The Evolution Mechanism of Human-Machine Trust Driven by the Dual Dimensions of Instrumentality and Value

Authors: Song Yu1 Hu Xiaoran2

Date: 2026-05-21T16:06:06+00:00

Abstract

随着人工智能技术的迅猛发展，人机关系已成为人们组织生活中的重要关系。人机信任是人机关系的核心，关乎人机交互的成败。如何建构人机信任，把握人机关系发展规律，在人机融合中实现优势互补，是人机信任研究领域的核心问题。本研究围绕上述问题，立足人机二元互动，探讨人机信任随时间变化的演化过程和机制。首先，本研究基于技术伦理视角，明确人机信任的内涵，提出工具信任和价值信任的两维度人机信任模型，并在此基础上编制人机信任测量量表。接着，采用动态发展视角，探寻工具信任与价值信任在时空情境耦合作用下时序演化的一般特征，打开人机信任动态演化的“黑箱”。最后，以人机协作视角为突破口，探讨工具信任和价值信任对个体创造力的差异化赋能机制，辩证分析数智时代的人机关系发展变化，为在数智时代人机信任如何塑造个体核心优势提供见解。

Full Text

The Evolution Mechanism of Human-Machine Trust Driven by the Dual Dimensions of Instrumentality and Value

Song Yu1 Hu Xiaoran2

(1 School of Economics and Management, Southeast University, Nanjing 211189)

(2 Department of Management, London School of Economics and Political Science, London WC2A 2AE, UK)

Abstract With the rapid development of artificial intelligence technologies, human-machine relationships have become an important form of relationship in people’ s organizational lives. Human-machine trust is the core of human-machine relationships and is closely related to the success or failure of human-machine interaction. How to build human-machine trust, grasp the developmental

laws of human-machine relationships, and achieve complementary advantages in human-machine integration are core issues in the field of human-machine trust research. Centering on these issues and grounded in dyadic human-machine interaction, this study explores the evolutionary process and mechanisms through which human-machine trust changes over time. First, from the perspective of technology ethics, this study clarifies the connotation of human-machine trust, proposes a two-dimensional model of human-machine trust comprising instrumental trust and value trust, and, on this basis, develops a scale for measuring human-machine trust. Next, adopting a dynamic developmental perspective, it investigates the general characteristics of the temporal evolution of instrumental trust and value trust under the coupled effects of spatiotemporal contexts, thereby opening the "black box" of the dynamic evolution of human-machine trust. Finally, taking the perspective of human-machine collaboration as a point of departure, it explores the differentiated enabling mechanisms through which instrumental trust and value trust affect individual creativity, dialectically analyzes the development and transformation of human-machine relationships in the digital-intelligence era, and offers insights into how human-machine trust can shape individuals' core advantages in the digital-intelligence era.

Keywords human-machine trust; instrumental trust; value trust; human-machine relationship; dynamic mechanism

Received: 2025-11-14
* Supported by the National Natural Science Foundation of China (72501059).
Corresponding author: Song Yu, E-mail: songyu897@hotmail.com

1 Problem Statement

With the vigorous development of technologies such as generative artificial intelligence, deep learning, and big data, the digital-intelligent era is arriving at an accelerated pace (Xu Wei et al., 2024; Zhang Zhixue et al., 2024). Against the current backdrop of the state' s strong efforts to promote the deep integration of the digital economy and the real economy, many enterprises are actively embracing intelligent technologies—artificial intelligence (AI)—and investing substantial resources in intelligent upgrading and digital transformation, in the expectation of bringing transformation and innovation to production, operations, management, and services. In this context, individuals' interaction with AI in work tasks has become a normal feature of individual work in the digital-intelligent era. As the key to intelligent interaction between humans and AI, human-AI trust is not only an important prerequisite for individuals' willingness to interact, but also a long-term guarantee for achieving effective interaction (Qi Yue et al., 2024; Dang & Li, 2026; Glikson & Woolley, 2020). Therefore, in work settings, how individual employees construct human-AI trust, grasp the developmental patterns of human-AI relationships, and realize complementary advantages in human-AI integration has become a topic of great concern in both academia

and practice (Xie Xiaoyun et al., 2021; Anthony et al., 2023; Korsgaard et al., 2025).

However, in the face of this urgent practical need, existing research on human-AI trust in the field of organizational management shows a certain degree of theoretical lag, and several key issues remain to be explored.

First, what is the structural connotation of human-AI trust? This question is the logical starting point for research on human-AI trust. Human-AI trust focuses on the interactive relationship between humans and AI, which differs from the interpersonal interactive relationships traditionally emphasized in management and psychology. Yet most current research on human-AI trust directly borrows from studies of interpersonal trust and introduces the classification framework of cognitive trust and affective trust directly into human-AI trust (e.g., Gkinko & Elbanna, 2023; Glikson & Woolley, 2020; Qin et al., 2024; Vanneste & Puranam, 2025), causing existing theories to face adaptation difficulties in AI contexts. On the one hand, the distinctive characteristics of human-AI relationships as opposed to interpersonal relationships have not received sufficient attention. As a non-human agent, AI operates on the basis of data and algorithms and lacks genuine emotional expression and social intentions in social activities (Pentina et al., 2023; Perry, 2023; Wang et al., in press). Directly transferring research on interpersonal interaction to the domain of human-AI interaction makes it difficult to truly reflect the unique logic and psychological foundations of human-AI trust, thereby greatly weakening the guiding role of theoretical research for management practice. On the other hand, the dichotomy between cognitive trust and affective trust remains subject to theoretical controversy in the field of interpersonal trust because of unclear boundaries and measurement ambiguity (Legood et al., 2021; Legood et al., 2023; Tomlinson et al., 2020; van Knippenberg, 2018). After this classification framework is introduced into AI contexts, the lack of clear boundaries and the resulting measurement confusion further amplify the points of contention, limiting its explanatory power in AI contexts (Valori et al., 2026; Wang & Ding, 2024). It can be seen that current research has not yet effectively defined the structural connotation of human-AI trust from the perspective of the uniqueness of human-AI interaction. This limitation not only reflects insufficient research on human-AI trust at the conceptual level, but also directly constrains the development of dynamic research on human-AI trust and the exploration of deeper mechanisms through which human-AI trust enables outcomes.

Second, what are the evolutionary trajectories of different types of human-AI trust across time and space? Trust is established on the basis of individuals' cognition and past experience, and it adjusts as time and space change (Dang & Li, 2026; de Visser et al., 2020; Hoff & Bashir, 2015). Existing research on dynamic trust mainly uses undertrust and overtrust to describe the extent to which trust deviates from an appropriate level (Huang Xinyu, Li Pan, 2024; Hoff & Bashir, 2015; Lee & See, 2004), emphasizing a quantitative description of the matching relationship between the level of trust and AI' s actual capabilities.

However,

Explaining the mechanism of the dynamic development of trust solely from the perspective of trust intensity has certain limitations. On the one hand, appropriate trust involves not only the intensity of trust, but also the structure of trust—that is, the dimensions on which individuals trust AI (for example, emphasizing AI' s functionality or attending to AI' s ethical norms). Different trust structures correspond to different psychological foundations and behavioral consequences (Lee & See, 2004; Malle & Ullman, 2021; Vuori et al., 2026). On the other hand, human–AI trust is embedded in a temporal sequence; that is, at different stages of interaction, individuals should assign AI trust with different structures and at different levels (Lewis et al., 2018; Skjuve et al., 2021). In other words, human–AI trust has reasonable levels of structural differentiation across different developmental stages. Therefore, the dynamic development of trust is a multidimensional, coordinated process encompassing degree, structure, and temporality. However, because existing research has not yet effectively identified the differences within the internal structure of human–AI trust, current research on the differentiated development of different types of human–AI trust has been impeded. Deepening research on the dynamics of trust can provide more explanatory and predictive theoretical support for understanding and optimizing the development of human–AI relationships. It not only helps individuals understand the logic of changes in their own trust during human–AI interaction, leverage AI effectiveness, and improve the quality of human–AI interaction; it can also help organizational managers anticipate the direction in which trust relationships evolve, implement effective interventions at critical junctures, and reduce the likelihood that investments in intelligent technologies will face risks arising from trust failure.

Third, how do different types of human–AI trust influence human–AI collaboration and enable the realization of complementary advantages in human–AI integration? As a bridge toward efficient human–AI relationships, the construction of human–AI trust is not an endpoint, nor is its significance limited to promoting AI use. What is more critical is to explore how individuals can use AI effectively (Chen Hui, Feng Chao, in press; Xu Hui et al., 2025; Lu & Yan, in press)—that is, how different types of human–AI trust are transformed into efficient modes of cooperation and substantive increments in individual capability. Creativity is an important manifestation of individuals' core competitiveness in the digital-intelligence era (Boussioux et al., 2024; Cheng & Zhang, 2025). However, because existing research has failed to effectively distinguish trust structures, current studies on the effects of human–AI trust mainly focus on individuals' willingness to use AI and performance improvement (Afroogh et al., 2024; Dang & Li, 2026; Ng & Zhang, 2025), and lack systematic investigation of the deeper enabling mechanisms of human–AI trust—namely, how different types of human–AI trust affect individuals' understanding and use of AI, and thereby influence individual creativity. An in-depth exploration of this issue will not only help respond to concerns that "AI leads to the degradation of human capabilities" (Gillespie et al., 2025; Korsgaard et al., 2025; Natali et al., 2025),

but also dialectically reveal the core role of human–AI trust in promoting the enhancement rather than the simple replacement of human capabilities. It can further provide a clear and reliable practical pathway for how individuals, in the digital-intelligence era, can construct differentiated trust, transform technological power into irreplaceable human advantages, and realize complementary advantages through human–AI integration.

In short, against the backdrop of the digital-intelligence era profoundly reshaping individuals' ways of working, it is of great theoretical value and practical significance to take the workplace as the setting, individual employees as the core unit of analysis, and to re-examine and construct a model of human–AI trust that accords with the essence of human–AI interaction; on this basis, it is also crucial to systematically reveal the laws of its dynamic evolution and the pathways through which its value is manifested. This not only helps break through existing research bottlenecks, but also constitutes a practical response to how individual employees can achieve self-adaptation and capability enhancement amid the wave of digital intelligence. By clarifying how individuals cultivate, maintain, and make good use of human–AI trust relationships in workplace interactions, and by revealing the internal logic through which the collaborative advantages of "AI + human" are generated, this article aims to provide systematic and forward-looking theoretical insights and practical guidance for individuals to build core competitiveness in an increasingly intense digital-intelligence competitive environment.

2 Literature Review

Relative to the pace of AI' s own development, research on human–AI trust in the field of organizational management has begun somewhat later (Cabiddu et al., 2022; Glikson & Woolley, 2020). Existing studies mainly focus on examining the types, influencing factors, and effects of human–AI trust.

2.1 Types of Human–AI Trust

The discussion of types of human–AI trust in organizational management has mainly been transferred from research on interpersonal trust. At present, the most common approach is to divide human–AI trust into cognitive trust and affective trust (Glikson & Woolley, 2020; Komiak & Benbasat, 2006). This classification is based primarily on McAllister' s (1995) distinction between cognitive trust and affective trust in interpersonal trust. In recent years, scholars have debated the distinction between interpersonal cognitive trust and affective trust, arguing that their definitions and measurements cannot accurately differentiate the two (Legood et al., 2021; Legood et al., 2023; Tomlinson et al., 2020; van Knippenberg, 2018). Affective trust, in particular, has been the focus of controversy. It is defined as an affective attitude generated by the trustor on the basis of the emotional bond between the trustor and the interaction target (McAllister, 1995). van Knippenberg (2018) argues that the definition of affective trust reflects the trustor' s evaluation of the trust relationship between the

two parties, and in essence also falls within the cognitive domain.

The controversy over the cognitive-affective dichotomy that originated in the field of interpersonal trust has inevitably affected research on human-AI trust, mainly in terms of measurement. At present, the measurement of cognitive trust and affective trust in human-AI relations is mainly based on three characteristics by which AI is perceived as trustworthy: ability, integrity, and benevolence. Some studies hold that cognitive trust represents individuals' recognition of AI' s ability and integrity, whereas affective trust reflects individuals' judgment of AI' s benevolence (Komiak & Benbasat, 2006); some studies hold that cognitive trust reflects individuals'evaluation of AI's ability, whereas affective trust focuses on integrity and benevolence (Choung et al., 2023; Hu et al., 2021); still other studies argue that cognitive trust should include the three aspects of ability, integrity, and benevolence, whereas affective trust reflects individuals' emotional experience of AI (Wang et al., 2016).

It can be seen that existing research on cognitive trust and affective trust in human-AI relations suffers from problems of unclear conceptual definitions and confused measurement. To facilitate the development of subsequent research and dialogue on human-AI trust, and to deeply explore the differentiated development mechanisms and enabling mechanisms of different types of trust, it is necessary to restate the internal structure of human-AI trust and to redevelop its scales.

2.2 Influencing Factors and Effects of Human-AI Trust

Existing research on the influencing factors of human-AI trust is relatively extensive and can be summarized into three categories: individual characteristics, AI characteristics, and environmental factors (Kaplan et al., 2023; Schaefer et al., 2016). Individual factors include people's AI awareness (Yin et al., 2024), AI self-efficacy (Dong et al., 2025), work experience (Wang et al., 2024), personality traits (Bawack et al., 2021; Huo et al., 2022), as well as demographic characteristics such as gender, age, educational attainment, and social class (Oksanen et al., 2020). For example, employees with high AI awareness and AI self-efficacy have greater tolerance for AI, hold more optimistic estimates of the prospects for AI applications, and are more willing to trust AI and engage in human-AI collaboration (Yin et al., 2024). Similarly, individuals with lower social class and educational levels have lower familiarity with and mastery of new technologies (Oksanen et al., 2020), and older employees are usually accompanied by declines in cognitive ability (Dutta et al., 2023). Because of their lower capacity for acceptance, this group generally finds it more difficult

establish a relationship of trust with AI.

