Review

From Tool to Social Actor: A Systematic Review of the Psychological Mechanisms Through Which Conversational AI Reshapes the Employee Experience

Li Liu , Ruijuan Zhang , Xin Su *

Posted Date: 25 June 2026

doi: 10.20944/preprints202606.1797.v1

Keywords: conversational AI; generative AI; employee experience; anthropomorphism; dark side of AI; systematic review; human-AI collaboration

 doi:10.20944/preprints202606.1797.v1

Review

From Tool to Social Actor: A Systematic Review of the Psychological Mechanisms Through Which Conversational AI Reshapes the Employee Experience

Li Liu [1], Ruijuan Zhang [1] and Xin Su [2,*]

[1] School of Management, China Women's University, Beijing, China

[2] School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing, China

* Correspondence: xin.su@bupt.edu.cn

Abstract

The rapid workplace diffusion of conversational artificial intelligence (CAI) introduces a new social actor, yet existing reviews often neglect the intrapsychic processes through which employees construe these systems. This systematic review synthesizes the psychological mechanisms through which CAI reshapes the employee experience. Following PRISMA guidelines, we analyzed 84 SSCI-indexed studies (2020–2026) using Sociotechnical Systems theory, the Computers as Social Actors paradigm, and the Job Demands-Resources model. Findings reveal a significant theoretical turn: CAI is increasingly conceptualized as a partner or supervisor rather than a tool, fundamentally driven by anthropomorphic cues. The synthesis identifies a double-edged reshaping of work: cognitively, human intelligence augmentation competes with skill threat; emotionally, constant support contrasts with social fabric erosion; and career-wise, inclusion coexists with work alienation and generative AI loafing. These dimensions are psychologically coupled, with cognitive threats often cascading into emotional anxiety and moral expediency. The review provides an integrated conceptual model linking multi-layered antecedents to these outcomes. By shifting the focus from productivity to psychological experience, this study offers theoretical foundations for human-AI collaboration and outlines a future research agenda on trust dynamics and algorithmic fairness.

Keywords: conversational AI; generative AI; employee experience; anthropomorphism; dark side of AI; systematic review; human-AI collaboration

1. Introduction

1.1. A New Social Actor in the Workplace

An unprecedented psychological transition is unfolding in contemporary organizations. When an employee opens ChatGPT, Microsoft 365 Copilot, or an internal chatbot to draft a report, answer customer questions, or support a subordinate, the interaction is not experienced as a conventional software command. It is experienced as a conversation with turn-taking, coherence, and at times apparent empathy (Baygi & Huysman, 2025; Nguyen, 2026). As Microsoft (2024) reports in its *Work Trend Index*, 75% of global knowledge workers now use generative artificial intelligence (AI) in daily tasks, yet nearly half adopted it within the preceding six months, representing a diffusion speed that has outpaced both organizational policy and psychological theory.

This conversational interface, which we refer to throughout as conversational artificial intelligence (CAI), represents more than a technological substitution. CAI systems manifest increasing levels of social presence (Short et al., 1976) and anthropomorphic cues, such as names, first-person pronouns, empathic phrasing, and adaptive learning, that reliably activate employees'

 doi:10.20944/preprints202606.1797.v1

interpersonal schemas (Epley et al., 2007; Nass & Moon, 2000; Pillai et al., 2024). The resulting psychological transition is both subtle and consequential: employees no longer simply use AI; they collaborate with, rely on, distrust, and in some cases emotionally attach to it (Gkinko & Elbanna, 2022; Le et al., 2025). Understanding this shift is therefore a question about the psychology of work in a post-generative-AI era.

Three features of the current moment distinguish it from previous waves of workplace automation. First, conversational AI are encountered directly by end users rather than through specialist intermediaries. While earlier information technologies, including enterprise resource planning systems, decision-support software, predictive-analytics platforms, were mediated by specialists, encountered through tightly scripted interfaces, and understood by employees as back-office infrastructure. Second, they are accessed through open natural-language interaction rather than fixed menus and scripted transactions. Third, employees confront them as counterparts in conversation rather than as back-office infrastructure. Sarkar (2026) argues that this shift changes the psychological locus of work from using a system to working with a system. The implication is not only technical. It affects trust, accountability, and the meaning of competent work. What is at stake is not the adoption of another productivity tool but a quiet renegotiation of what it means to be a competent, connected, and ethically accountable employee.

The psychological gravity of this moment is amplified by structure of contemporary knowledge work. CAI has entered knowledge work at an unprecedented speed: within roughly two years of ChatGPT's public release, it has been integrated into the routines of recruitment, performance feedback, knowledge search, creative composition, and emotional counselling (Budhwar et al., 2023; Hughes et al., 2026). At the same time, employees are often left to improvise the boundaries of appropriate delegation, trust, and disclosure. Because CAI produces fluent human-register outputs, employees may also over-extend trust, reduce verification, or redirect help-seeking away from colleagues and toward the system (Baygi & Huysman, 2025; Nguyen, 2026). These features jointly turn workplace CAI into a living experiment in the psychology of human–machine sociality, an area that deserves theoretical articulation beyond the adoption-intention paradigm that still dominates much of the field (Pillai et al., 2024).

1.2. Theoretical Lenses

We draw on three complementary theoretical lenses. First, Sociotechnical Systems (STS) theory (Trist & Bamforth, 1951; Cherns, 1976) treats organizations as jointly optimized socio-technical arrangements. When a new technology enters work, it changes not only task execution but also patterns of trust, help-seeking, and mutual recognition. From this perspective, CAI raises a direct question about the social fabric of work, which Baygi & Huysman (2025) operationalized as the density of interpersonal help relations vulnerable to being replaced by human–AI exchanges.

Second, the Computers-Are-Social-Actors (CASA) paradigm (Nass & Moon, 2000; Reeves & Nass, 1996) and the anthropomorphism framework (Epley et al., 2007) explain how a conversational interface becomes a social counterpart in the employee's mind. Even when users know they are interacting with software, linguistic fluency, apparent affect, and persona continuity can activate social scripts and elicit trust, reciprocity, attachment, moral expectation and disappointment. Nguyen (2026) provides direct support for this logic by showing that anthropomorphic features in a company generative-AI system increase employee engagement and knowledge contribution through perceived social presence.

Third, the Job Demands–Resources (JD-R) model (Bakker & Demerouti, 2017; Demerouti et al., 2001) helps explain why CAI produces both enabling and harmful consequences. As a job resource, CAI can offload cognitive load and scaffold creative problem finding, as shown in Lin et al.'s (2024). As a job demand, it can generate prompt burden, skill-threat anxiety, and work alienation, as shown in Hai et al.'s (2025).

Three additional psychological considerations deepen these lenses. Conservation of Resources theory (Hobfoll, 1989) complements the JD-R model by predicting that employees facing AI-induced

 doi:10.20944/preprints202606.1797.v1

skill threat enter a resource-protection mode. In this state, they defend their remaining expertise through vigilant investment or avoidance, a mechanism consistent with the GenAI Loafing pattern observed by Saluja et al. (2025). Self-Determination Theory (Ryan & Deci, 2000) suggests that the long-term motivational effects of CAI depend on whether the technology is experienced as autonomy-supportive or autonomy-thwarting. This aligns with Zhang et al.'s (2025) finding that generative-AI use simultaneously promotes both incremental and radical creativity through distinct intrinsic-motivation pathways. Finally, Social Identity Theory (Tajfel & Turner, 1986) sharpens the CASA lens by positing that when CAI is admitted into the ingroup of "colleagues," the boundary between human and non-human coworker becomes psychologically porous. This assimilation carries consequences for loyalty, reciprocity, and perceived group continuity that remain empirically underexplored (Le et al., 2025).

Integrating these perspectives yields the following proposition: Through anthropomorphic cues (CASA), CAI becomes embedded in the socio-technical fabric of organizations (STS); subsequently, via the dual pathways of resource-gain and demand-cost (JD-R), it simultaneously enables and erodes employees' cognitive, emotional, and ethical experiences. This integrative proposition structures our research questions and analytic architecture, which are empirically articulated through the synthesis of 84 studies below.

1.3. Research Gaps and Research Questions

While the intersection of AI and human resource management (HRM) has garnered significant scholarly attention, existing systematic literature reviews reveal critical theoretical and empirical gaps. Table 1 contrasts five published SLRs with the present review and shows a recurring pattern: prior syntheses map AI applications, strategic themes, ethical risks, or sustainability domains, but they do not systematically center the lived psychological experience of employees interacting with conversational AI as a social actor, nor do they map the underlying mechanisms shaping this transition.

Table 1. Comparison of Published SLRs and the Present Review.

