds critical evaluation and calibrated trust     |
| Mandatory Meta-Commentary       | Talent Management  | Require justification for accepting or rejecting AI outputs | Reinforces metacognition and active reasoning       |
| Decision Calibration Simulation | Talent Retention   | Leaders model when to override AI recommendations           | Strengthens confidence in experience-based judgment |
| Gap-Analysis Reflection         | Talent Exits       | Capture contextual insights not represented in data         | Preserves tacit knowledge and experiential insight  |

These practices ensure that AI functions as a cognitive amplifier rather than a substitute for human thinking. They introduce deliberate friction into decision-making processes, preserving the interpretive work that underpins expertise and effective judgment.

## 7. Embodiment Practices for HR Leaders: Operationalizing Human–AI Design Principles

Leaders play a critical role in reinforcing human cognitive contributions in AI-augmented environments. However, these contributions do not emerge passively; they must be intentionally designed and behaviorally modeled.

To ensure alignment with the Human–AI design principles outlined above, embodiment practices should not be treated as general leadership ideals, but as concrete, repeatable activities embedded across the Talent Cycle. In this way, leaders function as the mechanism through which cognitive science principles—such as sensemaking, chunking, and experience-based judgment—are translated into organizational practice.

The following practices map directly to the Human–AI design principles and illustrate how leaders operationalize them in context:

### 7.1. *Preserve-Then-Compare* → *Model Critical Engagement*

Talent Cycle Phase: Talent Acquisition, Talent Management

Leaders explicitly separate independent judgment from AI input. Before introducing AI-generated recommendations, they require teams to articulate their own assessments, whether evaluating candidates, diagnosing performance issues, or forming strategic hypotheses.

In practice, this includes:

- Drafting candidate evaluations or decision rationales prior to AI augmentation.

- Asking teams: “What do you see before we look at the AI output?”
- Publicly comparing human vs. AI conclusions and discussing differences.

For example: A retail organization is hiring a district manager to oversee multiple stores during a period of operational transformation. An AI-enabled hiring platform ranks candidates based on historical performance data, assessment scores, leadership experience, and interview responses. The system places Candidate A at the top of the list because the candidate closely matches the profiles of previous high-performing district managers. Before viewing the AI ranking, the hiring panel independently evaluates the finalists. Several leaders identify Candidate B as their preferred choice, citing the candidate’s ability to navigate ambiguity, build trust across diverse teams, and adapt communication styles to different stakeholders. When the AI results are revealed, the panel recognizes that Candidate A’s background aligns more closely with historical patterns of success, while Candidate B brings a less conventional career path and broader cross-functional experience. The discussion that follows helps the team distinguish between historical pattern recognition and contextual judgment. Rather than treating the AI output as a definitive answer, the panel uses it as one input into a broader sensemaking process. The preserve-then-compare practice reduces anchoring effects and preserves the organization’s ability to identify capabilities that may be valuable for future, not just past, conditions.

### 7.2. *Adversarial AI Use → Train for Intelligent Skepticism*

#### Talent Cycle Phase: Talent Management

Leaders normalize the expectation that AI outputs should be interrogated, not accepted. Rather than positioning AI as an authority, they design learning environments where employees actively identify errors, gaps, or misalignments.

In practice, this includes:

- Running “AI challenge” sessions where teams critique flawed outputs.
- Embedding prompts such as: “Where might this be wrong?”
- Incorporating AI-error detection into training and development programs.

For example, A hospital system introduces an AI-supported coaching tool designed to generate development plans for nurse managers. During a leadership development workshop, participants receive AI-generated recommendations that intentionally include several flawed assumptions, such as assigning training priorities based solely on productivity metrics while overlooking team communication challenges and patient experience concerns. Participants work in small groups to identify the weaknesses in the recommendations. They note that the AI failed to account for staffing shortages, recent changes in patient acuity, and the emotional demands placed on frontline teams. The exercise shifts the conversation from “What did the AI recommend?” to “What important context is missing?” By practicing adversarial evaluation, leaders develop calibrated trust in AI systems. They learn to recognize AI as a valuable analytical tool while maintaining responsibility for interpreting recommendations within the realities of patient care and team dynamics.