AI characteristics mainly include the transparency and reliability of AI technology, as well as the embodiment and anthropomorphism of its physical appearance (Glikson & Woolley, 2020). AI transparency not only helps reduce people' s algorithmic biases and gain trust, but also helps employees and organizations

trace and clarify responsibilities and obligations, enabling employees to understand the decisions and recommendations made by AI (Lehmann et al., 2022). AI reliability involves data security and privacy, as well as the consistency and stability of behavioral performance. Studies show that in human–AI collaboration, if AI makes more than three errors, or if AI' s performance is inconsistent over time, employees' trust in it will decrease significantly (Hu et al., 2021). Regarding the physical appearance attributes of AI, research has found that robots with embodied forms and a high degree of anthropomorphism are more likely to elicit people' s favorable impressions and trust (Yam et al., 2021). However, excessively anthropomorphized robots can also easily go too far and produce an "uncanny valley" effect (Wang Haizhong et al., 2021; Wykowska, 2021).

Environmental factors include national culture (Chi et al., 2023), organizational reputation (Hengstler et al., 2016), organizational climate (Chowdhury et al., 2023), leadership style (Yin et al., 2024), and so on. For example, Chi et al. (2023) and Gillespie et al. (2023) found that individuals from highly collectivistic cultures have higher trust in AI. An organizational climate characterized by tolerance, support, and encouragement toward employees' use of AI helps employees establish trust in AI (Chowdhury et al., 2023). A good organizational reputation is also conducive to the establishment of human–AI trust, and individuals are more willing to trust AI systems from organizations with good reputations (Hengstler et al., 2016). Employees under a transformational leadership style are more willing to trust and use AI to complete work tasks (Yin et al., 2024).

Existing research on the effects of human–AI trust can generally be classified into two types: one type emphasizes the facilitative function of trust, while the other emphasizes the calibration and appropriateness of trust. First, one stream of research starts from a relatively static and linear perspective, emphasizing the facilitative function of human–AI trust and focusing on its effects on willingness to use AI, continued-use behavior, and performance. Relevant studies generally find that the higher the level of trust, the more individuals tend to adopt and continue using AI (Chatterjee et al., 2021; Song & Lin, 2024; Suseno et al., 2022). Moreover, trusting AI helps increase work engagement (Chandra et al., 2022; Marikyan et al., 2022), improve individual performance (Chowdhury et al., 2022; Li & Zhou, 2025), and promote career development and occupational well-being (Kong et al., 2023; Salah et al., 2024). In this type of research, human–AI trust is usually regarded as a positive psychological resource, and its mechanism of action displays a typical linear facilitation logic: the higher the trust, the better the positive effects (Dang & Li, 2026; Ng & Zhang, 2025).

However, with the widespread application of AI in high-risk, complex decision-making, and highly automated scenarios, scholars have gradually recognized the theoretical and practical limitations of simply understanding human–AI trust as "the higher, the better." Therefore, in recent years, another stream of research has begun to emphasize the importance of appropriate trust, arguing that human–AI trust is not unidirectionally better when higher, but should re-

main matched with AI' s actual capability level (Huang Xinyu & Li Ye, 2024; Hoff & Bashir, 2015; Lee & See, 2004). Specifically, excessive trust may lead individuals to become overly dependent on AI, weaken supervision and critical judgment, trigger cognitive disengagement and responsibility transfer, and even produce catastrophic consequences in complex or high-risk situations (Ding et al., 2026; Küper & Krämer, 2025; Ullrich et al., 2021). By contrast, insufficient trust can likewise bring negative effects. Insufficient trust may cause individuals to maintain excessive vigilance toward AI outputs, increase false alarms and repeated checking behavior, reduce AI adoption rates and collaboration efficiency, and even prevent AI' s technical potential from being fully realized, causing losses (Wang Xinye et al., 2017; Ayoub et

al., 2022; Ding et al., 2026).

From the above analysis, it can be seen that, although existing studies have provided rich and valuable insights into the antecedent variables of human-AI trust and have recognized that trust calibration triggered by overtrust or undertrust is a manifestation of the dynamization of trust, current research has not yet effectively distinguished the internal structure of human-AI trust. Consequently, scholarly exploration of the dynamic development of human-AI trust and its enabling mechanisms remains limited. On the one hand, existing research on dynamic trust has focused on a quantitative perspective of trust intensity, describing the degree of trust deviation in terms of undertrust or overtrust, while neglecting the multidimensional coordination inherent in the dynamic development of trust. The dynamism of trust involves not only the level of trust, but should also include trust structure and temporal embeddedness (Lee & See, 2004; Skjuve et al., 2021; Vuori et al., 2026)—that is, how the structure and level of trust change over time. On the other hand, existing research on the effects of human-AI trust has focused on explicating its role in promoting AI use and individual performance. Yet the enabling value of human-AI trust should not stop at facilitating the acceptance and use of AI; the more central concern lies in helping individuals achieve effective AI use and higher-order collaboration (Chen Hui, Feng Chao, in press; Xu Hui et al., 2025; Lu & Yan, in press)—that is, how different structures of human-AI trust help individuals use AI effectively in differentiated ways and enhance individual capabilities.

In sum, current research on human-AI trust is still in its ascendant stage. However, given the three research gaps noted above—namely, unclear boundaries of the internal structure, the absence of a dynamic evolution mechanism, and an unclear pathway of deeper enablement—existing research has difficulty systematically explaining how individual employees in the digital-intelligence era can effectively construct, maintain, and make good use of human-AI trust. It is therefore necessary to reinterpret the internal structure of human-AI trust and, within this framework, systematically explore the laws governing its dynamic evolution and the mechanisms through which it enables individuals at a deeper level.

3 Research Framework

To address the limitations of existing research, this study will adopt a dynamic developmental perspective and focus on examining the evolution mechanism of human–AI trust in workplace settings, while dialectically analyzing changes in human–AI relationships in the digital-intelligence era. Specifically, the study has three research objectives: (1) theoretical construction: to propose a theoretically persuasive multidimensional construct of human–AI trust and develop a corresponding measurement scale; (2) dynamic analysis: to reveal the temporal evolution patterns of human–AI trust and the time-varying effects of trust cues; and (3) enabling pathway: to clarify the evolutionary pathway through which human–AI trust stimulates individual creativity, and to analyze the boundary conditions of individual thinking models and organizational management models. Corresponding to these objectives, this study comprises three interrelated substudies: Study 1 focuses on clarifying the connotation and measurement of human–AI trust in the workplace and serves as the logical starting point of the present research; from a dynamic developmental perspective, Study 2 focuses on examining the nonlinear developmental pathway of human–AI trust under the coupled influence of space–time contexts, and dialectically analyzes the developmental changes of human–AI trust in the digital-intelligence era, constituting the focus of this research; from the perspective of human–AI collaboration, Study 3 focuses on the evolutionary process through which human–AI trust stimulates individual creativity, representing the extension and elevation of this research. The overall research framework is shown in Figure 1.

3.1 Study 1: The Connotation, Measurement, and Nomological Network of Human–AI Trust

This study intends to explore the specific connotation of human–AI trust and to develop a measurement scale for human–AI trust in strict accordance with psychological scale construction procedures. Hinkin (1998) pointed out that, as a new construct, it is necessary to draw on a nomological network to seek corresponding construct evidence externally, thereby clarifying the conceptual characteristics of the new construct. Accordingly, this study will select antecedent and outcome variables that are theoretically highly related to human–AI trust to construct a nomological network.

Study 2: Temporal Evolution Mechanism of Human–AI Trust

- **Trust Cues**
  - Preset cues
  - AI cues

  →
- **Tool Trust**
- **Value Trust**

- **Tool Trust / Value Trust**

Study 1: The Connotation, Measurement, and Logical Relationship Network of Human-AI Trust

- **AI Characteristics**

  ↓

- **Human-AI Trust**
  - Tool trust
  - Value trust

  ↓

- **AI Role Positioning**

Study 3: The Evolution from Human-AI Trust to Creativity—From the Perspective of Human-AI Collaboration

- **Algorithmic Management**

  ↓

- **Human-AI Collaboration**
  - Divisional collaboration
  - Symbiotic collaboration

  →

- **Creativity**
  - Incremental creativity
  - Breakthrough creativity

- **Mindset**
  (Fixed vs. growth)

Figure 1. Overall Research Framework

3.1.1 The Connotation and Measurement of Human-AI Trust

Human-AI trust is a topic of shared concern across multiple disciplines. Social psychology and organizational behavior, as well as information management and human factors engineering, have all defined it. Among these definitions, social psychologist Mayer et al. (1995) proposed a widely cited definition of trust: trust is a positive psychological state in which an individual, based on positive expectations of another party, voluntarily exposes their own vulnerability and is willing to bear the risk of being harmed by that party. Although Mayer et al.'

s (1995) definition of trust originated in research on interpersonal relationships, this definition is not limited to human–human interaction (Glikson & Woolley, 2020; Wang et al., 2016) and can be extended to trust relationships between humans and technology. At present, a large portion of research on human–AI trust has been extended or transformed from the field of interpersonal trust research (Qi Yue et al., 2024; Lalot & Bertram, 2025). Mayer et al.' s (1995) definition is highly similar to another definition that is frequently cited in the field of trust in automation, namely Lee and See' s (2004) definition of human–AI trust: an attitude whereby an individual, under conditions of known uncertainty and vulnerability, believes that an agent can help the individual achieve their goals. In addition, although definitions of human–AI trust in other disciplines include some different assumptions, such as human social embeddedness, conceptualizing trust as the tendency to take meaningful risks when one believes there is a relatively high likelihood of positive outcomes is a general consensus across disciplines (Hoff & Bashir, 2015). Therefore, this study follows the concept of Mayer et al. (1995) and defines human–AI trust as: a psychological state in which an individual, based on positive expectations of AI, voluntarily exposes vulnerability and bears the risk of being harmed by AI.

There is broad consensus on the definition of human–AI trust, but in empirical research, the measurement dimensions of human–AI trust differ, mainly taking two forms. One is unidimensional measurement; commonly used scales include the 4-item scale of Choung et al. (2023) and the 11-item scale of Chowdhury et al. (2022). The other is multidimensional measurement; the common approach is to distinguish cognitive trust from affective trust in human–AI trust (Glikson & Woolley, 2020; Komiak & Benbasat, 2006). This distinction originates from McAllister' s (1995) distinction between affect and cognition in interpersonal trust and was introduced into research on human–AI trust by Komiak and Benbasat (2006). Cognitive trust refers to an individual' s rational evaluation of AI, believing that AI is useful; affective trust refers to an individual' s emotional response to AI and is irrational to a certain extent.

At present, both paradigms for measuring human–AI trust have certain limitations. First, unidimensional measures of human–AI trust in fact focus on trust at the cognitive level; for example, the 4-item scale of Choung et al. (2023) focuses on AI capability and data security. Second, within the two-dimensional approach, the content measured at the cognitive level is inconsistent. Existing scales of cognitive trust mainly measure whether AI is trustworthy. Drawing on research on interpersonal trust, existing studies generally hold that AI trustworthiness has three dimensions: ability, integrity, and benevolence. However, existing studies differ on whether cognitive trust needs to include all three dimensions. Some studies argue that cognitive trust should focus primarily on AI ability (Choung et al., 2023; Hu et al., 2021); some hold that it should include AI ability and AI integrity (Komiak & Benbasat, 2006); some believe it should include AI integrity and AI benevolence (Moussawi & Benbunan-Fich, 2021); and still others contend that AI ability, AI integrity, and AI benevolence should all be included (Wang et al., 2016). Third, within the two-dimensional

approach, measurement at the affective level also varies. Some studies focus on emotional experiences brought about by AI, such as satisfaction, excitement, and comfort (Komiak & Benbasat, 2006; Moussawi & Benbunan-Fich, 2021), referring to this as emotional trust; others emphasize the emotional bond between the individual and AI, whether AI shows concern for people, and whether the individual is attached to AI (Wang et al., 2016), referring to this as affective trust. It can thus be

It can be seen that although many studies use the concept of human–AI trust, their measurement methods are far from identical, which poses a major challenge to effective dialogue across different studies.

There are two reasons for this phenomenon. First, psychology divides human mentality and experience into two parts: cognition and emotion (Forgas, 2008). Cognition is undoubtedly the broader of the two concepts; the American Psychological Association defines cognition as the processes encompassing all conscious and unconscious aspects such as perception, reasoning, and judgment (American Psychological Association, 2020). It is reasonable to incorporate any component of AI capability, integrity, and benevolence into cognitive trust. Second, emotion is usually defined as an internal affective state that includes feelings and moods (Barsade & Gibson, 2007), yet common measures of emotional trust more often derive from perceptions of AI motives—believing that it is able to care about humans—which is essentially also a form of cognition about AI. Moreover, trust generated in any way is usually accompanied by emotional responses (Lee et al., 2023). Furthermore, in the field of interpersonal trust research, the classification of trust into cognition and emotion has also long been debated (Legood et al., 2023; Tomlinson et al., 2020; van Knippenberg, 2018).

In addition, some studies incorporate antecedents or bases of trust—such as an individual' s propensity to trust, task characteristics, institutional trust, and even self-efficacy—into the measurement dimensions of human–AI trust (Chi et al., 2021; Huang et al., 2022), thereby conflating the basis of trust with the form of trust.