Author & Year	Research Focus	Primary Level of Analysis	Theoretical Lens / Organizing Logic	Unresolved Gaps (Limitations)Title
Vrontis et al. (2022)	Intelligent automation strategies and organizational performance.	Macro (Organizational) & Workforce	Multidisciplinary HRM-IM-IB-GM synthesis.	Treats AI as infrastructure; lacks exploration of micro-level psychological role transitions.
Jatobá et al. (2023)	AI adoption in HRM functions (recruitment, training, etc.).	Meso (Strategic/Managerial)	Cluster-based content analysis of adoption paths.	Foreground managerial adoption, neglecting the fine-grained lived experience of employees.
Úbeda-García et al. (2025)	Thematic mapping of AI-HRM interactions (automation to personalization).	Macro & Meso (Field/Organizational)	Bibliometric science mapping.	Maps broad themes (e.g., trust, technostress) without an operationalized micro-level framework.
Kekez et al. (2025)	Bias and discrimination in AI-enabled HRM decisions.	Algorithmic & Policy	Fairness and ethical implications.	Isolates the ethical dimension; does not integrate it within a multi-dimensional employee experience architecture.

 doi:10.20944/preprints202606.1797.v1

Alherimi et al. (2025)	AI advancements in Green HRM and sustainability.	Macro (Sectoral/Organizational)	Sustainability and adoption models.	Domain-specific; fails to explain how AI becomes a social actor in general workplace routines.
The Present Review	Psychological mechanisms reshaping employee experience via CAI.	Micro (Interpersonal/ Psychological)	CASA + STS + JD-R framework.	Moves from "AI-as-infrastructure" to "AI-as-social-actor", providing an integrated antecedent-process-outcome model.

Three theoretical gaps in the extant literature motivate this review. First, published reviews have mapped AI across HR functions, strategic adoption, and organizational levels, but they largely retain an AI-as-infrastructure assumption. Vrontis et al. (2022) organize the field around advanced technologies, AI, and robotics in HRM strategies and activities; Jatobá et al. (2023) cluster the literature into strategic HR, recruitment, training, and future of work; and Úbeda-García et al. (2025) identify six fundamental strategic research themes through bibliometric mapping, ranging from automation and predictive analysis to the personalization of the employee experience. Even when employee well-being, trust, or technostress appear in these reviews, they are nested within broad field-level themes rather than analyzed as a role transition in which employees begin to treat Conversational AI (CAI) as a digital colleague or social actor (Le et al., 2025). No prior synthesis has systematically characterized the psychological profile of this shift or its anthropomorphic antecedents (addressed by RQ1).

Second, existing reviews treat ethical, social, and experiential strands in a segmented way. While Filippelli et al. (2026) proposed a three-dimensional well-being schema (emotional, social, cognitive), the ethical dimension remains under-theorized as a constitutive part of employee experience. This fragmentation is also visible across the published SLRs. Kekez et al. (2025) provide a focused review of bias and discrimination in AI-enabled HRM, but their contribution is intentionally bounded to conceptual definitions, topic distributions, and positive-versus-negative implications of fairness-related harms. Alherimi et al. (2025), by contrast, review AI in Green HRM and map technologies, sectors, and performance effects, yet employee experience appears mainly through sustainability and adoption lenses. Bankins et al. (2024), in their multilevel review of AI in organizations, likewise called for a micro-psychological agenda but provided no operationalized coordinate system. A framework that integrates cognitive, emotional-social, and career-ethical dimensions, and articulates their cross-dimensional coupling, is still missing (addressed by RQ2).

Third, published reviews are rich in thematic mapping but stop short of a process model linking antecedents, psychological mechanisms, and consequences. Jatobá et al. (2023) differentiate positive and negative approaches to AI adoption; Úbeda-García et al. (2025) identify strategic themes and recurring challenges such as technostress, trust, and fairness; and Alherimi et al. (2025) catalogue implementation factors such as organizational readiness, leadership support, and ethical governance. Yet these drivers and consequences remain catalogued rather than theorized as an employee-level antecedent-process-outcome chain. Likewise, the opinion pieces of Budhwar et al. (2023) and Dwivedi et al. (2023) foregrounded multiple promises and perils of generative AI but did not construct a testable antecedent–process–outcome model. The field therefore lacks a synthesis specifying which individual- and organizational-level antecedents push employees onto which psychological path, and under what conditions the double-edged effects of human-AI interaction tilt toward flourishing or harm (addressed by RQ3).

Accordingly, we pose three interlocking research questions, structured in a progressive logical sequence:

RQ1 (role evolution; CASA lens). How has CAI evolved from an automation tool to a partner or supervisor in employees' perceptions, and what anthropomorphic mechanisms underlie this transition?

 doi:10.20944/preprints202606.1797.v1

RQ2 (multi-dimensional experience; Integrative STS and JD-R lens). How is the employee experience reshaped across cognitive, emotional–social, and career–ethical dimensions through human–AI collaboration?

RQ3 (antecedents and double-edged outcomes; Integrative STS and JD-R lens). Which individual- and organizational-level antecedents drive this reshaping, and what positive (enabling) and negative (dark-side) consequences follow for employees' behavior and the social structure of work?

The three questions are progressive: RQ1 characterizes the new actor (what); RQ2 maps its operating structure on employees (how); RQ3 traces the antecedent–outcome chain (why and with what consequences).

1.4. Contributions and Roadmap

A note on terminology is warranted before proceeding. The boundaries between "conversational AI", "generative AI" and "large language models" are not entirely stable in the literature. In this review, CAI refers to systems whose primary interface is natural-language conversation, regardless of the underlying generative-model architecture. This definition aligns with the usage by Hughes et al. (2026) and Pillai et al. (2024). When a study refers specifically to generative AI without necessarily implying a conversational interface (for example, an image-generation system), we note this distinction explicitly; conversely, when a study explicitly examines chatbot features, we treat it as squarely within our scope. Readers should understand CAI as the conversational-interface instantiation of the broader generative-AI phenomenon. We emphasize this form because it is specifically the conversational interface that activates the psychological mechanisms motivating our review, namely the social actor default, anthropomorphic attribution, interpersonal-script recruitment.

This review advances the AI-at-work literature through three interconnected theoretical contributions. Fundamentally, it pivots the discourse from a macro-economic to a psychological-mechanism register by documenting how CAI becomes a social counterpart in employees' mental models, detailing both the resulting psychological costs and benefits. To capture this experiential complexity, the review introduces a three-dimensional coordinate system encompassing cognitive, emotional-social, and career-ethical dimensions. This framework not only operationalizes Bankins et al.'s (2024) call for micro-level research but also extends Filippelli et al.'s (2026) well-being schema by elevating the ethical dimension to a peer construct. Beyond this structural mapping, our synthesis articulates a set of emergent dark-side phenomena that were largely unforeseen by prior opinion-paper agendas. These unintended consequences include GenAI Loafing (Saluja et al., 2025), workplace cheating (Song et al., 2025), expediency (Hai et al., 2025), and social-fabric rupture (Baygi & Huysman, 2025). Following this introduction, section 2 details methodology, and Section 3 describes the resulting literature corpus. Sections 4 and 5 present the synthesized themes, leading into a comprehensive discussion in Section 6. finally, Section 7 concludes the review by proposing a forward-looking research agenda.

2. Methods

2.1. Search Strategy

This systematic review adheres to the PRISMA 2020 guidelines (Page et al., 2021). Following the methodology of prior systematic reviews in this domain (e.g., Úbeda-García et al., 2025), we limited our search exclusively to the Web of Science (WoS) Core Collection (SSCI/SCI-E/ESCI) database. This decision was made to ensure the inclusion of high-quality, peer-reviewed studies, which are widely recognized as the most authoritative sources for mapping knowledge structures and theoretical tensions in management research. The search was conducted on 9 February 2026. The protocol was not pre-registered, a limitation we revisit in Section 5.4. The Boolean search string combined two concept sets:

 doi:10.20944/preprints202606.1797.v1

TS = ((“chatbot*” OR “Chat bot*” OR “conversational Agent” OR “Virtual Assistant*” OR “Digital Assistant*” OR “conversational AI” OR “generative AI” OR “ChatGPT” OR “large language model*” OR “LLM”) AND (“human resource” OR “HRM” OR “Personnel Management” OR “workforce” OR “employee*” OR “Talent Management” OR “Staff*”))

The initial search yielded 1,210 records.

2.2. Inclusion Criteria and Screening

During the automated database screening phase, we sequentially applied several structural filters. We limited the publication window to 2015–2026 (n = 1,120) and strictly retained records indexed in the SSCI (n = 411). The corpus was further restricted to English-language peer-reviewed articles, review articles, and early-access papers (n = 400).

Next, to ensure thematic relevance, we conducted a disciplinary refinement using Web of Science subject categories. We retained records from domains closely aligned with organizational behavior and technology adoption, specifically: Management, Business, Applied Psychology, Information Science & Library Science, Computer Science Information Systems, and Industrial Relations & Labor (n = 194). Conversely, we explicitly excluded structurally misaligned fields such as clinical medicine, industrial/civil engineering, urban planning, and telecommunications, thereby reducing the pool to 161 records.