### 7.3. *Mandatory Meta-Commentary → Anchor Decisions in Context*

#### Talent Cycle Phase: Talent Management, Talent Retention

Leaders require explicit justification when AI is used in decision-making. Employees must articulate whether they accept, modify, or reject AI recommendations—and why.

In practice, this includes:

- Embedding justification fields in hiring, promotion, or performance systems.
- Requiring written or verbal rationale tied to contextual factors (team dynamics, culture, timing).
- Encouraging integration of qualitative insights alongside quantitative outputs.

For example, a technology company uses AI to identify employees who are likely to be ready for promotion to senior engineering roles. The system evaluates factors such as project outcomes, technical assessments, peer feedback, and tenure. Managers are required to document whether they accept, modify, or reject the AI recommendation and to explain their reasoning. In one case, the AI recommends promotion for an engineer with exceptional technical metrics. The manager agrees with the recommendation but adds context regarding the employee's recent success in mentoring junior staff and leading cross-functional collaboration efforts. In another case, the AI does not recommend promotion for an engineer whose quantitative indicators are less consistent. The manager overrides the recommendation, citing the employee's critical role in stabilizing a complex product launch and the strong influence the individual has on team learning and problem-solving. These justification requirements encourage managers to integrate contextual knowledge into the decision process. The result is not the rejection of data-driven insights, but a more transparent and reflective approach to interpreting them.

#### 7.4. *Decision Calibration* → *Model When to Override AI*

##### Talent Cycle Phase: Talent Retention

Leaders play a visible role in demonstrating when experience-based judgment takes precedence over algorithmic recommendations. These moments are critical signals that shape organizational norms.

In practice, this includes:

- Publicly overriding AI recommendations and explaining the reasoning.
- Highlighting cases where contextual or relational knowledge changed outcomes.
- Encouraging teams to escalate when intuition conflicts with AI outputs.

For example, a manufacturing company implements an AI system that predicts employees' risk of attrition using engagement survey results, attendance patterns, and productivity trends. The system flags a highly skilled plant supervisor as a high-risk employee likely to leave within six months. The supervisor's manager reviews the alert but knows that the employee has recently been caring for an aging parent and has temporarily reduced participation in optional company activities. Through regular coaching conversations, the manager understands that the employee remains committed to the organization and is actively discussing future career opportunities. The manager documents the contextual information and explains why immediate retention intervention is unnecessary. Later, the case is discussed in a leadership meeting as an example of calibrated trust. The AI correctly identified a change in behavioral patterns, but the manager's relational knowledge provided additional information that altered the interpretation of those patterns. This practice helps leaders distinguish between useful signals and definitive conclusions.

#### 7.5. *Gap-Analysis Reflection* → *Foster Reflection and Knowledge Transfer*

##### Talent Cycle Phase: Talent Exits (and Cross-Cycle Learning)

Leaders institutionalize reflection practices that capture what AI does not see—including tacit knowledge, relational dynamics, and situational nuance.

In practice, this includes:

- Post-decision or post-project debriefs focused on: "What did we notice that the data did not capture?"
- Exit interviews that emphasize experiential insights, not just process feedback.
- Knowledge-sharing forums that translate lived experience into organizational learning.