In sum, this study argues that it is necessary to rearticulate the connotative dimensions of human–AI trust and to develop a more scientific scale of human–AI trust, thereby providing a high-quality measurement tool for subsequent empirical research on the development theory of human–AI trust. Returning to the definition of human–AI trust, it emphasizes two core elements of trust. The first is "trust beliefs," namely positive expectations of AI, reflecting individuals' cognition and evaluation of AI and constituting the cognitive foundation and content of trust. The second is "trust intention," namely the willingness to expose oneself to vulnerability, reflecting an individual' s willingness to assume the risks of trust, and constituting the psychological tendency and behavioral readiness generated on the basis of belief-based judgments. Existing trust theories hold that trust beliefs are the direct antecedent of trust intention (proximal predictor), the primary element among the two elements of trust, and that trust intention is the implementation preparation for trust beliefs (Conchie et al.,

2012; Mayer et al., 1995; McKnight et al., 1998; McKnight & Chervany, 2006). Accordingly, based on differences in the value orientation of individual beliefs, we divide human–AI trust into instrumental trust and value trust. Specifically, instrumental trust is the trust generated from individuals' positive expectations regarding AI' s technical functions and task efficacy; from the "is" perspective, it concerns whether AI can accomplish established goals in a stable, reliable, and efficient manner. Value trust is the trust generated from individuals' positive expectations regarding AI' s adherence to and promotion of human values; from the "ought" perspective, it concerns whether AI can observe and promote human moral principles, social norms, and long-term welfare. Classifying trust according to beliefs can clearly delineate the conceptual boundaries of different types of trust, avoid conflating the content of trust with the implementation of trust, and thereby preserve the parsimony and explanatory power of the theory. This classification does not weaken the importance of trust intention; on the contrary, positioning trust intention as a downstream consequence of trust beliefs can provide a coherent foundation for future process-oriented research on trust dynamics (see Ballinger et al., 2025; McKnight & Chervany, 2006; van der Werff et al., 2019; Weber et al., 2004). Moreover, this research orientation is consistent with mainstream approaches in the field of interpersonal trust research. Existing interpersonal trust research has widely classified trust on the basis of different content dimensions of beliefs—for example, the classifications of ability-based trust, integrity-based trust, and benevolence-based trust (Kim et al., 2004; Mayer et al., 1995), as well as those based on knowl-

knowledge-based trust and identification-based trust (Lewicki & Bunker, 1996).

Classifying human–AI trust into instrumental trust and value trust has three advantages. First, the binary classification of instrumental trust and value trust reflects the dual logic of human acceptance of technology: instrumental trust addresses the question of "whether it can be used," whereas value trust answers the further question of "whether it should be used." Second, instrumental trust and value trust are positive states generated by individuals'identification with AI at the functional and value levels, respectively, thereby avoiding the problems of overly broad definitions of cognition and emotion and their mutual entanglement. Third, both instrumental trust and value trust are forms of trust generated on the basis of beliefs about AI itself; they are different forms of human–AI trust. This distinguishes them from approaches that conflate antecedents of human–AI trust—such as individual characteristics (e.g., propensity to trust) and environmental characteristics (e.g., institutional trust)—with different forms of human–AI trust.

A review of the literature shows that individuals' concern with the functional efficacy of AI involves two aspects: (1) basic capability, namely AI' s ability to achieve a given goal on its own, such as algorithmic accuracy, task reliability, and resistance to interference; and (2) collaborative capability, namely AI' s capability to collaborate with humans, such as iterative intention understanding, the expansion of communication bandwidth, and affective co-calibration. Basic

capability emphasizes the intrinsic stability of AI in achieving goals, ensuring that "nothing goes wrong"(the technical bottom line) and addressing "whether it can do it." Collaborative capability points to the bidirectional adaptive capacity of AI in collaboration with humans, pursues "being better together" (the upper bound of experience), and answers "how to do it efficiently."

Individuals' concern with the legitimacy of AI' s ethical values also has two aspects: (1) the moral baseline, namely AI' s adherence to basic moral bottom lines, including honesty, fairness, and resisting prejudice and discrimination; and (2) moral extension, namely the social responsibilities that AI proactively assumes, including promoting social inclusion, maintaining intergenerational fairness, and attending to long-term welfare. The moral baseline emphasizes defensive bottom-line control ("not transgressing"), whereas moral extension emphasizes constructive value creation ("proactive goodness"). Table 1 lists the dimensions of human–AI trust and possible measurement items.

Compared with existing mainstream scales (e.g., Chi et al., 2021; Choung et al., 2023), the human–AI trust scale proposed in this study not only clarifies the construct hierarchy but also effectively deepens the item dimensions. First, at the level of the construct hierarchy, this scale takes trust beliefs as its core and divides human–AI trust into two dimensions—instrumental trust and value trust. It clearly separates antecedents such as individual differences in trust and contextual dependence, focuses on measuring individuals' trust in AI attributes, and enhances the focus and parsimony of the construct structure. Second, in the dimension of instrumental trust, existing scales' measurement of AI technological functions has largely remained at the level of "basic capability," focusing on AI' s expertise and reliability. For example, "AI technologies are competent in their area of expertise"(Choung et al., 2023) and "AI technologies will provide me with the help I need" (Chi et al., 2021). By contrast, this scale operationalizes "collaborative capability" as an independent dimension, systematically depicting AI' s functions in understanding intentions, dynamically adapting, and improving capabilities during interaction with humans. The introduction of this dimension not only enriches the connotation of instrumental trust and captures key information about efficient collaboration in AI technological efficacy, but also better accords with the current trend in which human–AI interaction is increasingly moving toward deep integration and collaborative development (Tian Jianing & Luo Jinlian, 2025; Choudhary et al., 2025; Fügener et al., 2026). Third, in the dimension of value trust, measurements related to the "moral baseline" in existing scales are mostly presented through vague expressions of "AI integrity," such as "AI is honest" (Huang et al., 2022; Komiak & Benbasat, 2006) and "AI keeps its commitments and delivers on its promises" (Choung et al., 2023). This scale, however, will 将

It is concretized as observable and judgeable ethical bottom lines, including protecting privacy, avoiding biased and discriminatory content, and refusing to execute commands that violate ethics. This concrete expression reduces ambiguity in respondents' understanding and can more precisely measure individuals'

trust that AI will comply with basic moral norms. Moreover, most existing scales related to "moral extension" are limited to AI' s benevolence toward individual users, for example, "AI would act in my best interest" (Hu et al., 2021; Moussawi & Benbunan-Fich, 2021), and "AI cares about our well-being" (Choung et al., 2023; Wang et al., 2016). The present scale breaks through this limitation by extending moral expectations to broader and longer-term social and human-level values, including being oriented toward the long-term welfare of humanity and proactively identifying ethical risks. This reflects higher expectations for AI' s social responsibility and positive ethical role (Heyder et al., 2023). The addition of this dimension enables the scale to capture trust in major or long-range decisions that transcends immediate individual utility and concerns the overall interests of humanity, thereby enriching the value connotation of human–AI trust.

Table 1 Example Items of the Human–AI Trust Scale

Construct	Dimension	Example items
Instrumental trust	Basic capability	1. I believe AI is capable of providing professional advice.2. I believe AI is capable of providing information and services that interest me.3. I believe AI consistently maintains stable performance in routine tasks.……
	Collaborative capability	1. I believe AI can accurately identify and adapt to my work habits and preferences.2. I believe AI will dynamically adjust work strategies based on feedback.3. When interacting with AI, I believe AI can proactively compensate for my capability shortcomings.……

Construct	Dimension	Example items
Value trust	Moral baseline	1. I believe AI will not misuse my private data.2. I believe AI can effectively avoid producing biased and discriminatory content.3. I believe AI will refuse to execute instructions that violate ethical principles.……
	Moral extension	1. I believe the advice and decisions provided by AI are oriented toward the long-term welfare of humanity.2. I believe AI considers long-term impact factors such as environmental sustainability in decision-making.3. I believe AI will proactively identify and remind users of potential ethical risks. ……

3.1.2 The Logical Relational Network of Human–AI Trust

(1) Discriminant validity

To clarify the distinctiveness of the human–AI trust construct (including instrumental trust and value trust), based on its definition, we selected two constructs involving individuals' positive beliefs or attitudes toward AI—namely, AI self-efficacy and AI appreciation—and compared and distinguished them from the human–AI trust construct.

AI self-efficacy refers to the degree of confidence individuals have in their ability to use and interact with AI (桂橙林等, 2024; Dong

et al., 2025). Although individuals with a high sense of AI self-efficacy may be more inclined to interact with AI, thereby developing positive beliefs about certain aspects of AI and generating trust (Zhong Dingjing et al., 2025; Montag et al., 2023), the focal concerns of the two are not the same. AI self-efficacy is a self-centered belief about one' s own capability; its focus is on the individual,

rather than on whether AI itself is trustworthy. By contrast, human–AI trust is a relational psychological state oriented toward AI as the target object. Its core lies in whether an individual is willing, on the basis of certain positive expectations about AI, to voluntarily expose their own vulnerability and assume potential risks. Even if an individual is highly confident that they possess the ability to use AI, they may still be unwilling to trust AI in critical decision-making because of concerns about AI' s unreliability or ethical risks.

AI appreciation refers to an attitude in which individuals are willing to accept the judgment that AI is superior to humans, and are inclined to adopt advice and services provided by AI rather than by humans (Le Chengyi et al., 2024; Qin et al., 2025). Although high levels of AI trust may promote a high degree of AI appreciation (Huynh, in press; Logg et al., 2019), the frameworks by which the two are established are not the same. AI appreciation emphasizes a relatively preferential attitudinal tendency: that is, within an "AI-human" comparative framework, individuals are more willing to choose AI; whereas human–AI trust is the individual' s pure judgment and attitude toward AI. In other words, an individual may trust both AI and humans at the tool or value level, but AI appreciation reflects that the individual is more willing to choose AI. For example, in human–AI teams, individuals who trust AI may cooperate with both AI and human colleagues (Erengin et al., in press; Ulfert et al., 2024), whereas individuals who appreciate AI will be more willing to choose to cooperate with AI (Logg et al., 2019). Therefore, this study proposes:

Proposition 1-1: The construct of human–AI trust (tool trust and value trust) differs from AI self-efficacy and AI appreciation.

(2) Logical relationship network

To verify the differences and connections between the two constructs of tool trust and value trust, based on the definitions of the constructs themselves, we selected different antecedent and outcome variables for tool trust and value trust in the logical relationship network, as well as the same antecedent and outcome variables, in order to test the validity of the constructs (see Figure 2).

```
flowchart LR
    U[Usefulness] --> TT[Tool trust]
    TP[Trust propensity] --> TT
    VA[Value alignment] --> VT[Value trust]
    TP --> VT

    TT --> Tool[Viewing AI as a tool]
    TT --> Freq[AI use frequency]
    VT --> Freq
    VT --> Teammate[Viewing AI as a teammate]

    ASE[AI self-efficacy] -.-> Tool
    ASE -.-> Freq
```

```
ASE -.-> Teammate
APP[AI appreciation] -.-> Tool
APP -.-> Freq
APP -.-> Teammate
```

Figure 2. Schematic diagram of the logical relationship network of human-AI trust

Human-AI trust is, in essence, a psychological state based on positive expectations. Its formation depends on cues concerning AI attributes and on individuals' subjective judgments of these cues (Glikson & Woolley, 2020; Vanneste & Puranam, 2025; Wirz et al., 2025). In technological contexts, individuals' judgments about whether a technology is trustworthy usually first originate from their evaluations of the technology' s functionality and task performance (Afroogh et al., 2024; Davis & Granić, 2024; King & He, 2006). AI usefulness as

as one of the most intuitive and core functional cues, reflecting the extent to which individuals perceive that using AI can improve performance (Childers et al., 2001; Davis, 1989).

According to the definition of instrumental trust, instrumental trust is a psychological state in which individuals, based on positive expectations regarding the technical functions and task efficacy of AI, voluntarily expose themselves to vulnerability and accept the risk of possible harm from AI. Its core lies in judging whether AI can stably, reliably, and efficiently accomplish preset goals, embodying a rational judgment oriented toward "what is." When individuals perceive AI as highly useful, this means that they expect AI to effectively improve the efficiency and quality of task execution (Huang et al., 2022; Topsakal, 2025). The positive emotions and favorable evaluations elicited by such positive performance expectations reduce individuals' perceptions of uncertainty and risk regarding AI (Kim et al., 2021; Magni et al., 2023). At this point, even if individuals recognize that the use of AI entails potential risks (e.g., AI may make errors), they still tend to use AI in task execution (Huang et al., 2022; Singh & Sinha, 2020) and to trust its judgments and outputs, thereby forming instrumental trust. Existing research on technology acceptance and automation trust also indicates that system performance and usefulness are important preconditions for individuals to establish trust (Lee & See, 2004; Hoff & Bashir, 2015).

Furthermore, trust, as a combination of beliefs and intentions, influences individuals' cognition of human-AI relationships (Erengin et al., in press; Park & Yoon, 2024). Individuals' role cognition of AI in human-AI relationships has two orientations: one is to view AI as a tool used to assist humans in completing tasks; the other is to view AI as a teammate that collaborates with humans to complete tasks (Einola & Khoreva, 2023; Qi et al., 2025; Sedlakova & Trachsel, 2023). Positioning AI as a tool or teammate is a cognitive representation of individuals' human-AI relationships, reflecting the psychological positioning of

AI' s role attributes in human–AI interaction (Anthony et al., 2023; Kim et al., 2021; Xu & Li, 2022). When individuals' trust in AI is mainly anchored in its instrumental efficacy, this trust model shapes a one-way, utilitarian interaction script (Ding Shushu et al., 2024). In this script, the value of AI is defined by the functional outcomes it produces. The relationship between individuals and AI is analogous to the relationship between a person and an instrument or a piece of software (Li Kai, Hu Fangzhou, 2025; Kim et al., 2021). The primary purpose of human–AI interaction is to obtain expected outputs through inputting commands. The core features of this relational cognition are depersonalization and goal detachment (Anthony et al., 2023). Individuals do not expect to establish empathy or value consensus with AI; instead, they focus on whether AI' s responses to input are accurate and whether it can achieve the goals set by the individual. In other words, instrumental trust prompts individuals to be more inclined to place AI within the category of "object" rather than "subject," and to view AI in human–AI relationships as a tool for achieving preset goals rather than as a social actor with independent value claims. Therefore, this study proposes:

Proposition 1-2: AI usefulness enhances instrumental trust, thereby leading individuals to view AI as a tool in human–AI relationships.