Finally, in the manual eligibility assessment phase, we conducted a comprehensive title, abstract, and full-text review. Articles were retained only if they empirically or theoretically investigated employees, job applicants, or HR managers, and explicitly addressed the micro-psychological experience of interacting with Conversational AI. This manual evaluation excluded an additional 77 out-of-scope papers, resulting in a final high-quality corpus of 84 studies.

 doi:10.20944/preprints202606.1797.v1

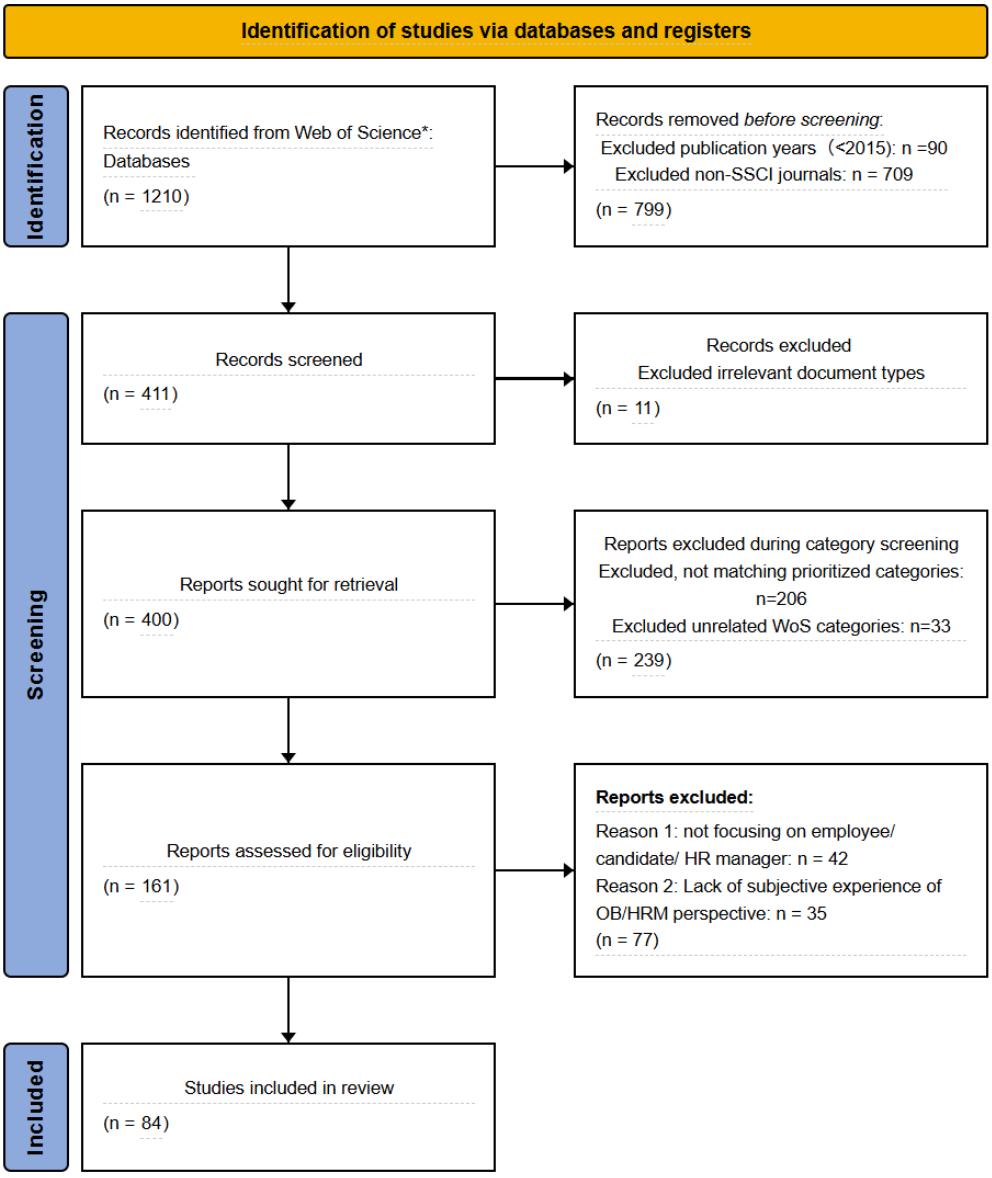

Figure 1. PRISMA 2020 flow diagram for the systematic review.

2.3. Theory-Driven Coding Framework

Two researchers (the authors) developed a 16-dimension coding framework grouped into five analytic categories, deductively anchored in the three theoretical lenses (STS / CASA / JD-R) and inductively refined through a ten-paper pilot (Table 2). The framework reflects three methodological commitments. First, the unit of analysis was fixed as the direct human party of the human-AI interaction (employees, job applicants, or HR managers) to avoid conflating HR professionals' experience with employees. Second, we distinguished employee-end dark-side behaviors (GenAI Loafing, alienation, cheating, social-fabric rupture) from technology-end boundary conditions (AI hallucination, knowledge-based incompleteness). Third, the framework deliberately includes a career–ethical dimension alongside cognitive and emotional-social dimensions identified by Filippelli et al. (2026), because roughly one-third of studies in our corpus substantively engaged with algorithmic fairness, alienation, or workplace cheating.

Coding was performed independently by two researchers. A ten-paper pilot calibrated the framework; the remaining 74 papers were coded blind and then reconciled. Across the 5 primary dimensions (comprising 16 specific coding categories) for the 84 selected papers, the coders made a total of 1,344 independent coding decisions. The initial inter-coder agreement was extremely high at

 doi:10.20944/preprints202606.1797.v1

93.6% (1,258 agreements and 86 disagreements). All 86 minor discrepancies were subsequently resolved through joint discussion and revisiting the source texts to reach a 100% final consensus. This rigorous process ensured high reliability of the extracted data.

Table 2. Theory-Driven 16-Dimension Coding Framework.

Level-1 Category	Level-2 Dimension	Theoretical Anchor & Coding Description
Basic Study Characteristics	1. Publication year	Year of publication
	2. Journal & discipline	the specific journal of publication
	3. Document type	Empirical / Review / Conceptual
Theoretical & Methodological Foundation	4. Dominant theory	The primary theoretical lens
	5. Research method	Quantitative / Qualitative / Mixed / Conceptual
	6. Sample & context	sample size and organizational setting
Role & Attributes of Conversational AI	7. Technology form	Generative AI / LLMs (ChatGPT)/ Enterprise Chatbot
	8. Anthropomorphism	Strong / Moderate / Weak / None
	9. AI role	Tool / Partner / Supervisor / Assistant /Competitor
Core Dimensions of Employee Experience (EX)	10. Cognitive EX	Cognitive Relief/ Cognitive Augmentation/ Cognitive Depletion /Cognitive Strain
	11. Emotional–social EX	Positive (Trust/Empathy) / Negative (Anxiety/Fear)/ Social Alienation /Psychological Safety
	12. Career–ethical EX	Fairness / Justice/ Privacy & Data Security /Job Insecurity / Displacement / Professional Identity
Antecedents & Outcomes of Human-AI Interaction	13. Individual antecedents	AI Literacy / Digital Skills /Trust Propensity/ Personality (Openness) /Prior Expectations
	14. Organizational & task antecedents	Task Complexity/Org Support / Climate / Governance & Policy/ Role/Workflow Design
	15. Positive outcomes	Productivity & Efficiency / Decision Quality/ Well-being & Engagement/ Creativity /Enhanced Opportunities
	16. Negative outcomes	Deskilling & Over-reliance / Resistance / Abandonment/ Job Displacement Anxiety/ Bias & Inequality

To strengthen interpretability, we also recorded the theoretical orientation, sampled population, and relationship to the three EX dimensions (cognitive, emotional and social, and career and ethical) for each study. Approximately half of the empirical studies explicitly invoked formal psychological, organizational, or information systems theories. The most frequently adopted frameworks included Technology Acceptance models (such as TAM and UTAUT), Affordance Theory, the Job Demands-Resources (JD-R) model, and Social Exchange Theory. Another substantial portion of the literature implicitly mobilized these constructs without formal theoretical anchoring. Furthermore, nearly a quarter of the empirical studies were atheoretical in a strict sense, reporting primarily descriptive, practical, or design-oriented findings

For the present review, we treated all three strata as legitimate contributions but synthesized them at different levels of theoretical abstraction. Specifically, explicitly theorized studies directly informed the integrated conceptual model (presented later in Section 4.3, Figure 5). Implicitly theorized studies supplied corroborative evidence for the structural patterns, whereas atheoretical studies furnished ecological grounding and contextual richness. This graded synthesis strategy aims to honor the full range of evidence while avoiding the methodological overreach of extracting causal theoretical conclusions from purely descriptive data.

 doi:10.20944/preprints202606.1797.v1

2.4. Quality Appraisal and Corpus Rigor

Given the highly interdisciplinary and rapidly evolving nature of this topic, traditional numerical scoring (e.g., MMAT) was adapted to focus on methodological transparency and source credibility. The quality of the 84 included studies is reflected in their rigorous empirical designs and publication outlets. Our corpus is overwhelmingly empirically driven, with 60.7% (51 studies) adopting strict quantitative or mixed-method designs (e.g., PLS-SEM, field experiments), and 15.5% (13 studies) employing in-depth qualitative case studies. Furthermore, the credibility of the findings is bolstered by their publication in top-tier outlets across Human Resource Management, Information Systems, and Organizational Behavior. All 84 studies met the inclusion criteria for methodological soundness. Consequently, all were retained to capture the full spectrum of the phenomenon, while the specific research designs (quantitative vs. qualitative vs. conceptual) were explicitly tracked in our coding matrix to weigh the evidence during synthesis.

3. Descriptive Results

3.1. Publication Trajectory and Methodological Mix

The 84 studies display a sharply accelerating publication trajectory, as illustrated in Figure 2. corpus includes two studies each in 2020, 2021, and 2022. This baseline rises to 7 publications in 2023, 14 in 2024, and 48 in 2025, with a further 9 early-access publications already indexed for 2026. Combined, the period from 2024 to 2026 accounts for 84.5% of the entire corpus. This pattern closely synchronized with the public release of ChatGPT in late 2022 and the subsequent wave of scholarly response, which was initially exemplified by rapid-turnaround opinion pieces such as Budhwar et al. (2023). Regarding the methodological mix shown in Figure 3 and 4, empirical studies dominate and account for exactly 71.4% of the corpus. Specifically, this comprises 31 quantitative studies (36.9%), 16 mixed-methods studies (19%), and 12 qualitative studies (14.3%). Conceptual and theoretical frameworks account for 15 studies (17.9%), while review-style pieces make up 8 studies (9.5%). This empirical density indicates the field has rapidly moved beyond agenda-setting and into testable propositions, providing a defensible evidence base for the qualitative synthesis below.