For example, an energy company faces the retirement of a senior field engineer with more than twenty years of experience managing critical infrastructure projects. The organization has extensive documentation of procedures, maintenance records, and technical

specifications, and AI systems can readily organize and search this information. Yet colleagues repeatedly seek the engineer's advice when unusual operational issues arise because of an accumulated understanding of historical decisions, supplier relationships, and prior failures. Before retirement, the company conducts structured reflection sessions focused on atypical situations the engineer encountered throughout the career. The conversations surface practical insights about negotiating with contractors during emergencies, recognizing early warning signs of equipment problems, and navigating local regulatory relationships—knowledge that is rarely captured in formal documentation. These narratives are added to the organization's knowledge repository and incorporated into training for newer engineers. The process preserves experiential knowledge that would otherwise leave with the employee and strengthens the organization's collective capacity for future sensemaking.

#### 7.6. *Experiential Expansion* → *Champion Varied Perspectives Through Design* Talent Cycle Phase: All Phases

Leaders actively design for exposure to varied experiences, recognizing that cognitive capacity expands through diversity of context, not just diversity of data.

In practice, this includes:

- Cross-functional rotations and stretch assignments.
- Deliberate inclusion of varied perspectives in decision-making forums.
- Structuring teams to include different tenure levels, domains, and thinking styles.

For example, a financial services firm develops a rotational program for emerging leaders. Participants spend time in retail banking, risk management, operations, digital product development, and customer experience teams before moving into senior leadership roles. During a market disruption, one participant draws on experience from multiple rotations to recognize connections between customer complaints, operational bottlenecks, and evolving risk indicators that were not immediately apparent within any single department. Because the individual has worked across diverse contexts, the leader is able to integrate information from multiple perspectives and propose a coordinated response. The organization treats these rotations not simply as career development opportunities, but as a way to expand the range of mental models available within leadership teams. By deliberately broadening experiential exposure, the firm increases its ability to detect weak signals, reinterpret emerging challenges, and adapt to novel situations.

#### 7.7. *From Individual Practices to Organizational Capability*

The embodiment practices described above are most effective when implemented as an integrated system rather than as isolated interventions. Each practice targets a different point within the Talent Cycle, but collectively they strengthen an organization's capacity for learning, interpretation, and adaptation. Viewed together, these practices create organizational conditions that support effective sensemaking (Weick, 1995). Rather than relying exclusively on historical patterns embedded within AI systems, organizations continuously generate, test, refine, and transfer new experiences through human interaction. In this way, the Human–AI Design Principles function as mechanisms for maintaining and expanding the organization's "large experience model" across the Talent Cycle.

## 8. Integrative Insight: Human-Centered Path Forward

AI collaboration is not merely a technological implementation challenge—it is a leadership capability. The examples presented across retail, healthcare, technology, manufacturing, energy, and financial services reveal a consistent pattern: AI contributes pattern

recognition, computational speed, and analytical support, while humans contribute contextual interpretation, relational understanding, embodied experience, and adaptive judgment.

For HR leaders, the implication is clear:

*The goal is not simply to adopt AI, but to design environments where human cognition is continuously exercised, challenged, and expanded in interaction with it.*

As organizations increasingly adopt AI-driven tools across the Talent Cycle, the strategic role of HR shifts from managing people to cultivating the conditions under which human expertise develops, transfers, and evolves. This perspective extends the Talent Cycle framework (Nebel & Damani, 2024) beyond talent management and into organizational cognition. Talent Acquisition introduces new experiences and perspectives. Talent Management transforms experience into expertise through learning and development. Talent Retention preserves critical knowledge and relationships. Talent Exits transfer experience beyond formal organizational boundaries. Together, these processes create and sustain what we have termed a large experience model—a distributed network of lived experiences that enables organizations to interpret complexity, navigate uncertainty, and adapt to changing conditions.

From this perspective, the future competitive advantage of organizations may not be determined by access to the most advanced AI systems, but by their ability to combine AI capabilities with the breadth and depth of human experience available within the organization. While AI can identify patterns from the past, people remain uniquely capable of generating meaning from the present and imagining possibilities that do not yet exist.