Unlike instrumental trust, which focuses on the functional attributes of AI, value trust emphasizes individuals' positive expectations regarding whether AI abides by and promotes human values, embodying a normative judgment oriented toward "what ought to be." In human–AI interaction contexts, individuals' evaluations of AI are no longer limited to whether it can complete tasks efficiently; rather, they pay greater attention to whether AI does the right thing—that is, whether its decisions and behaviors follow human moral principles, social norms, and long-term welfare orientations.

Value alignment, as the core embodiment of AI' s ethical attributes, refers to the extent to which AI can act in accordance with human goals and values (Saffarizadeh et al., 2024). When individuals perceive that AI aligns with human values in its decisions and behaviors, they are more likely to believe that AI will not engage in behaviors that violate human morality or social norms (McKee et al., 2023;

Omrani et al., 2022), thereby forming value trust in AI.

Furthermore, value trust shapes how individuals regard AI as an acting subject. Research in moral psychology indicates that when individuals believe that an object observes and respects core moral values, they include it within their own moral community—that is, they develop moral inclusion (moral inclusion; Passini, 2016; Passini & Morselli, 2017). Moral inclusion means that individuals believe the object should be treated according to the same moral values and rules of justice as humans, and is therefore viewed as an actor with a certain moral status. Value trust indicates that AI is not merely a technological object, but an acting subject capable of understanding, respecting, and practicing human

rules (Bonnefon et al., 2024; Ladak et al., 2024). In other words, value trust encourages individuals, at the moral level, to regard AI as an actor relatively equal to humans, rather than simply as a tool or means. This perception of equality helps individuals weaken the distinction between "master–subordinate" or "human–object" in human–AI relations, and instead makes them more inclined to regard AI as a work partner or teammate with whom they can collaborate and share goals. Viewing AI as a teammate means cognitively acknowledging it as a collaborator with intentionality, agency, and some form of subjectivity. The basic characteristics of teammate relationships are relative equality, shared goals, mutual dependence, and coordinated adjustment (Ding Shukan et al., 2024; Anthony et al., 2023; Seeber et al., 2020). Moral inclusion based on value trust makes individuals more willing to believe that AI' s intentions are benevolent and that, in unpredictable and complex situations, its behavior will still follow moral principles and social norms. This greatly reduces ethical risks in collaboration, prompting individuals no longer merely to issue commands to AI, but to begin sharing situational information with it, negotiating task goals, jointly assuming responsibility, and sharing blame when problems arise. AI thus transforms from a passive, value-neutral execution terminal into a partner that can actively contribute, is worthy of trust, and can co-create value with the individual. Therefore, this study proposes:

Proposition 1-3: AI value alignment enhances value trust, thereby prompting individuals to regard AI as a teammate in human–AI relationships.

In addition to the attributes of the trust object itself, individual characteristics are also an important source in the formation of trust. Trust propensity reflects an individual' s general tendency to hold trust when interacting with others or unfamiliar entities, and is regarded as a foundational antecedent of multiple types of trust relationships (Cabiddu et al., 2022; Kraus et al., 2020; Mayer et al., 1995). Trust propensity functions like a filter, influencing the speed and degree to which individuals form initial trust in new objects under conditions of incomplete information (Gefen, 2000; van der Werff & Buckley, 2017). In human–AI interaction contexts, trust propensity simultaneously strengthens individuals' instrumental trust and value trust in AI, thereby increasing the frequency of AI use.

First, along the instrumental-trust pathway, individuals with high trust propensity hold a more optimistic prior attitude toward AI' s technical functions (Cabiddu et al., 2022; Montag et al., 2023). When faced with emerging AI, they tend by default to assume that AI can operate effectively as designed or advertised, unless strong counterevidence emerges. This tendency reduces their need to verify AI' s functions in detail before using it. Therefore, individuals with high trust propensity can more quickly form positive expectations of AI' s usefulness, thereby accelerating the establishment of instrumental trust.

Second, along the value-trust pathway, high trust propensity is manifested as a basic belief in AI' s benevolence (Chi et al., 2021; Lalot & Bertram, 2025). When facing AI, high trust propensity makes individuals more willing to believe

in the ethical intentions of AI designers and developers, and to assume that AI' s underlying logic and decision-making rules conform to social ethical standards. They are more willing to accept ethical commitments concerning AI (such as privacy statements and fairness reports), which leads them, when encountering AI, to perceive lower ethical risks and moral threats, and thus to establish value trust more easily.

Furthermore, instrumental trust addresses individuals'pragmatic concerns about AI. High instrumental trust means that individuals believe using AI can improve efficiency and performance, thereby increasing AI use behavior (Wu Jun et al., 2024; Kim et al., 2021). Value trust, by contrast, addresses individuals' concerns about the ethical value of AI. High value trust means that individuals do not have to weigh the pursuit of efficiency against adherence to ethics; using AI will not induce moral discomfort or concerns about social image (Heyder et al., 2023). In other words, when individuals simultaneously or separately develop relatively high levels of instrumental trust and value trust, their perceived uncertainty and risk when facing AI decrease, making them more willing to use AI frequently. Therefore, this study proposes:

Proposition 1-4: Individuals' trust propensity increases (a) instrumental trust and (b) value trust, thereby increasing the frequency with which individuals use AI.

(3) Incremental validity

On the basis of verifying that instrumental trust and value trust are not conceptually redundant with similar constructs (i.e., AI self-efficacy and AI appreciation) and are related to human–AI relationships and frequency of AI use, we further propose that instrumental trust and value trust can explain unique variance in human–AI relationships and frequency of AI use beyond these similar constructs. As noted above, these similar constructs do not specifically focus on a relational, purely psychological state whose referent object is AI. Therefore, we predict that, compared with similar constructs, instrumental trust and value trust can further predict individuals' human–AI relationships and frequency of AI use on top of them.

Proposition 1-5: After controlling for AI self-efficacy and AI appreciation, (a) instrumental trust and (b) value trust can not only still significantly and positively predict frequency of AI use, but can also respectively significantly and positively predict the human–AI relationship of viewing AI as a tool and viewing AI as a teammate.

3.2 Study 2: The Temporal Evolution Mechanism of Human–AI Trust

In the field of interpersonal trust research, many scholars have proposed theoretical models of trust development (Lewicki et al., 2006; McEvily, 2011; McKnight et al., 1998; Rousseau et al., 1998), emphasizing that the development of

trust changes as the types and modes of information processing change (Jones & Shah, 2016; van der Werff & Buckley, 2017). Drawing on theories of interpersonal trust development, this study will analyze the temporal evolution mechanism of human–AI trust from an information-processing perspective: (1) the general characteristics of the temporal evolution of human–AI trust: as time passes, what are the evolutionary trajectories of instrumental trust and value trust? How does the human–AI trust structure composed of instrumental trust and value trust evolve over time? (2) the time-varying effects of types of trust cues on the construction of human–AI trust: how does the relative importance of different types of trust cues in the process of constructing human–AI trust evolve over time?

3.2.1 Temporal Evolution Characteristics of Human–AI Trust

The establishment of instrumental trust is grounded in the technical functionality of AI and is directly influenced by individual cognition. As an intelligent technology capable of performing cognitive functions related to human thought, AI produces outputs characterized by timeliness and intuitiveness (Bucinca et al., 2021). For example, when users employ autonomous driving technology, they can observe in real time, through the in-vehicle screen, the system' s recognition results for pedestrians and vehicles, and can simultaneously receive voice prompts for obstacle avoidance during emergency lane changes. The functional experience of intuitiveness and timeliness enables instrumental trust to be rapidly established at the beginning of human–AI interaction. As human–AI interaction increases, individuals' exploration of AI functions gradually extends from surface-level operations to deeper logic. Taking generative AI as an example, employees initially use ChatGPT for simple question-and-answer information collection; as frequency of use increases, individuals gradually uncover its complex functions, using ChatGPT for document production, code

code writing, trend forecasting, and other complex tasks (You Junzhe, 2023). The increase in tool trust is accompanied by the continual exploration, excavation, and use of AI functions. However, technical efficacy has objective upper limits (Townsend et al., 2025). For example, algorithmic accuracy cannot break through 100% (Shen Chang, 2024). When individuals perceive that the technology has reached its capability threshold, the growth of tool trust tends to stagnate, entering a stable period and forming a "ceiling effect" (see Figure 3a). Research by Manchon et al. (2022) shows that autonomous-driving users' trust in a system' s obstacle-avoidance capability rises rapidly during the initial stage of use, but after a period of use the level of trust tends to stabilize. Based on this, this study proposes:

Proposition 2-1: The evolution of tool trust exhibits a "ceiling effect," and its evolutionary trajectory is an inverted L-shaped curve. That is, individuals' tool trust in AI increases over time in the early stage; after reaching a certain threshold, it remains in a relatively stable state in the later stage and does not change with the passage of time.

The formation of value trust is rooted in individuals' process of cognitive internalization of AI ethical principles. Unlike the direct observability of technical functions, AI ethical principles are usually embedded in algorithmic black boxes in the form of implicit rules (Kong Xiangwei et al., 2022; Yue & Li, 2023), resulting in a double unobservability in their moral decision-making logic: first, the lack of transparency at the technical level, such as the uninterpretable parameters of deep learning models; and second, the temporal lag in ethical verification. Taking a human-resource scenario as an example, the technical efficacy of AI-based résumé screening can be verified through performance indicators such as hiring rates, but people cannot directly judge whether AI exhibits discrimination in résumé screening, such as implicit bias against female job applicants; this requires recognition and verification through multiple rounds of interaction (Budhwar et al., 2022). In addition, individuals' judgments of ethical principles are often characterized by prudence. Only when AI demonstrates a stable value orientation across continuous ethical scenarios are individuals willing to confirm the credibility of AI' s values (Wang Chen et al., 2024; Telkamp & Anderson, 2022). For example, if an autonomous vehicle chooses a braking strategy that protects pedestrians in a single pedestrian-crossing scenario, individuals may attribute it to chance or coincidence; but if the autonomous vehicle consistently adopts a conservative braking strategy across multiple pedestrian-crossing scenarios, individuals will confirm that its values are trustworthy. It can be seen that individuals' cognition of AI ethics must undergo an iterative verification cycle: through systematic performance in continuous moral scenarios, they gradually construct a cognitive schema of value trust. Therefore, the establishment of value trust involves a "threshold effect" : in the early stage, individuals' value trust in AI does not change over time; after reaching a certain threshold, it gradually increases over time (see Figure 3b). Based on this, this study proposes:

Proposition 2-2: The evolution of value trust exhibits a "threshold effect," and its evolutionary trajectory is a J-shaped curve. That is, individuals' value trust in AI remains in a relatively stable state in the early stage and does not change with the passage of time; after reaching a certain threshold, it increases over time in the later stage.

Although tool trust and value trust are mutually independent in psychological mechanisms and evolutionary paths, the two jointly constitute individuals' overall trust system toward AI. Therefore, understanding the temporal evolutionary characteristics of human–machine trust requires not only examining the respective trends of change in the two types of trust, but also attending to how their relative importance is dynamically adjusted over time. This paper uses changes in the ratio between tool trust and value trust to reflect changes in the relative importance of the trust structure.

As time develops, the ratio between tool trust and value trust will show stage-based differences. In the initial stage of human–machine relations, individuals' needs for AI are concentrated mainly on functional realization. Timely and

observable feedback on AI' s technical efficacy can rapidly reduce uncertainty and promote the rapid formation of tool trust. By contrast, value trust depends on the accumulation of ethical situations and the verification of cross-situational consistency. At this point, individuals still find it difficult to judge whether AI can comply with and promote human value norms. Therefore,

In the initial trust system, value trust has not yet emerged because of the temporal lag in ethical validation, whereas instrumental trust continues to accumulate and occupies a dominant position. This forms a "technology-first" trust structure, manifested as a continuous rise in the ratio of instrumental trust to value trust.

(a) Instrumental trust —Time

(b) Value trust —Time

(c) Instrumental trust / Value trust —Time

Figure 3. Schematic diagram of the temporal evolutionary trajectories of human-machine trust

Thereafter, as human-machine interaction increases, individuals' functional understanding of AI gradually deepens and stabilizes, and instrumental trust enters a relatively moderate stage of development—that is, a stage of diminishing marginal growth. At this stage, although the absolute level of instrumental trust may continue to rise, its growth rate slows markedly and may even tend toward stagnation. At this point, individuals' attention begins to shift toward value alignment that has not yet been satisfied. For example, users of autonomous driving are no longer satisfied merely with obstacle-avoidance success rates, but instead require the system to prioritize protecting pedestrians in moral dilemmas (Telkamp & Anderson, 2022). In this process, the threshold effect of value trust begins to appear. After the accumulation of earlier ethical scenarios, individuals gradually confirm the ethical stability of AI, and value trust enters an upward trajectory.

Therefore, as value trust increases, the growth of instrumental trust slows down or even remains stable, and the ratio of instrumental trust to value trust begins to decline.