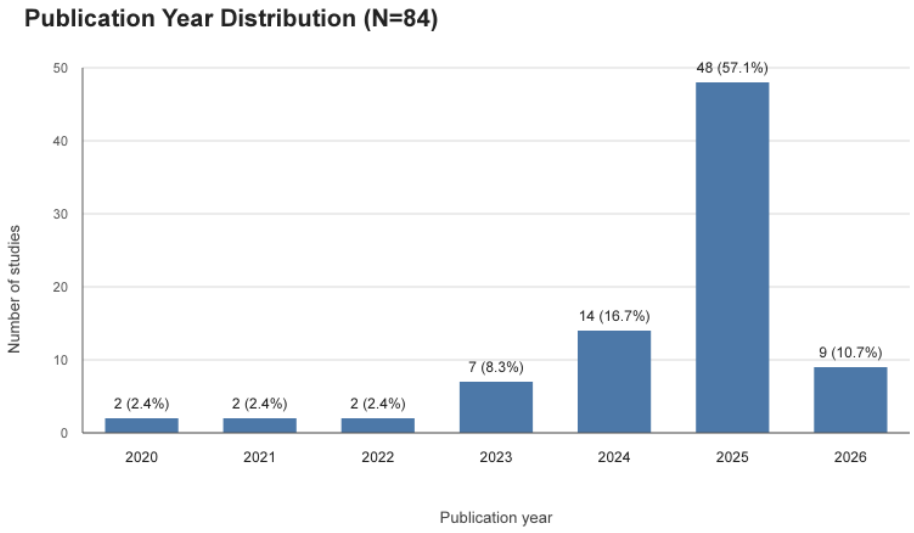

Figure 2. Publication trend of research on CAI and employee experience (2020–2026).

 doi:10.20944/preprints202606.1797.v1

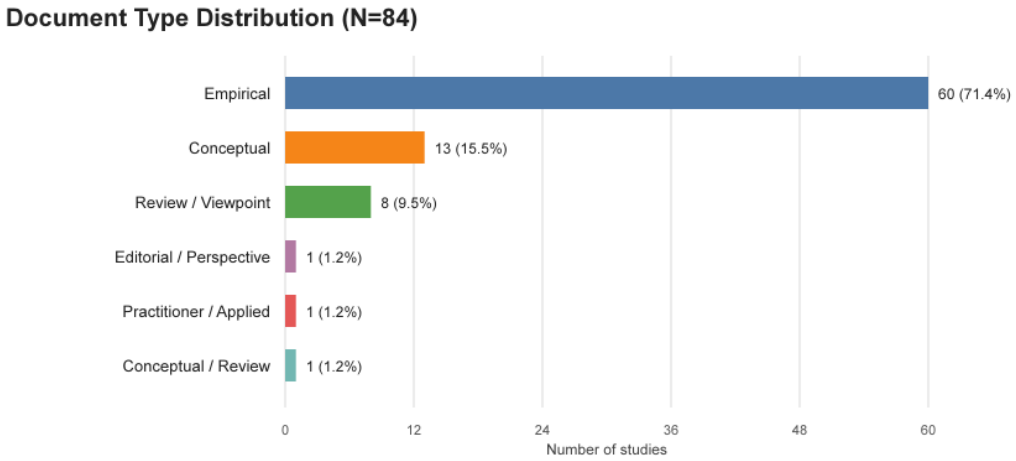

Figure 3. Distribution of document type across included studies (N = 84).

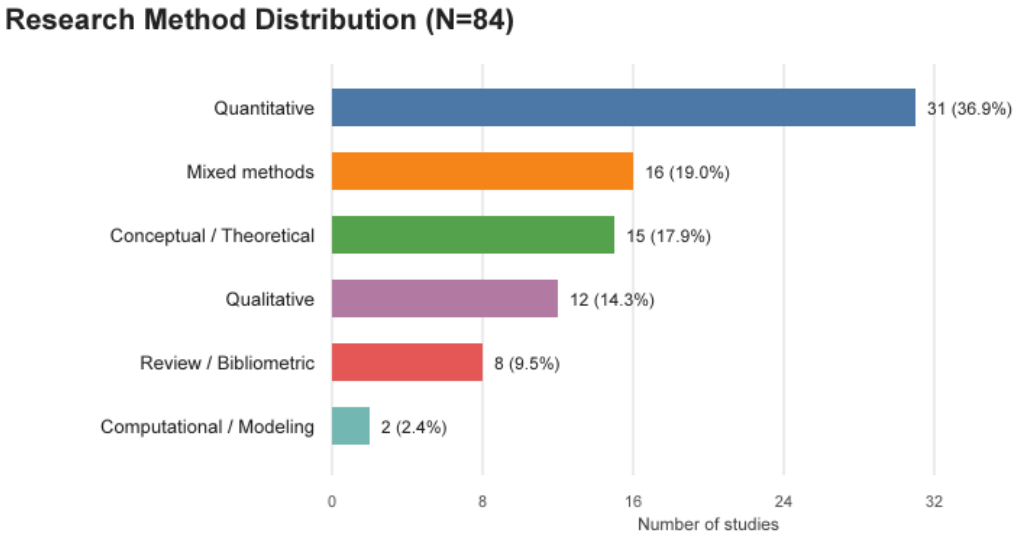

Figure 4. Distribution of research methods across included studies (N = 84).

3.2. Cross-Disciplinary Outlets and the Role Shift

The corpus spans approximately 50 SSCI-indexed outlets displaying a long-tailed distribution. The International Journal of Information Management leads with 7 articles, followed by Organizational Dynamics with 5. Several prominent journals published 3 articles each, including the Journal of Innovation & Knowledge, International Journal of Selection and Assessment, Human Resource Management, The International Journal of Human Resource Management, Personnel Review, Journal of Organizational Change Management, and European Journal of Innovation Management. Three distinct intellectual clusters emerge from this distribution: Information Systems journals, organizational behaviors and general management journals, and human resource management and applied psychology journals. Pure psychology outlets remain a small minority. A core motivation for the present review is to translate cross-disciplinary evidence into a psychologically coherent framework that occupational health and industrial-organizational psychologists can productively adopt and extend.

For studies identifying multiple AI roles, the primary role representing the study's core theoretical contribution or the highest level of anthropomorphic agency was coded to ensure mutually exclusive categories. The role distribution presented in Figure 5 offers a direct answer to RQ1 regarding role evolution. In most studies, CAI remains positioned in a utilitarian capacity. Specifically, 55 studies frame the technology strictly as a tool, and 8 as an assistant, agent, or resource.

 doi:10.20944/preprints202606.1797.v1

These instrumental conceptualizations are typically theorized through affordance theory or the job demands-resources model. This default posture is exemplified by Barba et al.'s (2025) aerospace industry case study, where large language models served explicitly as content generation engine rather than social counterpart. However, a clear theoretical turn regarding role elevation has emerged. Fifteen studies treat the CAI as a partner, teammate, or companion, an approach most strikingly visible in Le et al.'s (2025) service sector investigation of human-digital employee collaboration. Furthermore, 6 studies (7.2%) elevate the technology to the role of supervisor, coach, manager, or mentor, as observed in Terblanche's (2024) study on how AI coaching is redefining people development. Together, these elevated non-tool roles account for 34.5% of the corpus.

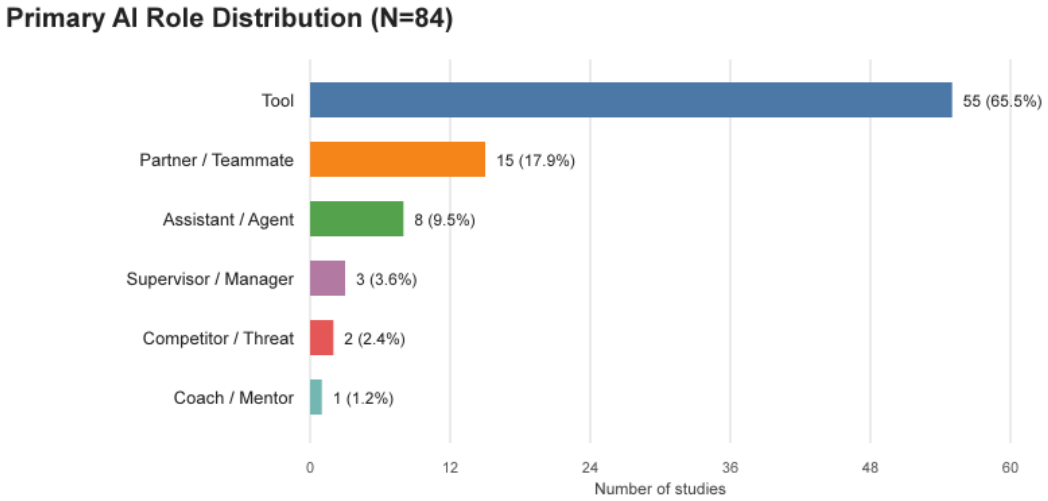

Figure 5. Distribution of CAI roles in the workplace (N = 84).