AI has the potential to improve organizational outcomes through efficiency, scale, and pattern recognition. However, it cannot replace the embodied cognition that emerges from lived experience, social interaction, contextual awareness, and reflective judgment. The framework proposed here offers one approach for designing human–AI collaboration that preserves and amplifies these capabilities rather than diminishing them.

Future research should empirically examine the proposed Human–AI Design Principles across organizational contexts. Questions remain regarding which practices are most effective under varying levels of uncertainty, how they influence decision quality, and whether they improve organizational learning, innovation, and adaptability over time. Such work would help establish evidence-based guidance for responsible AI integration at scale.

Ultimately, the integration of AI into HRM is not a question of replacement but of design. The Talent Cycle provides the structural foundation, while learning science explains the mechanisms through which experience becomes expertise. Together, they position HR not as an administrative function, but as the architect of how organizations think, learn, and adapt in the age of artificial intelligence.

**Author Contributions:** Conceptualization, D.A., Z.D. and C.N.; writing—original draft preparation, D.A., Z.D. and C.N.; writing—review and editing, D.A., Z.D. and C.N.; supervision, C.N.; project administration, C.N.; funding acquisition, D.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Agudo, U., Liberal, K. G., Arrese, M., & Matute, H. (2024). The impact of AI errors in a human-in-the-loop process. *Cognitive Research: Principles and Implications*, 9(1), 1. [CrossRef] [PubMed]
- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556–559. [CrossRef] [PubMed]
- Bassett, D. S., & Gazzaniga, M. S. (2011). Understanding complexity in the human brain. *Trends in Cognitive Sciences*, 15(5), 200–209. [CrossRef] [PubMed]
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March 3–10). *On the dangers of stochastic parrots: Can language models be too big?* FAccT 2021—Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623), Virtual Event. [CrossRef]
- Bhattacharya, P., Prasad, V. K., Verma, A., Gupta, D., Sapsomboon, A., Viriyasitavat, W., & Dhiman, G. (2024). Demystifying ChatGPT: An in-depth survey of OpenAI's robust large language models. *Archives of Computational Methods in Engineering*, 31(8), 4557–4600. [CrossRef]
- Boden, M. A. (2014). Aaron Solman: A bright tile in AI's mosaic. *Cognitive Systems Monographs*, 22, 9–30. [CrossRef]
- Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics*, 140(2), 889–942. [CrossRef]
- Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn and what does it mean for occupations and the economy? *AEA Papers and Proceedings*, 108, 43–47. [CrossRef]
- Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 188. [CrossRef]
- Buscema, M. (1998). Theory: Foundations of artificial neural networks. *Substance Use and Misuse*, 33(1), 17–199. [CrossRef]
- Carter, L., & Liu, D. (2025). How was my performance? Exploring the role of anchoring bias in AI-assisted decision making. *International Journal of Information Management*, 82, 102875. [CrossRef]
- Chanda, S. S., & Banerjee, D. N. (2022). Omission and commission errors underlying AI failures. *AI & Society*, 39(3), 937–960. [CrossRef] [PubMed]
- Chavlis, S., & Poirazi, P. (2021). Drawing inspiration from biological dendrites to empower artificial neural networks. *Current Opinion in Neurobiology*, 70, 1–10. [CrossRef] [PubMed]
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152. [CrossRef] [PubMed]
- Chiu, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., & Zimmel, R. (2023). *The economic potential of generative AI: The next productivity frontier*. Available online: <https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier> (accessed on 22 March 2026).
- Cohen, M., Khavkin, M., Davidow, D. M., & Toch, E. (2026). ChatGPT in the public eye: Ethical principles and generative concerns in social media discussions. *New Media & Society*, 28(1), 5–31. [CrossRef]
- Cowan, N. (2010). The magical mystery four. *Current Directions in Psychological Science*, 19(1), 51–57. [CrossRef] [PubMed]
- Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, 26(2), 197–223. [CrossRef] [PubMed]
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. [CrossRef] [PubMed]
- Dratsch, T., Chen, X., Mehrizi, M. R., Kloeckner, R., Mähringer-Kunz, A., Püsken, M., Baeßler, B., Sauer, S., Maintz, D., & dos Santos, D. P. (2023). Automation bias in mammography: The impact of artificial intelligence BI-RADS suggestions on reader performance. *Radiology*, 307(4), e222176. [CrossRef] [PubMed]
- Dreyfus, H. L. (1996). Response to my critics. *Artificial Intelligence*, 80(1), 171–191. [CrossRef]
- Dreyfus, H. L., & Wrathall, M. A. (2014). *Skillful coping: Essays on the phenomenology of everyday perception and action*. OUP Oxford. [CrossRef]
- Elmasry, M. I. (2024). *iMind: Artificial and real intelligence* (1st ed.). CRC Press. [CrossRef]
- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. Research is needed to estimate how jobs may be affected. *Science*, 384(6702), 1306–1308. [CrossRef] [PubMed]
- Fischer, S. W. S., Schraffenberger, H., Thill, S., & Haselager, P. (2025). A taxonomy of questions for critical reflection in machine-assisted decision-making. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(1), 940–954. [CrossRef]
- Froese, T. (2026). Sense-making reconsidered: Large language models and the blind spot of embodied cognition. *Phenomenology and the Cognitive Sciences*, (2026). [CrossRef]
- Ghosh, A., Sarkar, A., Lindley, S., & Poelitz, C. (2026). An experimental comparison of cognitive forcing functions for execution plans in AI-assisted writing: Effects on trust, overreliance, and perceived critical thinking. *arXiv*. [CrossRef]
- Glickman, M., & Sharot, T. (2024). How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour*, 9(2), 345–359. [CrossRef] [PubMed]