Ultimately, as individuals' understanding of AI functions and values tends to stabilize, the complementary characteristics of instrumental trust and value trust in evolutionary sequence and psychological mechanism gradually become salient. The ratio between the two remains stable, forming a composite trust structure of "technological reliability—ethical trustworthiness." For example, in autonomous-driving contexts, users require the system to possess stable obstacle-avoidance capabilities, while also expecting it to embody clear ethical priorities in situations such as the "trolley problem" ; in medical AI scenarios, individuals attend not only to diagnostic accuracy but also place great importance on long-term data privacy protection (Xu Wei et al., 2024).

Overall, the dynamic change in the ratio between instrumental trust and value trust essentially reflects a shift in individuals' cognitive focus from "verification of technological usability" to "confirmation of value consistency." This process is simultaneously influenced by the technological ceiling and the threshold of ethical verification, thereby driving the trust structure from dominance by a single dimension toward multidimensional dynamic equilibrium (see Figure 3c). Based on this, this study proposes:

Proposition 2-3: As time progresses, the ratio of instrumental trust to value trust follows an inverted U-shaped curve that first increases and then decreases, and tends to stabilize after reaching a certain threshold.

3.2.2 Time-Varying Effects of Trust Cues on the Construction of Human-Machine Trust

After an overall exploration of the general pattern of the temporal evolution of human-machine trust, another core issue in the study of trust dynamics emerges: in the process of trust development, the time-varying effects of different trust cues (antecedents) on the construction of human-machine trust—that is, the relative importance of different trust cues across the various stages of the evolution of human-machine trust. Exploring this issue not only helps to reveal the evolutionary process of human-machine trust in a detailed and comprehensive manner, but also responds to recent calls by scholars for research on the temporal scope of the effects of trust cues (Tian Jiayu & Luo Jinlian, 2025; Dang & Li, 2026; Wirz et al., 2025).

In the field of interpersonal trust research, many scholars have proposed theoretical models of trust development (e.g., Lewicki et al., 2006; McEvily, 2011; McKnight et al., 1998; Rousseau et al., 1998; van der Werff et al., 2019). These models commonly point to a core view: the types of information used to form trust judgments and the ways in which such information is processed change as trust develops (Jones & Shah, 2016; van der Werff & Buckley, 2017). Specifically, early trust formation is described as a judgment process dominated by heuristic processing (Levin et al., 2006; van der Werff et al., 2019; Williams, 2001). Individuals mainly form initial trust judgments on the basis of macro-level cues such as subjective preferences, social categories, or situational labels. However, as time develops, trust judgments gradually shift from heuristic processing to analytic processing (Levin & Cross, 2004; Lewicki & Bunker, 1996; McEvily, 2011). At this point, individuals are able to directly observe and evaluate the trustee through repeated interactions, thereby making relatively rational trust decisions.

Drawing on the logic of the dynamic development of interpersonal trust, in human-machine trust, during the initial stage of interaction individuals find it difficult to directly obtain information about AI, and their trust judgments are more likely to rely on heuristic processing triggered by macro-level environmental cues or individuals' preexisting tendencies. As interactive experience

accumulates, individuals gradually obtain direct feedback on the performance and behavioral manifestations of AI systems; at this point, trust judgments are based more on analytic processing of the specific attributes of AI. Based on the above analysis of the current state of domestic and international research, existing studies have identified, at different levels, multiple trust cues that influence the construction of human–machine trust, including national culture at the macro level (Chi et al., 2023), organizational reputation (Hengstler et al., 2016), and individual trust propensity at the micro level (Riedl, 2022) and AI

attribute characteristics (such as transparency and reliability; Cabiddu et al., 2022; Schaefer et al., 2016), among others. Because this paper focuses on the evolution of individuals' trust in AI at work within organizational contexts, the analysis of trust cues in this study concentrates on the organizational and individual levels. Drawing on the concept of "presumptive trust" proposed by Kramer and Lewicki (2010)—individuals' overall expectations of all members within their environment—this study divides trust cues into presumptive cues and AI cues. Presumptive cues refer to information from the organizational environment (including institutional trust and organizational identification) and individuals' propensity to trust. Such cues are relatively stable and exist prior to specific human–AI interactions. By contrast, AI cues refer to AI' s attribute characteristics, including transparency and reliability; these cues must be gradually acquired by individuals during the process of human–AI interaction.

According to the principle of bounded rationality, in the process of trust construction, although individuals may strive to acquire various trust cues, they may be unable to obtain all information simultaneously, much less process all information at the same time (Bijlsma & Koopman, 2003; Simon, 1955). In practice, constrained by time and energy, individuals are more likely to choose to attend to a limited number of cues during a given period. In the early stage of trust construction, individuals have relatively little interaction with AI, and thus fewer opportunities to obtain direct cues about AI. At this point, individuals are more likely to engage in preliminary trust construction through presumptive cues—that is, by obtaining indirect information from the macro environment, or on the basis of personal preferences in their propensity to trust.

Institutional trust is an individual' s trust in the effectiveness of an organization' s formal or informal rules and systems, together with the belief that people and affairs within the organization operate under the effective constraints of institutional norms (Liao, 2008; Möllering, 2006). Organizational identification refers to the degree to which individuals perceive consistency with their organization in terms of behavior, values, and related aspects (Dutton et al., 1994). Institutional trust reflects an individual' s recognition of the organization' s rule system, helping the individual view AI as a controllable entity under institutional constraints; organizational identification reflects an individual' s recognition of their fit with the organization, helping the individual extend emotional attachment to the organization to recognition of the AI promoted by that organization. Propensity to trust is an individual' s general willingness to trust others; in

ambiguous situations, it can serve as a positive "filter" that influences how individuals interpret the external world and provides them with the capacity for a "leap of faith" in trusting others (Baer et al., 2018; Grant & Sumanth, 2009).

The advantage of relying on presumptive cues lies in the fact that, by freeing cognitive resources (Mayer & Gavin, 2005), individuals can begin cooperating with AI as quickly as possible. In essence, presumptive cues shape individuals' initial cognitive schemas of AI and guide their initial interactions with AI through heuristic information-processing modes, thereby facilitating the rapid establishment of the relationship.

As time passes and individuals' interactions with AI increase, they are able to accumulate more direct cues about the transparency and reliability of AI in operation. At this point, individuals can rely more on verifiable information obtained through interactional experience for analysis and processing. When individuals obtain verifiable information through direct interactional experience, the initial trust stage comes to an end (Gao Zaifeng et al., 2021; McEvily, 2011). Therefore, presumptive cues exert the primary influence in the early stage of trust construction, but over time, the influence of AI cues gradually comes to occupy a dominant position. Based on the above analysis, this study proposes:

Proposition 2-4: Presumptive cues (institutional trust, organizational identification, and propensity to trust) have a positive effect on early-stage instrumental trust and value trust; over time, this effect will gradually diminish in later stages.

Proposition 2-5: AI cues (transparency and reliability) have a relatively small positive effect on early-stage instrumental trust and value trust; over time, this effect will gradually increase in later stages.

To capture the dynamics of the temporal evolution of human–AI trust described above and delineate its potential nonlinear developmental trajectory, this

Some studies are designed methodologically to adopt a multi-wave longitudinal research design. Specifically, through multi-stage data collection, this study will employ the univariate latent growth model and the augmented latent growth model, respectively, to estimate the overall developmental trajectories of tool trust and value trust over time, and further examine the time-varying effects of different trust cues on the construction of human–AI trust. The advantage of using latent growth models is that they can not only characterize the average growth trend of variables over time, but also identify nonlinear features, changes in trajectory speed, and potential inflection points in the developmental process through changes in growth slopes, thereby providing statistical support for the temporal evolution of human–AI trust (Duncan et al., 2013; van der Werff & Buckley, 2017). In terms of implementation, this study plans to take the formal deployment of a newly introduced AI software system in a company as the starting point, and to conduct a four-month, nine-wave longitudinal survey of its employees, with measurements administered every two weeks. Existing research indicates that the development of human–AI relationships from an initially un-

familiar stage to a relatively stable stage generally requires approximately 2–3 months of continuous interaction (Croes & Antheunis, 2021; Skjuve et al., 2022; Xu & Li, 2022). A four-month tracking period can provide individuals with rich accumulated interaction experience, offering sufficient space for the formation and development of human–AI trust. In addition, according to the recommendation of Ployhart and Vandenberg (2010), robust estimation of nonlinear growth trajectories typically requires at least four or more measurement time points. The nine-wave longitudinal design adopted in this study not only meets the statistical requirements for nonlinear growth modeling, but also helps to more finely depict the evolutionary patterns of human–AI trust across different interaction stages and its potential stage-specific characteristics.

3.3 Study 3: The Evolution from Human–AI Trust to Creativity—A Human–AI Collaboration Perspective

Study 2 reveals, from an information-processing perspective, the differentiated temporal-evolution mechanisms of tool trust and value trust during continuous interaction, answering the question of how human–AI trust develops dynamically over time. However, merely depicting the temporal-evolution trajectory of human–AI trust is still insufficient to explain a more critical theoretical and practical question: after individuals form differently structured trust in AI, how does this trust structure further shape the way they collaborate with AI, and ultimately influence their capability performance? In fact, trust is not the endpoint of human–AI interaction, but an important precondition that determines whether, and how, humans incorporate AI into their own cognitive and behavioral systems. Trust not only affects whether individuals use and rely on AI, but also profoundly influences how they define the division of labor boundaries and depth of interaction between humans and machines.

Therefore, this study further shifts the analytic focus from how trust forms and evolves to how trust structures are transformed into specific human–AI collaboration patterns and their consequences. By introducing human–AI collaboration mode as a key mechanism, it aims to reveal how different trust structures shape human–AI collaboration patterns and individual creative performance in work contexts, and to systematically answer how, in the era of digital intelligence, human–AI trust shapes individuals' core advantages. In doing so, it realizes a contextual expansion and an outcome-level extension of the trust dynamics mechanism revealed in Study 2. At the same time, unlike Study 1, which focuses on the influence of trust structure on role cognition in human–AI relationships, this study examines how trust structure affects human–AI collaborative behavioral patterns, thereby extending the cognitive-level exploration of Study 1 to the behavioral level. Specifically, this study will focus on analyzing the following two questions: How do tool trust and value trust influence individual creativity? Furthermore, how do individuals' ways of thinking and organizations' management approaches affect the above process? The theoretical model of Study 3 is shown in Figure 4.

Figure 4. Schematic model of the effects of human–AI trust on creativity

- **Algorithmic management**
- **Human–AI trust**
  - Tool trust → **Segmented collaboration** → Incremental creativity
  - Value trust → **Symbiotic collaboration** → Breakthrough creativity
- **Mindset** *(fixed vs. growth)*
- **Human–AI collaboration**
  - Segmented collaboration
  - Symbiotic collaboration
- **Creativity**
  - Incremental creativity
  - Breakthrough creativity

3.3.1 Mechanisms Through Which Human–AI Trust Influences Individual Creativity

Human–AI collaboration refers to the sustained interaction and coordinated behavior formed between humans and AI in the process of achieving shared goals (Sowa et al., 2021). Unlike studies that emphasize how AI participates at a single decision node, such as sequential decision-making and joint decision-making, this study defines human–AI collaboration as a collaborative behavioral structure characterized by continuous interaction throughout the entire task process, focusing on the division of labor boundaries and the degree of interactive coupling between humans and AI during task execution. Based on the mode of division of labor and the level of interactive integration in collaboration, human–AI collaboration can be divided into segmented collaboration and symbiotic collaboration (Hentout et al., 2019; Shrestha et al., 2019; Wang et al., 2019). Segmented collaboration refers to humans and AI undertaking relatively independent submodules with clear boundaries in the task process, completing the overall task through serial or staged linkage. Its core features are a clear functional division of labor, stable responsibility boundaries, and information exchange dominated by the transmission of results. For example, in customer-service scenarios, standardized and rule-based questions are handled by intelligent customer-service systems, while complex or highly contextualized questions are completed through follow-up by human customer-service agents. This type of collaboration emphasizes modular decomposition and sequential integration (Jia et al., 2024). Symbiotic collaboration, by contrast, refers to humans and AI maintaining continuous participation at all stages of a task and realizing task co-creation through shared information resources, dynamic feedback, and iterative adjustment. Its core features are a high level of interactive coupling, dynamic negotiation of responsibility boundaries, and parallel integration. For example, in financial investment decision-making, investment advisors and AI systems continuously share market data and analytical models,

and jointly formulate and optimize investment strategies through repeated feedback and revision. This type of collaboration reflects structural characteristics of high integration and flexible symbiosis (丁述磊等, 2024; Burridge, 2017). It should be noted that the segmented/symbiotic collaboration in this study differs from the conceptual demarcation in Study 1 between viewing AI as a tool and viewing AI as a teammate; the two have different focal points. Viewing AI as a tool/teammate is a cognitive relational representation, focusing on the individual' s psychological positioning of AI' s role attributes (Anthony et al., 2023; Kim et al., 2021; Xu & Li, 2022); segmented/symbiotic collaboration is a behavioral collaborative structure, focusing on the division-of-labor pattern and degree of interactive coupling actually manifested in task execution between humans and AI (陈慧, 丰超, 印刷中; Inga et al., 2023; Wang et al., 2019). In other words, the former answers "how individuals view AI," whereas the latter answers "how individuals collaborate with AI" ; the two are respectively located at the cognitive level and the behavioral level and are distinct from one another.