This section triangulates the empirical findings and theoretical perspectives across the 84 reviewed studies to directly address our three research questions. By integrating CASA paradigm, STS theory, and JD-R model, we map the psychological mechanisms through which CAI reshapes the employee experience.

Complementing this role shift, the data on anthropomorphism reveals significant variations in design cueing. Within the corpus (Figure 6), 46 studies (54.8% of the total) explicitly addressed or manipulated anthropomorphic characteristics. Among this specific subset, 14.3% coded the CAI with strong or high anthropomorphic cues. As detailed in the dimensions of anthropomorphism, these highly anthropomorphized systems frequently combined identity cues (e.g., formal names, avatars, and persistent personas) with emotional or empathic expressions, analogous to the paradigm case in Nguyen (2026). Weak or low cues constituted the largest category at 32.1%, typically relying on basic conversational and linguistic fluency (the most prevalent design dimension overall, n=27) without constructing a fully realized persona. Moderate cues accounted for 7.1%. The co-occurrence of role elevation and anthropomorphic cueing provides dense empirical support for the Computers as Social Actors hypothesis. As a CAI system increasingly signals human-like qualities, employees more readily recategorize it from a utilitarian tool to a social counterpart.

 doi:10.20944/preprints202606.1797.v1

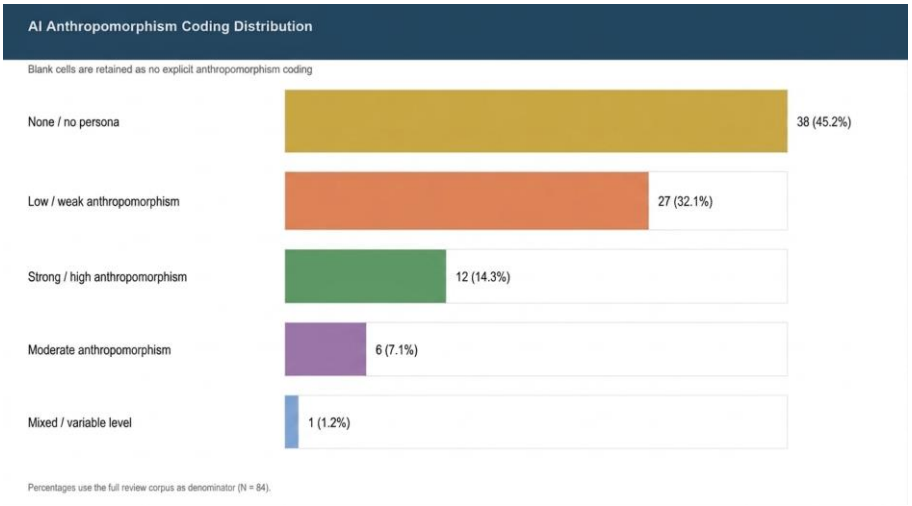

Figure 6. Distribution of CAI anthropomorphism in the workplace (N = 84).

4. Thematic Findings and Psychological Mechanisms

This section triangulates the empirical findings and theoretical perspectives across the 84 reviewed studies to directly address our three research questions. By integrating CASA paradigm, STS theory, and JD-R model, we map the psychological mechanisms through which CAI reshapes the employee experience.

4.1. From Tool to Social Actor: The Psychological Mechanisms of Anthropomorphism

The quantitative trend from Section 3.2 requires a qualitative explanation. Historically, organizational technology studies defaulted to an instrumentalist reading in which AI functions primarily as a resource to optimize tasks (Barba et al., 2025). Our corpus indicates that this reading now characterizes only about two-thirds of the field. The remaining third reflects a psychological recategorization driven by anthropomorphic cues. Consistent with Epley et al.'s (2007) three-factor theory of elicited agent knowledge, effectance motivation, and sociality motivation, conversational AI triggers anthropomorphic processing whenever linguistic fluency, persona continuity, and affective expression co-occur. In these instances, employees apply social scripts to a system they know is artificial. Pillai et al. (2024), surveying 415 service-sector employees, reported that anthropomorphic chatbot features predicted employee-experience satisfaction through perceived social presence. Nguyen (2026) demonstrated that anthropomorphic features in company-deployed generative-AI systems increased knowledge-contributing behaviors and engagement. Furthermore, Gkinko & Elbanna (2022) documented in a longitudinal qualitative study of an enterprise chatbot rollout that employees expressed hope, tolerance, and empathy towards a struggling system. These are emotional stances typically reserved for novice colleagues. Le et al. (2025) extended this observation with service-sector evidence showing that when a digital employee possesses a cohesive persona, coworkers coach its errors and advocate for its inclusion in team rituals, representing a form of anticipatory socialization directed at a machine.

he corpus also highlights non-trivial moderation of this transition. Strong anthropomorphic cues such as name, voice, stable personas, and empathic phrasing are psychologically potent but contextually sensitive. In high-stakes decision settings such as personnel selection, employees simultaneously expect machine-like neutrality and human-like accountability. This creates a double-bind that Fisher et al. (2023) identified as a source of distrust among applicants with disabilities. By contrast, weaker cues such as turn-taking and acknowledgement tokens often trigger social interpretation at a lower psychological cost and without generating the same accountability conflict. This dynamic may explain why the largest anthropomorphism category in our corpus is weak rather

 doi:10.20944/preprints202606.1797.v1

than strong, representing 36.9% versus 21.4% of the studies respectively. This distribution suggests that anthropomorphic design is not a simple intensity effect but a contextual calibration problem.

Task stakes and decisional authority further shape the role transition and moderate anthropomorphic processing. In contexts where employees retain clear decisional authority and the AI is framed as an assistive counterpart, such as customer service handoffs (Lin et al., 2024) or virtual assistant deployments (Dutta & Mishra, 2025), anthropomorphic cues foster engagement and perceived support. However, when conversational AI is interposed between employees and consequential decisions such as evaluative appraisals, hiring screens, and compensation calculations, the exact same cues can provoke a defensive psychological stance. Song et al. (2025) demonstrated this experimentally by showing that under high-stakes AI appraisals, employees with higher job complexity responded with heightened psychological empowerment, whereas those with lower job complexity responded with increased workplace cheating. This indicates that anthropomorphic framing alone cannot guarantee constructive outcomes when accountability asymmetries loom large. Role elevation therefore depends not only on interface design but also on how responsibility is distributed around the system.

Organizational legitimacy serves as a third recurring moderator in the corpus. Employees engage conversational AI more constructively when its use is openly authorized, supported by training, and embedded in accepted workflow norms. Cho et al. (2025) established this in Korean government agencies, where organizational support mediated the relationship between AI trust and employee adoption. Conversely, where AI use is tacitly permitted but lacks formal governance, employees often rely on the technology privately to meet deadlines while concealing the practice from peers and supervisors. This concealed reliance, often identified as shadow AI use, suppresses knowledge sharing and distorts learning (Chen, 2025). The transition from tool to social actor is therefore partly a design phenomenon and partly a governance phenomenon. Legitimacy operates not merely as an administrative concern but as a psychological precondition for articulating and improving the relationship between employee and machine

The theoretical significance of this shift is that it reopens the social actor default at an industrial scale. Employees frequently enter the interaction with a social-counterpart hypothesis even while fully aware they are interacting with software. This transition leads to two primary corollaries. First, findings regarding trust, betrayal, and repair in AI interactions must be theorized using interpersonal trust frameworks (McKnight et al., 2002; Rousseau et al., 1998) alongside traditional technology acceptance models. Second, qualitative differences among AI systems, ranging from names and voices to empathic phrasing, function as psychologically decisive design variables rather than mere cosmetic choices, as empirically demonstrated by Pillai et al. (2024). Ultimately, conversational AI becomes a social actor when anthropomorphic cues, task structure, and organizational legitimacy jointly encourage employees to interpret it as a socially relevant counterpart. The underlying mechanism is not technological sophistication alone, but the activation of interpersonal expectations surrounding trust, reciprocity, accountability, and support.

4.2. The Three-Dimensional Reshaping of Employee Experience

Once recategorized as a social counterpart, CAI reshapes the employee experience in psychologically distinct yet coupled ways. Drawing on the corpus evidence and extending prior schemas such as Filippelli et al.'s (2026), we distinguish three interdependent dimensions of the employee experience: cognitive, emotional-social, and career-ethical. Table 3 synthesizes the double-edged effects across these dimensions, while the subsequent analysis explains how they operate and reinforce one another.

 doi:10.20944/preprints202606.1797.v1

Table 3. The Double-Edged Effects of Conversational AI on Employee Experience.