- Guo, G., & Hu, J. (2025). Making tacit knowledge explicit: Generative AI's role in enhancing apprenticeship systems. *Intelligent Decision Technologies*, 20(1), 431–445. [CrossRef]
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586. [CrossRef]
- Jiang, Y., Cai, Z., & Wang, X. (2025). Leverage Generative AI for human resource management: Integrated risk analysis approach. *International Journal of Human Resource Management*, 36(11), 1929–1959. [CrossRef]
- Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv*. [CrossRef]
- Kumar, A., Devi, M., Singh, S. K., Singh, K., & Mishra, D. (2026). Evolution of neural network models and computing-in-memory architectures. *Archives of Computational Methods in Engineering*, 33, 5805–5838. [CrossRef]
- Kupfer, C., Prassl, R., Fleiß, J., Malin, C., Thalmann, S., & Kubicek, B. (2023). Check the box! How to deal with automation bias in AI-based personnel selection. *Frontiers in Psychology*, 14, 1118723. [CrossRef] [PubMed]
- Kücking, F., Hübner, U., Przysucha, M., Hannemann, N., Kutza, J. O., Moelleken, M., Erfurt-Berge, C., Dissemmond, J., Babitsch, B., & Busch, D. (2024). Automation bias in AI-decision support: Results from an empirical study. *Studies in Health Technology and Informatics*, 317, 298–304. [CrossRef] [PubMed]
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation* (pp. 27–44). Cambridge University Press. [CrossRef]
- Li, B. (2025). Qualia as preemptive constructs: A functional theory of consciousness distinct from cognitive intelligence. *Social Sciences & Humanities Open*, 12, 101876. [CrossRef]
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. [CrossRef]
- Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., Capstick, E., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., Walsh, T., Hamrah, A., Santarlasci, L., . . . Oak, S. (2025). Artificial intelligence index report 2025. *arXiv*. [CrossRef]
- Mayer, H., Yee, L., Chui, M., & Roberts, R. (2025). *Superagency in the workplace: Empowering people to unlock AI's full potential*. Available online: <https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/superagency-in-the-workplace-empowering-people-to-unlock-ais-full-potential-at-work> (accessed on 24 March 2026).
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97. [CrossRef]
- Nebel, C., & Damani, Z. (2024). *Uniting learning science and talent management: Org scholars*. Routledge, Taylor & Francis Group.
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192. [CrossRef] [PubMed]
- Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020, January 27–30). *Mitigating bias in algorithmic hiring: Evaluating claims and practices*. FAT\* 2020—Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 469–481), Barcelona, Spain. [CrossRef]
- Rastogi, C., Zhang, Y., Wei, D., Varshney, K. R., Dhurandhar, A., & Tomsett, R. (2022). Deciding fast and slow: The role of cognitive biases in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1), 83. [CrossRef]
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. [CrossRef] [PubMed]
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, 61(4), 394–403.
- Steyvers, M., & Kumar, A. (2024). Three challenges for AI-assisted decision-making. *Perspectives on Psychological Science*, 19(5), 722–734. [CrossRef] [PubMed]
- Straňák, P. (2026). What artificial intelligence may be missing—And why it is unlikely to attain it under current paradigms. *Philosophies*, 11(1), 20. [CrossRef]
- Su, B. C., & Luvaanjalba, B. (2021). The effect of Hubert Dreyfus's epistemological assumption on the philosophy of artificial intelligence. In *HCI in business, government and organizations* (Vol. 12783, pp. 630–644). Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer. [CrossRef]
- Susser, D. (2013). Artificial intelligence and the body: Dreyfus, bickhard, and the future of AI. *Studies in Applied Philosophy, Epistemology and Rational Ethics*, 5(5), 277–287. [CrossRef]
- Şeker, Ş. E. (2024). Editorial: Large language models in work and business. *Frontiers in Artificial Intelligence*, 7, 1516832. [CrossRef] [PubMed]
- Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2024). The metacognitive demands and opportunities of generative AI. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1, 1–24. [CrossRef]
- Turing, A. M. (1950). I.—Computing machinery and intelligence. *Mind*, LIX(236), 433–460. [CrossRef]
- Ugail, H., & Howard, N. (2025). Quantifying the dynamics of consciousness using hierarchical integration, organised complexity and metastability. *arXiv*. [CrossRef]