Tool trust indicates that individuals believe AI possesses stable and verifiable technical advantages in a specific domain or task module. This type of trust focuses on capability boundaries and functional reliability. When individuals develop a relatively high level of tool trust, they are more likely to rely on the capabilities of both parties—

rationally divide labor according to differences in capabilities, assigning structured, rule-based, and computation-intensive subtasks to AI, while retaining for themselves those parts that require contextual judgment, complex integration, or value trade-offs (Gao Jinping & Li Yiqi, 2024; Freisinger & Schneider, 2025; Song & Lin, 2024). In other words, tool trust strengthens individuals' clear awareness of the boundaries of human–AI capabilities and prompts them to break down tasks and achieve complementary advantages through modular configuration logic, thereby improving overall efficiency (Kim et al., 2021). Under this configuration, the boundaries of responsibilities between humans and AI are relatively clear, the degree of interactive coupling is relatively low, and information exchange mainly takes the form of phased outputs, giving rise to divided collaboration. Research by Song and Lin (2024) shows that employees who believe in the technological capabilities of AI are more inclined to assign objective tasks to AI while reserving subjective tasks for themselves. Therefore, tool trust promotes the formation of divided collaboration by strengthening functional division of labor and modular integration mechanisms.

Value trust refers to individuals' belief that AI can follow stable ethical principles and value orientations in the decision-making process, reflecting norm consistency and a long-term welfare orientation. Unlike tool trust, which focuses on capability advantages, value trust emphasizes the predictability of decision principles and moral reliability. When individuals develop a high level of value trust, its core psychological basis is that AI will not deviate from basic norms or trigger ethical risks in complex situations. This value-level trust helps enhance individuals' psychological safety in deep interaction, making them more willing

to invest time and cognitive resources and to engage in continuous information exchange and bidirectional feedback with AI throughout the entire task process (Olan et al., 2022; Weisz et al., 2025). In this process, humans and AI achieve complementarity of advantages through dynamic consultation and continuous iteration (Ding Shulei et al., 2024; Jarrahi et al., 2023; Othman & Yang, 2023); the boundaries of division of labor tend to become flexible and blurred, and task responsibilities are generated through interaction rather than being preset and divided in advance, thereby forming highly coupled symbiotic collaboration. Based on the above analysis, this study proposes:

Proposition 3-1: Tool trust has a positive effect on divided collaboration.

Proposition 3-2: Value trust has a positive effect on symbiotic collaboration.

Creativity is usually defined as ideas about products, services, or processes that are novel and useful (Amabile et al., 1996). According to the degree of novelty and usefulness of ideas, creativity is divided into incremental creativity and radical creativity. Incremental creativity refers to individuals' minor improvements, within existing modes of thinking, to an organization' s current practices or products, and is more strongly related to usefulness; radical creativity refers to ideas that differ essentially from individuals' and organizations' existing practices or products, and is more strongly related to novelty (Madjar et al., 2011). By analogy, if radical creativity concerns an individual' s ability to innovate "from 0 to 1," incremental creativity refers to an individual' s ability to make improvements "from 1 to N" (Luo Nanfeng et al., 2024).

Divided collaboration is the sequential cooperation between humans and AI in the work process. Individuals fully authorize AI to perform certain work that is suitable for AI, freeing themselves from complex work tasks, thereby conserving cognitive resources and enabling greater focus on work modules suited to humans (Daniel & Zhan, 2023; Jia et al., 2024). Focusing on work content motivates individuals to attend to work details, making it more likely that they will improve and refine existing small-scope work and enhance work quality (Li et al., 2018; Petrou & Jongerling, 2024), thus achieving better coordination between their own work and AI' s work. Therefore, divided collaboration is conducive to stimulating incremental creativity.

Symbiotic collaboration involves a high degree of interaction and integration between humans and AI at work, with in-depth communication and cooperation based on shared goals. In this process, rich knowledge sharing and communication feedback between individuals and AI continuously reconstruct individuals' cognition (Taylor & Greve, 2006; Yu & Choi, 2022). At this point, individuals have the capacity to make previously clearly irrelevant or

integrating heterogeneous knowledge across domains, thereby creating conditions for the generation of novel ideas capable of overturning existing practical experience (Huang Xiaozhi et al., in press; Ren & Song, 2024); that is, symbiotic collaboration helps stimulate breakthrough creativity. Therefore, based on the above analysis, this study proposes:

Proposition 3-3: Segmented collaboration mediates the positive relationship between tool trust and incremental creativity.

Proposition 3-4: Symbiotic collaboration mediates the positive relationship between value trust and breakthrough creativity.

3.3.2 The Moderating Role of Individual Mindset and Organizational Algorithmic Management

The choice of human-AI collaboration mode is inseparable from the influence of individual traits and the environment in which the individual is situated (Haesevoets et al., 2021; Yin et al., 2024). To comprehensively understand the impact of human-AI trust on human-AI collaboration, it is necessary to consider both individual characteristics and the characteristics of the organizational culture environment in which the individual is embedded.

Mindsets are the core assumptions people hold about the malleability of personal traits and the attributes of things. They mark the beginning of a series of psychological processes and influence individuals' interpretation of and responses to specific situations, goals, behaviors, motivations, and other events and psychological activities. Dweck (2006) divides individuals' mindsets into two types: growth mindset and fixed mindset. Individuals with a growth mindset hold an "incremental view of ability," believing that personal ability is characterized by being improvable, malleable, and controllable, and can be continuously enhanced through effortful learning and training; whereas individuals with a fixed mindset hold an "entity view of ability," believing that personal ability is a fixed and unchanging trait.

This study argues that a fixed mindset can strengthen the relationship between tool trust and segmented collaboration, whereas a growth mindset strengthens the relationship between value trust and symbiotic collaboration. First, a fixed mindset tends to be conservative, favors a sense of control and following established procedures, and seeks to avoid mistakes as much as possible (Blackwell et al., 2007). Therefore, individuals with a fixed mindset are more inclined toward work arrangements with clear structures and explicit responsibilities, assigning tasks to AI and engaging in segmented collaboration so that they occupy a dominant position in human-AI collaboration and retain control. In contrast, individuals with a growth mindset tend to cooperate with others, hoping to acquire knowledge and achieve growth through cooperation (Fraune et al., 2019). Therefore, individuals with a growth mindset are more likely to regard AI as a source of cognitive expansion, engage in deeply integrated interactions with AI, carry out symbiotic collaboration, obtain feedback from AI, and improve their own capabilities. Second, individuals with a fixed mindset have a stronger performance-goal orientation and usually regard challenges as threats rather than opportunities (Dweck, 2012). They are more willing to believe that AI is threatening, and they believe that AI may surpass humans in capability (Dang & Liu, 2022a). Therefore, individuals with a fixed mindset are more

willing to choose segmented collaboration with AI, which can not only improve task-completion efficiency but also allow them to control the task process. In contrast, individuals with a growth mindset have greater openness to experience, tend to regard AI as an opportunity for their own learning and growth (Chen & Yi, 2024; Dang & Liu, 2022b), have a stronger willingness to interact with AI, attend to acquiring knowledge from AI and learning from its strengths to offset their own weaknesses, and are more willing to engage in symbiotic collaboration with AI. Based on this, this study proposes:

Proposition 3-5: (a) Mindset moderates the relationship between tool trust and segmented collaboration. Compared with a growth mindset, a fixed mindset strengthens the positive relationship between tool trust and segmented collaboration. (b) Mindset moderates the mediating role of segmented collaboration between tool trust and incremental creativity. Compared with a growth mindset, a fixed mindset strengthens the mediating role of segmented collaboration between tool trust and incremental creativity.

Proposition 3-6: (a) Mindset moderates the relationship between value trust and symbiotic collaboration. Compared with a fixed

···mindset; a growth mindset strengthens the positive relationship between value trust and symbiotic collaboration. (b) Mindset moderates the mediating role of symbiotic collaboration between value trust and breakthrough creativity. Compared with a fixed mindset, a growth mindset strengthens the mediating role of symbiotic collaboration between value trust and breakthrough creativity.

As an emerging practice of digital innovation management, algorithmic management refers to organizations' use of algorithms to perform managerial functions in a highly automated and data-driven manner (Liu Shanshi et al., 2022; Duggan et al., 2020). Algorithmic management provides employees with standardized guidance by setting work standards and offering informational support; at the same time, it tracks and evaluates employees and constrains their behavior by monitoring task progress and recording employees' work logs and behavioral habits (Zhan Xiaohui & Zhao Lijing, 2024). In organizations with high algorithmic management, individuals' work processes are precisely monitored and controlled by algorithms, and their work autonomy is restricted (Liu Shanshi et al., 2021; Ma Jun & Zhao Shuang, 2022). To avoid disputes, individuals tend to allocate low-skill tasks to AI and engage in segmented collaboration with AI. Moreover, through data analysis, organizations with high algorithmic management proactively provide employees with workflow-optimization solutions (Norlander et al., 2021), encouraging employees to engage in segmented collaboration with AI and improve efficiency. Conversely, in organizations with low algorithmic management, employees have greater role breadth (Wang Hongli et al., 2025), higher job security and work autonomy (Pei Jialiang et al., 2024), and individuals have the time and energy to improve themselves through their work. In addition, in organizations with low algorithmic management, the optimization and improvement of work methods rely more on individuals' autonomous exploration (Wei Wei & Liu Beini, 2023; Liu & Yin, 2024). At this point,

individuals tend to interact and communicate with AI, enhancing themselves through two-way feedback and exploring the optimization of work content and methods. Based on this, this study proposes:

Proposition 3-7: (a) Algorithmic management strengthens the positive relationship between tool trust and segmented collaboration. (b) Algorithmic management strengthens the mediating role of segmented collaboration between tool trust and incremental creativity.

Proposition 3-8: (a) Algorithmic management weakens the positive relationship between value trust and symbiotic collaboration. (b) Algorithmic management weakens the mediating role of symbiotic collaboration between value trust and breakthrough creativity.

To verify the theoretical model above and avoid the potential problems of model complexity and insufficient statistical power that may arise from testing multiple mechanism paths simultaneously in a single sample, this study plans to adopt a two-study design and verify the theoretical framework step by step in stages. First, a scenario experiment will be used, focusing on identifying the causal relationship between types of human–AI trust and human–AI collaboration structures. By manipulating scenarios to activate tool trust and value trust separately, the study will examine their effects on the human–AI collaboration mode selected by individuals. The experimental task will be designed as a human–AI collaborative decision-making scenario in which the division of labor can be freely configured; behavioral choice indicators will be used to operationally measure the collaboration structure, thereby identifying the causal effects of different trust types on collaboration-structure configuration. This study aims to verify the direct mechanism through which trust type affects collaboration mode, establishing the first link in the theoretical chain. Next, a three-wave, multi-source questionnaire design will be adopted, with a two-week interval between each measurement occasion (e.g., Chen Liping et al., 2025; Tu et al., in press), to test the complete theoretical model in real organizational settings. Specifically, at Time 1, participants will self-report human–AI trust, mindset, algorithmic management, and control variables; at Time 2, participants will self-report human–AI collaboration variables; and at Time 3, participants' supervisors will evaluate participants' creativity, thereby constructing a time-separated testing framework for moderated mediation paths. Through time-separated, multi-wave, multi-source data collection, it will be possible to control for common method bias while more robustly testing the complex mechanism model in which dual mediation and dual moderation coexist. By adopting a complementary research design that combines scenario experiments with longitudinal questionnaire surveys, this part of the study not only ensures the internal validity of causal inference, but also enhances the external validity and ecological validity of the theoretical model in real organizational contexts, thereby systematically testing

demonstrate how human–AI trust influences the generative pathway of creativity through collaborative structures.

4. Theoretical Construction

Human–AI trust is key to the success or failure of human–AI interaction. Existing research, by directly drawing on studies of interpersonal interaction and classifications of interpersonal trust, divides human–AI trust into cognitive trust and affective trust. However, human–AI interaction has distinctive features that differentiate it from interpersonal interaction, and the classification of interpersonal trust itself has limitations; as a result, existing research cannot adequately reflect the connotations and developmental laws of human–AI trust. Therefore, grounded in human–AI dyadic interaction, this study examines the evolutionary process and mechanisms by which human–AI trust changes over time in workplace contexts. Through a three-stage progressive research design of "theoretical reconstruction—dynamic laws—practical empowerment," it systematically answers the questions of "what it is—how it changes—how it can be used" with regard to human–AI trust in the digital-intelligent era. First, from the perspective of technology ethics, this study clarifies the connotation of human–AI trust, proposes a two-dimensional model of human–AI trust comprising instrumental trust and value trust, and, on this basis, develops a measurement scale for human–AI trust. Next, adopting a dynamic developmental perspective, it explores the general characteristics of the temporal evolution of instrumental trust and value trust under the coupled effects of spatiotemporal contexts, thereby opening the "black box" of the dynamic evolution of human–AI trust. Finally, taking the perspective of human–AI collaboration as a breakthrough point, it explores the differentiated mechanisms through which instrumental trust and value trust empower individual creativity, and explicates the development and transformation of human–AI relations in the digital-intelligent era, providing insights into how human–AI trust shapes individuals' core advantages in this era. This study makes three theoretical contributions.

First, it breaks through the traditional cognitive framework of human–AI trust. Grounded in the classic definition of trust and in the classic analytical framework that decomposes trust into trust beliefs and trust intentions, this study proposes a two-dimensional structural model of "instrumental trust—value trust" based on differences in the value orientations of individuals' trust beliefs. It systematically analyzes humans' dual logic of acceptance toward AI—functional reliability and value desirability—overcoming the theoretical and measurement limitations of existing studies that directly introduce the categories of cognitive trust and affective trust from interpersonal-trust research, and providing a theoretical framework for human–AI trust that is more contextually compatible and explanatory.

Second, it reveals the dynamic evolutionary mechanism of human–AI trust. From an information-processing perspective, this study integrates the three dimensions of trust structure, intensity, and time to analyze the nonlinear evolutionary trajectory of human–AI trust. It reveals the unbalanced co-evolutionary laws of instrumental trust and value trust and their critical-threshold effects, as well as the time-varying effects of different trust cues on the construction of

human–AI trust. In doing so, it remedies the insufficiency of existing dynamic research on human–AI trust, which focuses on the single dimension of trust intensity; refines the temporal scope of action of different types of trust cues; and enriches and improves dynamic research on human–AI trust.