Dimension	Bright Side	Dark Side	Representative Studies
Cognitive	Human intelligence augmentation; reduced cognitive load; problem-finding	Skill threat; prompt-engineering burden; judgement dependence	Lin et al. (2024); Brachten et al. (2020); Sabbah & Li (2025); Chen (2025); Sarkar (2026)
Emotional–Social	24/7 emotional support; burnout relief; tolerance & empathy	Social-fabric rupture; declining help-seeking	Qi et al. (2025); Gkinko & Elbanna (2022); Baygi & Huysman (2025); Li et al. (2025)
Career–Ethical	Meaningful work; bias reduction; inclusion	Algorithmic exclusion; work alienation; GenAI Loafing; workplace cheating	Callari & Puppione (2025); Dutta & Mishra (2025); Singh et al. (2025); Fisher et al. (2023); Hai et al. (2025); Saluja et al. (2025); Song et al. (2025)

At the cognitive level, conversational AI frequently functions as an augmentation resource that produces a human intelligence augmentation effect. Lin et al. (2024) demonstrated in a mixed methods study that chatbot support increased customer service agents' diagnostic accuracy and task efficiency. Similarly, Brachten et al. (2020) reported measurable reductions in cognitive load during routine inquiry handling. Freed from routine processing, employees can redirect cognitive resources toward higher-order activity such as problem-finding, as Sabbah & Li (2025) documented in human-LLM collaboration scenarios. However, this reconfiguration carries psychological cost. Chen (2025) found that skill threat perception, defined as the anticipation that accumulated domain expertise will be devalued, significantly suppressed knowledge-sharing behavior even when AI self-efficacy remained high. Furthermore, Sarkar (2026) argues that the cognitive locus shifts from executing tasks to crafting prompts and verifying probabilistic outputs, crafting a new layer of cognitive burden and decision fatigue. The cognitive experience is therefore characterized by simultaneous augmentation and vigilance.

The emotional and social evidence reveals a similar duality. CAI furnishes a form of interactional warmth unusual in organizational life by being using continuously available, non-judgement, and infinitely patient. Qi et al. (2025) found that generative AI use was positively associated with digital performance partly through reduced emotional exhaustion. Gkinko & Elbanna (2022) even documented employees extending hope, tolerance, and empathy to a struggling enterprise chatbot. While these findings validate the technology as a psychological resource, reliance on artificial systems can fray the organizational social fabric. Baygi & Huysman (2025) warned that when employees learn to consult the system in preference to colleagues, the informal lattice of help-seeking and reciprocity weakens. Li et al. (2025) corroborated this with evidence that employee collaboration with generative AI, while instrumentally useful, reduced spontaneous helping among colleagues. The critical factor governing these outcomes is whether the technology is positioned as an augmenting presence that enriches peer exchange or a substituting presence that replaces human relationships.

The career and ethical dimension profound shifts in meaningful work, fairness perception, and behavioral adaptation. Callari & Puppione (2025) identified configurations where employees successfully reconstructed meaningful work around AI-assisted practices. Dutta & Mishra (2025) showed that supportive AI virtual assistant increased perceived procedural fairness and work engagement. Singh et al. (2025) similarly demonstrated that conversational tools could enhance leaders' ability to translate inclusion strategies into daily behaviors. Nevertheless, the opacity and asymmetry of these systems produce emergent ethical risks. Fisher et al. (2023) highlighted algorithmic exclusion, arguing that AI-enabled screening systems can systematically disadvantage applicants with disabilities whose communication profiles deviate from training data norms. Hai et al. (2025) found that high-intensity digital demands created by generative AI produced work

 doi:10.20944/preprints202606.1797.v1

alienation, which subsequently increased employee expediency. Saluja et al. (2025) identified generative AI loafing, a phenomenon where employees transfer cognitive effort wholesale to the machine, leading to the atrophy of critical thinking. Additionally, Song et al. (2025) experimentally demonstrated that high-stakes AI appraisal contexts can raise employees' propensity to engage in workplace cheating. Crucially, these negative outcomes represent employee adaptations to asymmetrical collaboration conditions rather than mere technical failures.

These three dimensions do not operate independently. A key synthetic observation across the corpus is that they are psychologically coupled. A cognitive skill threat appraisal can amplify emotional career anxiety, which in turn lowers the threshold for ethical expediency. Read together, studies by Chen (2025), Hai et al. (2025), and Saluja et al. (2025) trace a recurring cascade where the perceived devaluation of expertise increases anxiety, weakens identification with work, and lowers resistance to loafing or expedient behaviors. This coupled experience structure implies that single-dimensional interventions, such as isolated AI literacy training, are insufficient. Instead, organizations require integrated governance strategies that address the cognitive, emotional, and ethical strands simultaneously.

4.3. Antecedents, Pathways, and the Integrated Model

4.3.1. The Three-Layered Antecedent Structure

Entry into deep human-AI collaboration is conditioned by a three-layered antecedent structure. At the individual layer, AI literacy, AI self-efficacy, and identity security emerge as central drivers that influence whether employees treat conversational AI as an aid or a threat. Chen (2025) reported that AI self-efficacy buffered the negative effect of skill threat on knowledge sharing, while Saluja et al. (2025) suggested that limited critical engagement can intensify over-reliance. At the organizational layer, digital trust climate, training, and perceived organizational support govern whether the use of the technology is experienced as legitimate and developmental or as imposed and risky. Cho et al. (2025) demonstrated that organizational support mediated the link between trust in AI and chatbot adoption. Conversely, Baygi & Huysman (2025) document that laissez-faire organizational stances can precipitate both system misuse and inter-employee trust erosion. At the task layer, the complexity of work and the degree of discretion retained by y employees shape the role the system is permitted to take. Large-scale recruitment screening, document generation, IT self-service, and creative composition each produce a distinct collaboration mode. These three layers jointly determine employees' initial construal of the technology and shape the downstream psychological pathway.

A further antecedent that requires explicit theoretical attention is the employee's professional identity prior to the technology encounter. Studies focusing on knowledge-intensive professions consistently report that employees with strong, stable professional identities engage the system as a tool or collaborator within a well-defined repertoire. In contrast, employees with weaker or transitional identities appear more vulnerable to the identity threat of skill devaluation (Chen, 2025). This pattern suggests that professional identity stability functions as a psychological resource. It buffers the demand-like features of the technological encounter and enables employees to process novelty without destabilization.

Importantly, these antecedents operate as configurations rather than isolated predictors. High AI literacy in a low-trust climate often produces concealed or cynical use, where employees utilize the system but protect their core work from it (Chen, 2025). High trust without sufficient literacy can produce over-reliance that eventually erodes skill. The most constructive forms of collaboration emerge when employee capability, supportive climate, and task fit develop simultaneously (Callari & Puppione, 2025). This configurational dynamic helps explain why outcomes vary so sharply across different organizational settings.

 doi:10.20944/preprints202606.1797.v1

4.3.2. Positive and Negative Outcomes

The outcomes of these interactions are distinctly double-edged. Positive outcomes include measurable performance gains (Qi et al., 2025), enhanced incremental and radical creativity (Zhang et al., 2025), psychological empowerment (Song et al., 2025), and strong procedural fairness perceptions (Dutta & Mishra, 2025; Singh et al., 2025). Negative outcomes cluster along three recuring pathways. The first is a strain pathway, where intensified digital demands yield alienation and emotional exhaustion (Hai et al., 2025). The second is a moral pathway, where alienation, combined with hindrance stressors to produce GenAI Loafing and workplace cheating (Saluja et al., 2025; Song et al., 2025). The third is a relational pathway, where the system substitutes for peer help-seeking, thereby eroding spontaneous helping and knowledge-sharing (Baygi & Huysman, 2025; Li et al., 2025). Technical limitations, such as hallucination or knowledge-based gaps, function primarily as boundary conditions that shape how these pathways unfold rather than directly constituting the psychological outcomes themselves.

The temporal unfolding of these outcomes represents a critical emerging theme. Gkinko & Elbanna's (2022) captured an early phase characterized by hope and tolerance, essentially a psychological honeymoon where employees extended generous benefit of the doubt to a struggling system. Later phases revealed a shift toward calibrated expectation and occasional disillusionment. This trajectory echoes established literature on interpersonal-trust development (Rousseau et al., 1998) and demonstrates that relationship between employees and artificial systems evolves through identifiable stages.

From a psychological mechanism standpoint, the negative pathways are unified by a common feature: they responses to structural asymmetries. CAI introduces or intensifies the asymmetry of information between humans and algorithms, the asymmetry of pace between machine output and human verification, and the asymmetry of accountability between opaque systems and transparent employees. When these asymmetries are balanced by explanation, employee discretion, and shared norms, augmentation is highly likely. When they remain unbalanced, harm typically follows.