- Vasconcelos, H., Jörke, M., Grunde-Mclaughlin, M., Gerstenberg, T., Bernstein, M. S., & Krishna, R. (2023). Explanations can reduce overreliance on AI systems during decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW1), 38. [\[CrossRef\]](#)
- Waite, S., Farooq, Z., Grigorian, A., Siström, C., Kolla, S., Mancuso, A., Martinez-Conde, S., Alexander, R. G., Kantor, A., & Macknik, S. L. (2020). A review of perceptual expertise in radiology-how it develops, how we can test it, and why humans still matter in the era of artificial intelligence. *Academic Radiology*, 27(1), 26–38. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wang, J., Cheng, G. H. L., Chen, T., & Leung, K. (2019). Team creativity/innovation in culturally diverse teams: A meta-analysis. *Journal of Organizational Behavior*, 40(6), 693–708. [\[CrossRef\]](#)
- Weick, K. E. (1993). The collapse of sensemaking in organizations: The Mann gulch disaster. *Administrative Science Quarterly*, 38(4), 628. [\[CrossRef\]](#)
- Weick, K. E. (1995). *Sensemaking in organizations*. Sage Publications.
- Wischnewski, M., Krämer, N., & Müller, E. (2023, April 23–28). *Measuring and understanding trust calibrations for automated systems: A survey of the state-of-the-art and future directions*. Conference on Human Factors in Computing Systems—Proceedings, Hamburg, Germany. [\[CrossRef\]](#)
- Zheng, Q., Tang, Y., Liu, Y., Liu, W., & Huang, Y. (2022, April 29–May 5). *UX research on conversational human-AI interaction: A literature review of the ACM digital library*. Conference on Human Factors in Computing Systems—Proceedings, New Orleans, LA, USA. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.