Third, it clarifies the deep empowerment mechanism of human–AI trust. From the perspective of human–AI collaboration, this study explains the mechanisms by which instrumental trust and value trust differentially shape modes of human–AI collaboration and thereby differentially empower individual creativity, as well as the contextual effects of individual mindsets and organizational management models. It further deepens the general conclusion that "human–AI trust promotes AI acceptance and use" into "differentiated human–AI trust shapes differentiated AI use." This provides a concrete pathway for individuals to build creativity centered on human–AI collaboration, addresses concerns about cognitive degradation caused by excessive reliance on AI, and offers a dialectical view of the development and transformation of human–AI relations.

This study also has certain practical value. First, by proposing a two-dimensional model of "instrumental trust—value trust," it provides enterprises with an actionable pathway for building trust when introducing AI. Managers can accordingly start from the two aspects of improving technological reliability and strengthening value alignment, helping employees establish rational cognition of and value identification with AI, thereby enhancing the acceptance and effectiveness of human–AI collaboration. Second, the nonlinear evolutionary laws of human–AI trust and their critical thresholds revealed in this study provide key indicators for managers to dynamically manage human–AI trust relationships. Organizations can, based on the different developmental stage[[unclear: continuation off page]]

of stages, predict and identify critical points of trust, and prevent the emergence of trust collapse or overreliance. Finally, this study reveals the enabling pathway through which human–machine trust promotes individual creativity, providing a basis for individuals and enterprises to build innovation-management models centered on human–machine collaboration. By scientifically shaping human–machine trust relationships, individuals can integrate algorithmic insights with their own experience in the collaborative process and thereby enhance creativity; at the same time, on the basis of ensuring employees' psychological safety and cognitive vitality, enterprises can realize the synergistic enhancement of digital-intelligent technologies and human resources, helping enterprises achieve sustainable innovation and high-quality development in the digital-intelligent era.

References

Chen, H., & Feng, C. (in press). When employees encounter AI: Research on the construct measurement, antecedent configurations, and influence mechanisms of

employee-AI collaboration. *Advances in Psychological Science*.

Chen, L. P., Xu, M. Y., & Liu, S. M. (2025). The dual influence paths of workplace generative AI use on employee creativity. *Chinese Journal of Management*, *22*(2), 326-335.

Ding, S. L., Qi, F. D., Liu, C. H., & Li, J. Q. (2024). Evolution of labor forms, transformation of human-machine relations, and reconstruction of labor relations. *Economist*, (4), 45-55.

Gao, J. P., & Li, P. Y. (2024). Accounting task complexity, perceived trustworthiness, and the choice of human-machine collaboration mode. *Collected Essays on Finance and Economics*, (11), 80-91.

Gao, Z. F., Li, W. M., Liang, J. W., Pan, H. X., Xu, W., & Shen, M. W. (2021). Human-machine trust in autonomous vehicles. *Advances in Psychological Science*, *29*(12), 2172-2183.

Gui, C. L., Zhao, X. H., Zhang, P. C., Liu, Z. Q., & Zhou, R. (2024). The influence mechanism of employees' AI awareness on their innovation performance in the context of digital intelligence. *Human Resource Development of China*, *41*(8), 6-22.

Huang, X. Z., Zeng, L. M., Zhang, H., & Cao, X. (in press). How loose-tight culture influences consumers' radical and incremental creativity: A psychological adaptation perspective. *Nankai Business Review*.

Huang, X. Y., & Li, Y. (2024). Dual pathways of human-machine trust calibration: Trust inhibition and trust enhancement. *Advances in Psychological Science*, *32*(3), 527-542.

Kong, X. W., Wang, Z. M., Wang, M. Z., & Hu, X. P. (2022). Trustworthy decision-making in artificial-intelligence-enabled systems: Progress and challenges. *Journal of Industrial Engineering and Engineering Management*, *36*(6), 1-14.

Le, C. Y., Wang, Z. X., & Kong, W. W. (2024). Algorithm appreciation vs. algorithm aversion: Users' "algorithm paradox" under intelligent short-video recommendation. *Journal of Intelligence*, *43*(8), 170-181.

Li, K., & Hu, F. Z. (2025). From human-machine collaboration to human-intelligence collaboration: Conceptual clarification and future issues. *Nankai Business Review*, *28*(12), 48-60.

Liu, S. S., Pei, J. L., Ge, C. M., Liu, X. L., & Zhan, Y. B. (2022). Algorithmic management on online labor platforms: Theoretical exploration and research prospects. *Management World*, *38*(2), 225-239.

Liu, S. S., Pei, J. L., & Zhong, C. Y. (2021). Is platform work autonomous? The impact of algorithmic management on online labor platforms on work autonomy. *Foreign Economics & Management*, *43*(2), 51-67.

Luo, N. F., Li, T. J., Chen, W., Zhang, H. J., Liu, J. C., & Shen, Z. W. (2024). Have radical creativity and incremental creativity truly been distinguished?

An analysis based on literature from 2011 to 2024. *Advances in Psychological Science*, 32(11), 1882–1897.

Ma Jun, Zhao Shuang. (2022). An integrated analytical framework of algorithmic management and employee creativity. *Studies in Science of Science*, 40(10), 1811–1820.

Pei Jialiang, Liu Shanshi, Zhang Zhipeng, Xie Yu. (2024). Good algorithms, bad algorithms? A study of excessive labor among gig workers under algorithmic logic. *Journal of Industrial Engineering and Engineering Management*, 38(1), 101–115.

Qi Yue, Chen Junting, Qin Shaotian, Du Feng. (2024). Human–AI trust in the era of artificial general intelligence. *Advances in Psychological Science*, 32(12), 2124–2136.

Shen Yang. (2024). Why can generative large models not achieve 100% accuracy? Retrieved 2025-11-11 from https://t.cj.sina.com.cn/articles/view/6419993500/17ea9539c019019wrs?cref=

Tian Jianing, Luo Jinlian. (2025). From tools to mutualism: A study on the construction and dynamic evolution of human–AI collaborative relationships. *Management World*, 41(12), 179–197.

Wang Chen, Chen Weicong, Huang Liang, Hou Suyu, Wang Yiwen. (2024). Does robot compliance with ethics promote human–machine trust? The reversal effect of decision type and the human–machine projection hypothesis. *Acta Psychologica Sinica*, 56(2), 194–212.

Wang Haizhong, Xie Tao, Zhan Chunyu. (2021). The negative impact of intelligent customer-service avatar anthropomorphism in service-failure situations: The mediating mechanism of disgust. *Nankai Business Review*, 24(4), 194–204.

Wang Hongli, Chen Zhengren, Li Zhen, Liu Zhiqiang, Liang Cuiqi, Zhao Binjie. (2025). How to escape the temporal predicament: The subjective time boundary of the effects of algorithmic control on gig workers. *Acta Psychologica Sinica*, 57(2), 275–300.

Wang Xinye, Li Yuan, Chang Ming, You Xuqun. (2017). The hazards of automation trust and reliance to aviation safety and their improvement. *Advances in Psychological Science*, 25(9), 1614–1622.

Wei Wei, Liu Beini. (2023). Can algorithmic management improve digital gig workers' platform commitment? The double-edged sword effects of "controlism" and "decisionism." *Business Management Journal*, 45(4), 116–132.

Wu Jun, Zhang Di, Liu Tao, Liu Mantian, Zhao Shinan. (2024). Research on humans' acceptance of trust in artificial intelligence and its brain cognitive mechanisms: A meta-analysis of empirical studies and neuroscience experiments. *Journal of Industrial Engineering and Engineering Management*, 38(1), 60–73.

Xie Xiaoyun, Zuo Yuhan, Hu Qiongjing. (2021). Human resource management in the digital era: A perspective based on human-technology interaction. *Management World*, 37(1), 200-216.

Xu Hui, Long Yang, Li Yang, Lu Huibei. (2025). How manufacturing enterprises realize intelligent decision-making from the perspective of human-machine joint cognition. *China Industrial Economics*(4), 174-192.

Xu Wei, Gao Zaifeng, Ge Liezhong. (2024). New paradigmatic orientations and priorities for human factors science research in the intelligent era. *Acta Psychologica Sinica*, 56(3), 363-382.

You Junzhe. (2023). Application risks and control measures of ChatGPT-like generative artificial intelligence in scientific research scenarios. *Information Studies: Theory & Application*, 46(6), 24-32.

Zhan Xiaohui, Zhao Lijing. (2024). "Empowerment" or "burden" ? The double-edged sword effect of algorithmic management on workers' job performance on digital labor platforms. *Soft Science*, 38(7), 101-106.

Zhang, Z., Hua, Z., & Xie, X. (2024). Research status and future directions of human-machine collaboration in the digital-intelligence era. *Journal of Management Engineering*, *38*(1), 1-13.

Zhong, D., Wu, F., & Qiu, R. (2025). Anthropomorphization and intelligentization: An empirical study on AI anchors' mediation and the construction of human-machine trust relationships. *International Journalism*, *47*(02), 49-71.

Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: Progress, challenges, and future directions. *Humanities and Social Sciences Communications*, *11*(1), Article 1568.

Amabile, T. M., Conti, R., Coon, H., Lazenby, J., & Herron, M. (1996). Assessing the work environment for creativity. *Academy of Management Journal*, *39*(5), 1154-1184.

American Psychological Association. (2020). *Publication manual of the American psychological association* (7th ed.). American Psychological Association.

Anthony, C., Bechky, B. A., & Fayard, A. (2023). "Collaborating" with AI: Taking a system view to explore the future of work. *Organization Science*, *34*(5), 1672-1694.

Ayoub, J., Avetisyan, L., Makki, M., & Zhou, F. (2022). An investigation of drivers' dynamic situational trust in conditionally automated driving. *IEEE Transactions on Human-Machine Systems*, *52*(3), 501-511.

Baer, M. D., van der Werff, L., Colquitt, J. A., Rodell, J. B., Zipay, K. P., & Buckley, F. (2018). Trusting the "look and feel" : Situational normality, situational aesthetics, and the perceived trustworthiness of organizations. *Academy of Management Journal*, *61*(5), 1718-1740.

Ballinger, G. A., Schoorman, F. D., & Sharma, K. (2025). What we do while waiting: The experience of vulnerability in trusting relationships. *Academy of Management Review*, *50*(4), 768–787.

Barsade, S. G., & Gibson, D. E. (2007). Why does affect matter in organizations? *Academy of Management Perspectives*, *21*(1), 36–59.

Bawack, R. E., Wamba, S. F., & Carillo, K. D. A. (2021). Exploring the role of personality, trust, and privacy in customer experience performance during voice shopping: Evidence from SEM and fuzzy set qualitative comparative analysis. *International Journal of Information Management*, *58*, Article 102309.

Bijlsma, K., & Koopman, P. (2003). Introduction: Trust within organisations. *Personnel Review*, *32*(5), 543–555.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*(1), 246–263.

Bonnefon, J., Rahwan, I., & Shariff, A. (2024). The moral psychology of artificial intelligence. *Annual Review of Psychology*, *75*, 653–675.

Boussioux, L., Lane, J. N., Zhang, M., Jacimovic, V., & Lakhani, K. R. (2024). The crowdless future? Generative

AI and creative problem-solving. *Organization Science, 35*(5), 1589–1607.

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), Article 188.

Budhwar, P., Malik, A., De Silva, M. T. T., & Thevisuthan, P. (2022). Artificial intelligence–challenges and opportunities for international HRM: A review and research agenda. *The International Journal of Human Resource Management, 33*(6), 1065–1097.

Burridge, N. (2017). Artificial intelligence gets a seat in the boardroom: Hong Kong venture capitalist sees AI running Asian companies within 5 years, Retrieved November 11, 2025, from https://asia.nikkei.com/Business/Artificial-intelligence-gets-a-seat-in-the-boardroom

Cabiddu, F., Moi, L., Patriotta, G., & Allen, D. G. (2022). Why do users trust algorithms? A review and conceptualization of initial trust and trust over time. *European Management Journal, 40*(5), 685–706.

Chandra, S., Shirish, A., & Srivastava, S. C. (2022). To be or not to be ··· human? Theorizing the role of human-like competencies in conversational artificial intelligence agents. *Journal of Management Information Systems, 39*(4), 969–1005.

Chatterjee, S., Chaudhuri, R., Vrontis, D., Thrassou, A., & Ghosh, S. K. (2021). Adoption of artificial intelligence-integrated CRM systems in agile organizations in India. *Technological Forecasting and Social Change, 168*, Article 120783.

Chen, Q. Q., & Yi, Y. (2024). Mindsets and mirrors: How growth mindsets shape anthropomorphism in AI-enabled technologies. *Psychology & Marketing, 41*(12), 3072–3090.

Cheng, X., & Zhang, L. (2025). Inspiration booster or creative fixation? The dual mechanisms of LLMs in shaping individual creativity in tasks of different complexity. *Humanities and Social Sciences Communications, 12*(1), Article 1563.

Chi, O. H., Chi, C. G., Gursoy, D., & Nunkoo, R. (2023). Customers' acceptance of artificially intelligent service robots: The influence of trust and culture. *International Journal of Information Management, 70*, Article 102623.

Chi, O. H., Jia, S., Li, Y., & Gursoy, D. (2021). Developing a formative scale to measure consumers' trust toward interaction with artificially intelligent (AI) social robots in service delivery. *Computers in Human Behavior, 118*, Article 106700.