4.3.3. Integrated Conceptual Model

Figure 7 synthesizes the corpus evidence into an integrated conceptual model. The model identifies a logical progression where three-layered antecedents (individual, organizational, and task) feed into the system's role profile and anthropomorphic features. Through the activation of social scripts, the technology then engages the cognitive, emotional-social, and career-ethical dimensions of the employee experience. Operating through the dual pathways of resource gain and demand cost, this process generates either enabling or dark-side outcomes. The model is deliberately psychologically anchored; its focal point is the employee's internal psychological experience rather than the firm's operational efficiency.

 doi:10.20944/preprints202606.1797.v1

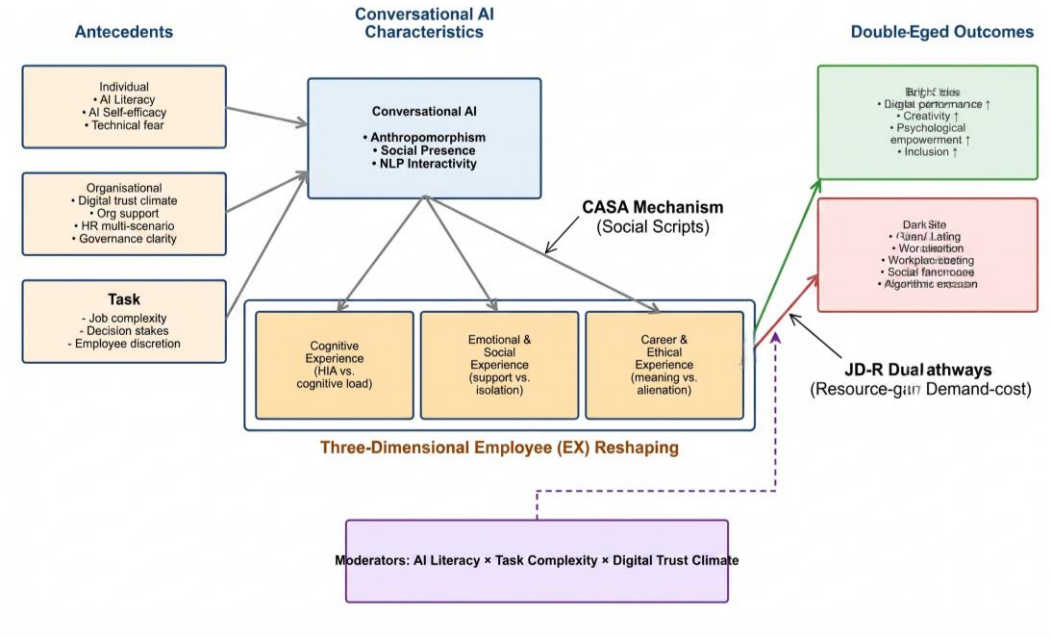

Figure 7. Integrated conceptual model of CAI reshaping employee experience.

4.3.4. Research-Question Cross-Mapping

Table 4 maps each research question to its core theoretical lens, the corresponding result section, and the key evidence pattern, making the analytical logic of the synthesis fully transparent.

Table 4. Cross-Mapping of Research Questions, Theoretical Lenses, and Result Sections.

RQ	Theoretical Lens	Result Section	Key Evidence
RQ1 Role evolution	CASA and Anthropomorphism	Section 3.2 and Section 4.1	CAI increasingly shifts from tool to partner, evaluator, or social actor under anthropomorphic cueing and supportive governance
RQ2 Experience reshaping	STS and JD-R	Section4.2 and Table 3	The technology reshapes employee experience across cognitive, emotional-social, and career-ethical dimensions, demonstrating coupled bright-side and dark-side effects.
RQ3 Antecedents and outcomes	JD-R and STS	Section 4.3 and Figure 5	Individual, organizational, and task antecedents channel employees toward distinct pathways of augmentation, strain, moral adaptation, or social-fabric erosion.

5. Discussion

5.1. Dialogue with Prior Syntheses

Our findings engage substantively with anchor reviews in the field, indicating that the ground has shifted conceptually. For example, Vrontis et al. (2022) and Kambur & Yildirim (2023) mapped artificial intelligence primarily as an enabling infrastructure for human resource management processes, often celebrating intelligent efficiency. Our corpus, reflecting post-ChatGPT realities, reveals for roughly a third of studies, CAI is no longer infrastructure but an actor. Consequently, efficiency gains are sometimes purchased with social-fabric erosion (Baygi & Huysman, 2025; Li et al., 2025), as documented by Baygi & Huysman (2025). This necessitates an internalized social-cost audit within the intelligent management discourse. Furthermore, Bankins et al. (2024) called for

 doi:10.20944/preprints202606.1797.v1

employee-level psychological research, which our synthesis operationalizes by providing a three-dimensional coordinate system that reveals emergent cross-dimensional coupling. While Budhwar et al. (2023) and Dwivedi et al. (2023) correctly forecast opportunities and dangers in opinion pieces, our empirical synthesis specifies these mechanisms, documenting unforeseen phenomena such as generative AI loafing and work alienation. Finally, while Filippelli et al.'s (2026) three-dimensional well-being schema informed our early coding, we propose extending it to elevate career-ethical concerns to a peer dimension rather than a residual category, given its prominence in our corpus.

Taken together, these dialogues suggest the field is undergoing a theoretical reorientation. The relevant question is no longer whether employees adopt the technology, but how they interpret, negotiate, and morally accommodate a conversational system that increasingly resembles a workplace counterpart. A purely technology acceptance framing is now inadequate. It must be supplemented by theories of social categorization, interpersonal trust, moral disengagement, and resource conservation. Crucially, while previous anchor reviews were written from distinct disciplinary bases such as information systems or innovation management, the evidence suggests no single discipline can fully encompass this phenomenon. Psychology occupies a privileged analytical position because the core mechanisms, such as construal, attribution, trust, and moral disengagement, are fundamentally psychological rather than strictly organizational or technological.

5.2. Theoretical Contributions

This synthesis contributes to the psychology of technology at work in four ways. First, its re-centers the field on psychological mechanism by demonstrating that workplace impact of CAI is mediated by how employees construe the system, a construal systematically predicted by anthropomorphic design cues (Nguyen, 2026; Pillai et al., 2024). This positions the computers as Social Actors paradigm as an essential import into organization behavior theorizing. Second, it develops an integrated, three-dimensional coordinate system for the employee experience, detailing empirically demonstrated cross-dimensional coupling and thereby advancing Bankins et al.'s (2024) conceptual agenda into testable propositions. Third, it consolidates dispersed evidence on dark-side outcomes such as work alienation (Hai et al., 2025), GenAI Loafing (Saluja et al., 2025), workplace cheating (Song et al., 2025), and social-fabric rupture (Baygi & Huysman, 2025), delineating these employee-side behavioral adaptations from technology-end boundary conditions like system hallucinations. Fourth, by fixing the unit of analysis strictly at the direct human party of the human-AI interaction, we establish a methodological discipline that prevents disparate outcomes from being conflated.

5.3. Practical Implications

Practical translation of these findings should proceed with psychological design logic rather than efficiency logic alone. Organizational efforts to "roll out" CAI framed solely around productivity targets tend to accelerate dark-side pathways because they intensify structural asymmetries without creating psychological slack for sense-making. Three practice principles follow from this review. First, organizations should design for system-plus-peer collaborative architectures rather than individualistic ones, ensuring that task engagement preserves opportunities for spontaneous collegial interaction and prevents the displacement of peer exchange (Baygi & Huysman, 2025). Second, systems must be deployed strategically to counter algorithmic exclusion. This requires auditing high-stakes evaluative tools for disparate impacts on applicants with disabilities (Fisher et al., 2023) or non-standard communication profiles, ensuring technology enhances rather than merely preserves inclusion (Singh et al.,2025). Third, organizations should institute social-cost audits alongside standard performance metrics to routinely track help-seeking frequency, cross-team consultation, and knowledge sharing. Furthermore, governance matters deeply. Employees respond more constructively when responsibility for verification is clear and when using CAI is explicitly authorized rather than covertly practiced.

 doi:10.20944/preprints202606.1797.v1

5.4. Limitations

Six limitations warrant acknowledgement. The literature search was restricted to Web of Science Core Collection and English-language publications, meaning relevant studies indexed elsewhere or published in other languages may have been missed. The review protocol was not pre-registered, an increasingly important norm in psychological research. Quality appraisal inevitably involved subjective elements, particularly when evaluating conceptual papers. Coding was performed exclusively by the authors; while inter-rater reliability was strong, external third-party coders would further enhance independence. Finally, most primary studies in the corpus rely on cross-sectional designs, which constrains inferences regarding the temporal dynamics of trust, anthropomorphic attachment, and alienation.

6. Conclusions and Forward Look

Amid a wave of conversational AI rapidly reorganizing everyday work, the deep question for the psychology of work is not whether artificial systems will replace employees, but what kind of colleagues humans are becoming to the machines and, through them, to one another. Our synthesis of 84 empirical studies demonstrates that through anthropomorphic cues and advanced interactivity, these systems have already begun to renegotiate the social boundaries of the workplace. Three critical integrative insights emerge. First, cognitive, emotional-social, and career-ethical dimensions are tightly coupled; perceived skill threats routinely amplify anxiety, which sequentially lowers resistance to moral expediency (Chen, 2025; Hai et al., 2025; Saluja et al.,2025). Second, efficiency-first deployments risk silently depleting the informal mutual-help networks upon which organizational resilience relies (Baygi & Huysman,2025; Li et al., 2025). Third, emergent dark-side phenomena such as generative AI loafing and workplace cheating are not accidents, but systematic products of asymmetric human-AI arrangements and high digital demands (Fisher et al., 2023; Saluja et al., 2025; Song et al., 2025).

These insights rest on a rapidly evolving evidence base that demands vigorous future research across several frontiers. The field urgently requires longitudinal designs using diary, experience sampling, and social network methods to trace how human-AI trust is formed, violated, and repaired across the collaboration lifecycle. Additionally, research must investigate work-family spillover to determine whether the reduction of workplace emotional labor translates into family enrichment or generates persistent digital anxiety. Algorithmic fairness and cross-cultural comparisons represent further critical frontiers. Researchers must specifically study the experiences of older workers, applicants with disabilities, and non-mainstream language users, while also examining how differing institutional and collectivistic-individualistic contexts shape psychological pathways beyond the predominantly Western and Chinese samples currently dominating the literature.

A cross-cutting frontier concerns measurement and individual differences. The construct of anthropomorphism, alongside conversational AI-specific trust and digital work expediency, is currently operationalized with significant heterogeneity. The psychological validity of cross-study comparisons depends entirely on the development of shared, robust psychometric instruments adapted for occupational settings. Furthermore, future inquiry must move beyond average effects to consider how individual traits, such as an inherent tendency toward anthropomorphism or the need for decisional autonomy, moderate employee responses to artificial systems.

The wave of conversational artificial intelligence will not slow. How it reshapes the psychology of work depends on whether the academic community can study it with the theoretical discipline and social imagination the moment demands. The central empirical lesson of this synthesis is that these systems are already producing both flourishing and harm in measurable proportions. The conceptual lesson is that these outcomes are patterned by whether organizations and researchers recognize conversational AI as a new class of social counterpart and accept the psychological responsibilities such recognition entails. By relocating the research conversation from a narrow focus

 doi:10.20944/preprints202606.1797.v1

on productivity toward a comprehensive account of experience reshaping, this review offers a conceptual and empirical foundation for that collective work.

Author Contributions: Conceptualization and methodology, L.L. and X.S.; validation and formal analysis, R.J.Z. and X.S.; investigation, data curation and visualization, L.L.; resources, supervision, project administration and funding acquisition, L.L. and X.S.; writing—original draft preparation, L.L. and R.J.Z.; writing—review and editing, all authors. 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: The original contributions presented in this study are included in the article/Supplementary Materials. Further enquiries can be directed to the corresponding author.

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

References

Alherimi, N., Abdulmaksoud, S., Ahmed, V., & Bahroun, Z. (2025). A Systematic Literature Review of Artificial Intelligence Advancements in Green Human Resource Management. *Sustainability, 17*(22), 10283. https://doi.org/10.3390/su172210283

Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology, 22*(3), 273–285.

Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2024). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *Journal of Organizational Behavior, 45*(2), 159–182.

Barba, G., Corallo, A., Lazoi, M., & Lezzi, M. (2025). Large language models for competence-based HRM: A case study in the aerospace industry. *Journal of Innovation & Knowledge, 10*(4), 100780.

Baygi, R. M., & Huysman, M. (2025). Generative AI and the social fabric of organizations. *Strategic Organization*. Advance online publication.

Brachten, F., Brünker, F., Frick, N. R. J., Ross, B., & Stieglitz, S. (2020). On the ability of virtual agents to decrease cognitive load: An experimental study. *Information Systems and e-Business Management, 18*(2), 187–207.

Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G. J., Beltran, J. R., ... Varma, A. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal, 33*(3), 606–659.

Callari, T. C., & Puppione, L. (2025). Meaningful work as shaped by employee work practices in human–AI collaborative environments: A qualitative exploration through ideal types. *Qualitative Research in Organizations and Management, 20*(2), 245–269.

Chen, Q. Q. (2025). Enhancing knowledge sharing in generative AI integration: The impact of AI self-efficacy and skill threat perceptions. *Journal of Knowledge Management, 29*(6), 1532–1551.

Cherns, A. (1976). The principles of sociotechnical design. *Human Relations, 29*(8), 783–792.

Cho, S., Hur, J.-Y., & Kim, D. (2025). Bridging trust in AI and its adoption: The role of organizational support in AI chatbot implementation in Korean government agencies. *Government Information Quarterly, 42*(1), 101987.

Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands–resources model of burnout. *Journal of Applied Psychology, 86*(3), 499–512.

Dutta, D., & Mishra, S. K. (2025). Artificial intelligence-based virtual assistant and employee engagement: An empirical investigation. *Personnel Review, 54*(3), 913–934.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and

 doi:10.20944/preprints202606.1797.v1

implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642.

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886.

Filippelli, S., Popescu, I. A., Verteramo, S., Corvello, V., & Tani, M. (2026). Generative AI and employee well-being: Exploring the emotional, social, and cognitive impacts of adoption. *Journal of Innovation & Knowledge, 11*(1), 100812.

Fisher, S. L., Connelly, C. E., & Bonaccio, S. (2023). Reactions of applicants with disabilities to technology-enabled recruitment and selection: A research agenda. *International Journal of Selection and Assessment, 31*(3), 366–381.

Gkinko, L., & Elbanna, A. (2022). Hope, tolerance and empathy: Employees' emotions when using an AI-enabled chatbot in a digitalised workplace. *Information Technology & People, 35*(6), 1714–1736.

Hai, S., Long, T., Honora, A., Japutra, A., & Guo, T. (2025). The dark side of employee–generative AI collaboration in the workplace: An investigation on work alienation and employee expediency. *International Journal of Information Management, 82*, 102855.

Hobfoll, S. E. (1989). Conservation of resources: A new attempt at conceptualizing stress. American Psychologist, 44(3), 513–524.

Hughes, L., Davies, F., Li, K., Gunaratnege, S. M., Malik, T., & Dwivedi, Y. K. (2026). Beyond the hype: Organisational adoption of generative AI through the lens of the TOE framework—A mixed method investigation. *International Journal of Information Management, 84*, 102912.

Jatobá, M. N., Ferreira, J. J., Fernandes, P. O., & Teixeira, J. P. (2023). Intelligent human resources for the adoption of artificial intelligence: a systematic literature review. *Journal of Organizational Change Management, 36*(7), 1099-1124.

Kambur, E., & Yildirim, T. (2023). From traditional to smart human resources management. *International Journal of Manpower, 44*(3), 422–452.

Kekez, I., Lauwaert, L., & Ređep, N. B. (2025). Is artificial intelligence (AI) research biased and conceptually vague? A systematic review of research on bias and discrimination in the context of using AI in human resource management. *Technology in Society, 81*, 102818.

Le, K. B. Q., Sajtos, L., Kunz, W. H., & Fernandez, K. V. (2025). The future of work: Understanding the effectiveness of collaboration between human and digital employees in service. *Journal of Service Research, 28*(3), 412–431.

Li, J.-M., Wu, H.-Y., Zhang, R.-X., & Wu, T.-J. (2025). How employee–generative AI collaboration affects employees' work and family outcomes? The relationship instrumentality perspective. *The International Journal of Human Resource Management, 36*(10), 1782–1810.

Lin, X., Wang, X., Shao, B., & Taylor, J. (2024). How chatbots augment human intelligence in customer services: A mixed-methods study. *Journal of Management Information Systems, 41*(4), 1016–1041.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research, 13*(3), 334–359.

Microsoft. (2024). *Work Trend Index Annual Report: AI at work is here. Now comes the hard part*. Microsoft & LinkedIn.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103.

Nguyen, M. (2026). Enhancing employee engagement and knowledge collecting: Impact of anthropomorphic features in company generative AI systems. *Knowledge Management Research & Practice, 24*(1), 95–112.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*:n71.

Pillai, R., Ghanghorkar, Y., Sivathanu, B., Algharabat, R., & Rana, N. P. (2024). Adoption of artificial intelligence (AI) based employee experience (EEX) chatbots. *Information Technology & People, 37*(2), 449–478.

Qi, Z., Zhang, X., Ma, L., & Wang, G. (2025). Generative artificial intelligence use and employees' digital performance. *Management Decision, 63*(7), 1892–1915.

 doi:10.20944/preprints202606.1797.v1

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge, UK, 10(10), 19-36.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393–404.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78.

Sabbah, J., & Li, F. (2025). When humans and large language models collaborate, problem-finding illuminates. *Innovation, 28*(2), 374–407..

Saluja, S., Sinha, S., & Goel, S. (2025). Loafing in the era of generative AI. *Organizational Dynamics, 54*(3), 101101.

Sarkar, A. (2026). From AI hype to workflow reality: A strategic framework for integrating generative AI across organizational functions. *Organizational Dynamics, 55*(1), 101202.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.

Singh, V., Rivin, J. M., Wagoner, H. P. v., Keplinger, K., & Barbuto, J. (2025). Chatting towards inclusivity: A digital approach to inclusion action plans and leader development. *Human Resource Management, 64*(4), 621–640.

Singh, V., Rivin, J. M., Wagoner, H. P. v., Keplinger, K., & Barbuto, J. (2025). Chatting towards inclusivity: A digital approach to inclusion action plans and leader development. *Human Resource Management, 64*(4), 621–640.

Song, X., Shen, J., & Yang, Z. (2025). The double-edged impact of AI appraisals on workplace cheating: The roles of psychological empowerment and job complexity. *Journal of Business Ethics, 198*(2), 355–378.

Tajfel, H., & Turner, J. C. (1986). *The social identity theory of intergroup behavior.* In J. T. Jost & J. Sidanius (Eds.), Political psychology: Key readings (pp. 276–293). Psychology Press.

Terblanche, N. H. D. (2024). Artificial intelligence (AI) coaching: Redefining people development and organizational performance. *The Journal of Applied Behavioral Science, 60*(4), 631–638.

Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting. *Human Relations, 4*(1), 3–38.

Úbeda-García, M., Marco-Lajara, B., Zaragoza-Sáez, P. C., & Poveda-Pareja, E. (2025). Artificial intelligence, knowledge and human resource management: A systematic literature review of theoretical tensions and strategic implications. *Journal of Innovation & Knowledge, 10*(6), 100809.

Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2022). Artificial intelligence, robotics, advanced technologies and human resource management: A systematic review. *The International Journal of Human Resource Management, 33*(6), 1237–1266.

Zhang, X., Yu, P., & Ma, L. (2025). How and when generative AI use affects employee incremental and radical creativity: An empirical study in China. *European Journal of Innovation Management, 28*(4), 891–912.