Childers, T. L., Carr, C. L., Peck, J., & Carson, S. (2001). Hedonic and utilitarian motivations for online retail

shopping behavior. *Journal of Retailing, 77*(4), 511–535.

Choudhary, V., Marchetti, A., Shrestha, Y. R., & Puranam, P. (2025). Human-AI ensembles: When can they work? *Journal of Management, 51*(2), 536–569.

Choung, H., David, P., & Ross, A. (2023). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human-Computer Interaction, 39*(9), 1727–1739.

Chowdhury, S., Budhwar, P., Dey, P. K., Joel-Edgar, S., & Abadie, A. (2022). AI-employee collaboration and business performance: Integrating knowledge-based view, socio-technical systems and organisational socialisation framework. *Journal of Business Research, 144*, 31–49.

Chowdhury, S., Dey, P., Joel-Edgar, S., Bhattacharya, S., Rodriguez-Espindola, O., Abadie, A., & Truong, L. (2023). Unlocking the value of artificial intelligence in human resource management through AI capability framework. *Human Resource Management Review, 33*(1), Article 100899.

Conchie, S. M., Taylor, P. J., & Donald, I. J. (2012). Promoting safety voice with safety-specific transformational leadership: The mediating role of two dimensions of trust. *Journal of Occupational Health Psychology, 17*(1), 105–115.

Croes, E. A. J., & Antheunis, M. L. (2021). Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships, 38*(1), 279–300.

Dang, J., & Liu, L. (2022a). Implicit theories of the human mind predict competitive and cooperative responses to AI robots. *Computers in Human Behavior, 134*, Article 107300.

Dang, J., & Liu, L. (2022b). A growth mindset about human minds promotes positive responses to intelligent technology. *Cognition, 220*, Article 104985.

Dang, Q., & Li, G. (2026). Unveiling trust in AI: The interplay of antecedents, consequences, and cultural dynamics. *Ai & Society, 41*, 669–692.

Daniel, V. L., & Zhan, Y. (2023). Wearing different hats enriches "outside the box"thinking: Examining the relationship between personal life activity breadth and creativity at work. *Journal of Applied Psychology, 108*(11), 1881–1901.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Davis, F. D., & Granić, A. (2024). *The technology acceptance model: 30 years of TAM.* Springer Cham.

de Visser, E. J., Peeters, M. M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R., & Neerincx, M. A. (2020).

Towards a theory of longitudinal trust calibration in human-robot teams. *International Journal of Social Robotics, 12*(2), 459–478.

Ding, S., Hu, L., Pan, X., Liu, J., & Guo, F. (2026). Probabilistic risk uncertainty assessment for driver over-trust and under-trust in level 3 human-automated driving systems cooperative driving based on the drift-diffusion model. *Reliability Engineering & System Safety, 271*, Article 112212.

Dong, X., Jiang, L., Li, W., Chen, C., Gan, Y., Xia, J., & Qin, X. (2025). Let' s talk about AI: Talking about AI is positively associated with AI crafting. *Asia Pacific Journal of Management, 42*, 1453–1484.

Duggan, J., Sherman, U., Carbery, R., & McDonnell, A. (2020). Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. *Human Resource Management Journal, 30*(1), 114–132.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2013). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (2nd ed.). Psychology Press.

Dutta, D., Mishra, S. K., & And Tyagi, D. (2023). Augmented employee voice and employee engagement using artificial intelligence-enabled chatbots: A field study. *The International Journal of Human Resource Management, 34*(12), 2451–2480.

Dutton, J. E., Dukerich, J. M., & Harquail, C. V. (1994). Organizational images and member identification. *Administrative Science Quarterly, 39*(2), 239–263.

Dweck, C. S. (2006). *Mindset: The new psychology of success.* Random House.

Dweck, C. S. (2012). Mindsets and human nature: Promoting change in the middle east, the schoolyard, the racial divide, and willpower. *American Psychologist, 67*(8), 614–622.

Einola, K., & Khoreva, V. (2023). Best friend or broken tool? Exploring the co-existence of humans and artificial intelligence in the workplace ecosystem. *Human Resource Management, 62*(1), 117–135.

Erengin, T., Briker, R., & de Jong, S. B. (in press). You, me, and the AI: The role of third-party human teammates for trust formation toward AI teammates. *Journal of Organizational Behavior*.

Forgas, J. P. (2008). Affect and cognition. *Perspectives on Psychological Science, 3*(2), 94–101.

Fraune, M. R., Sherrin, S., Šabanović, S., & Smith, E. R. (2019). Is human-robot interaction more competitive between groups than between individuals? In J. Kim, D. Sirkin, A. Tapus, M. Jung, & S. S. Kwak (Eds), *HRI' 19 The 14th ACM/IEEE international conference on human-robot interaction* (pp. 104–113). Institute of Electrical and Electronics Engineers.

Freisinger, E., & Schneider, S. (2025). Decoding decision delegation to artificial intelligence: A mixed-methods

study on the preferences of decision-makers and decision-affected in surrogate decision contexts. *European Management Journal, 43*(6), 958–969.

Fügener, A., Walzner, D. D., & Gupta, A. (2026). Roles of artificial intelligence in collaboration with humans: Automation, augmentation, and the future of work. *Management Science, 72*(1), 538–557.

Gefen, D. (2000). E-commerce: The role of familiarity and trust. *Omega, 28*(6), 725–737.

Gillespie, N., Lockey, S., Curtis, C., Pool, J., & Akbari, A. (2023). *Trust in artificial intelligence: A global study.* The University of Queensland and KPMG.

Gillespie, N., Lockey, S., Ward, T., Macdade, A., & Hassed, G. (2025). *Trust, attitudes and use of artificial intelligence: A global study 2025.* The University of Melbourne and KPMG.

Gkinko, L., & Elbanna, A. (2023). Designing trust: The formation of employees' trust in conversational AI in the digital workplace. *Journal of Business Research, 158*, Article 113707.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660.

Grant, A. M., & Sumanth, J. J. (2009). Mission possible? The performance of prosocially motivated employees depends on manager trustworthiness. *Journal of Applied Psychology, 94*(4), 927–944.

Haesevoets, T., De Cremer, D., Dierckx, K., & Van Hiel, A. (2021). Human-machine collaboration in managerial decision making. *Computers in Human Behavior, 119*, Article 106730.

Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust—the case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change, 105*, 105–120.

Hentout, A., Aouache, M., Maoudj, A., & Akli, I. (2019). Human-robot interaction in industrial collaborative robotics: A literature review of the decade 2008–2017. *Advanced Robotics, 33*(15-16), 764–799.

Heyder, T., Passlack, N., & Posegga, O. (2023). Ethical management of human-AI interaction: Theory development review. *The Journal of Strategic Information Systems, 32*(3), Article 101772.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104–121.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407–434.

Hu, P., Lu, Y., & Gong, Y. Y. (2021). Dual humanness and trust in conversational AI: A person-centered approach. *Computers in Human Behavior, 119*, Article 106727.

Huang, R., Kim, M., & Lennon, S. (2022). Trust as a second-order construct: Investigating the relationship

between consumers and virtual agents. *Telematics and Informatics, 70,* Article 101811.

Huo, W., Zheng, G., Yan, J., Sun, L., & Han, L. (2022). Interacting with medical artificial intelligence: Integrating self-responsibility attribution, human-computer trust, and personality. *Computers in Human Behavior, 132,* Article 107253.

Huynh, M. (in press). Using generative AI as decision-support tools: Unraveling users' trust and AI appreciation. *Journal of Decision Systems.*

Inga, J., Ruess, M., Robens, J. H., Nelius, T., Rothfuß, S., Kille, S., ···Kiesel, A. (2023). Human-machine symbiosis: A multivariate perspective for physically coupled human-machine systems. *International Journal of Human-Computer Studies, 170,* Article 102926.

Jarrahi, M. H., Askay, D., Eshraghi, A., & Smith, P. (2023). Artificial intelligence and knowledge management: A partnership between human and AI. *Business Horizons, 66*(1), 87–99.

Jia, N., Luo, X., Fang, Z., & Liao, C. (2024). When and how artificial intelligence augments employee creativity. *Academy of Management Journal, 67*(1), 5–32.

Jones, S. L., & Shah, P. P. (2016). Diagnosing the locus of trust: A temporal perspective for trustor, trustee, and dyadic influences on perceived trustworthiness. *Journal of Applied Psychology, 101*(3), 392–414.

Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2023). Trust in artificial intelligence: Meta-analytic findings. *Human Factors, 65*(2), 337–359.

Kim, J., Merrill Jr., K., & Collins, C. (2021). AI as a friend or assistant: The mediating role of perceived usefulness in social AI vs. Functional AI. *Telematics and Informatics, 64,* Article 101694.

Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: The effects of apology versus denial for repairing competence- versus integrity-based trust violations. *Journal of Applied Psychology, 89*(1), 104–118.

King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management, 43*(6), 740–755.

Komiak, S. Y. X., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quarterly, 30*(4), 941–960.

Kong, H., Yin, Z., Baruch, Y., & Yuan, Y. (2023). The impact of trust in AI on career sustainability: The role of employee-AI collaboration and protean career orientation. *Journal of Vocational Behavior, 146,* Article 103928.

Korsgaard, M. A., Cooper, C. D., Mayer, K. J., Poppo, L., & Zaheer, A. (2025). The boundaries of trust in a new

era. *Academy of Management Review, 50*(4), 687–697.

Kramer, R. M., & Lewicki, R. J. (2010). Repairing and enhancing trust: Approaches to reducing organizational trust deficits. *Academy of Management Annals, 4*(1), 245–277.

Kraus, J., Scholz, D., & Baumann, M. (2020.) What' s driving me? Exploration and validation of a hierarchical personality model for trust in automated driving. *Human Factors, 63*(6), 1076–1105.

Küper, A., & Krämer, N. (2025). Psychological traits and appropriate reliance: Factors shaping trust in AI. *International Journal of Human–Computer Interaction, 41*(7), 4115–4131.

Ladak, A., Loughnan, S., & Wilks, M. (2024). The moral psychology of artificial intelligence. *Current Directions in Psychological Science, 33*(1), 27–34.

Lalot, F., & Bertram, A. (2025). When the bot walks the talk: Investigating the foundations of trust in an artificial intelligence (AI) chatbot. *Journal of Experimental Psychology: General, 154*(2), 533–551.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80.

Lee, J. I., Dirks, K. T., & Campagna, R. L. (2023). At the heart of trust: Understanding the integral relationship between emotion and trust. *Group & Organization Management, 48*(2), 546–580.

Legood, A., van der Werff, L., Lee, A., den Hartog, D., & van Knippenberg, D. (2023). A critical review of the conceptualization, operationalization, and empirical literature on cognition-based and affect-based trust. *Journal of Management Studies, 60*(2), 495–537.

Legood, A., van der Werff, L., Lee, A., & den Hartog, D. (2021). A meta-analysis of the role of trust in the leadership- performance relationship. *European Journal of Work and Organizational Psychology, 30*(1), 1–22.

Lehmann, C. A., Haubitz, C. B., Fügener, A., & Thonemann, U. W. (2022). The risk of algorithm transparency: How algorithm complexity drives the effects on the use of advice. *Production and Operations Management, 31*(9), 3419–3434.

Levin, D. Z., Whitener, E. M., & Cross, R. (2006). Perceived trustworthiness of knowledge sources: The moderating impact of relationship length. *Journal of Applied Psychology, 91*(5), 1163–1171.

Levin, D. Z., & Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science, 50*(11), 1477–1490.

Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management, 32*(6), 991–1022.

Lewicki, R. J., & Bunker, B. B. (1996). Developing and maintaining trust in work relationships. In R. M. Kramer,

& T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 114–139). Sage Publications.

Lewis, M., Sycara, K., & Walker, P. (2018). The role of trust in human-robot interaction. In H. A. Abbass, J. Scholz, & D. J. Reid (Eds.), *Foundations of trusted autonomy*, Vol. 117 (pp. 135–159). Springer International Publishing.

Li, C., Lin, C., & Liu, J. (2018). The role of team regulatory focus and team learning in team radical and incremental creativity. *Group & Organization Management, 44*(6), 1036–1066.

Li, Z., & Zhou, Y. (2025). Starting with trust: Unraveling the impact of AI trust on employee digital performance. *Baltic Journal of Management, 20*(5), 637–655.

Liao, L. (2008). Knowledge-sharing in R&D departments: A social power and social exchange theory perspective. *International Journal of Human Resource Management, 19*(10), 1881–1895.

Liu, R., & Yin, H. (2024). How algorithmic management influences gig workers' job crafting. *Behavioral Sciences, 14*(10), Article 952.

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.

Lu, L., & Yan, B. (in press). Syncing minds and machines: Hybrid cognitive alignment as an emergent coordination mechanism in human-AI collaboration. *Academy of Management Review*.

Madjar, N., Greenberg, E., & Chen, Z. (2011). Factors for radical creativity, incremental creativity, and routine, noncreative performance. *Journal of Applied Psychology, 96*(4), 730–743.

Magni, D., Del Gaudio, G., Papa, A., & Della Corte, V. (2023). Digital humanism and artificial intelligence: The role of emotions beyond the human-machine interaction in society 5.0. *Journal of Management History, 30*(2), 195–218.

Malle, B. F., & Ullman, D. (2021). A multidimensional conception and measure of human-robot trust. In C. S. Nam, & J. B. Lyons (Eds.), *Trust in human-robot interaction* (pp. 3–25). Elsevier Academic Press.

Manchon, J. B., Bueno, M., & Navarro, J. (2022). How the initial level of trust in automated driving impacts drivers' behaviour and early trust construction. *Transportation Research Part F: Traffic Psychology and Behaviour, 86*, 281–295.

Marikyan, D., Papagiannidis, S., Rana, O. F., Ranjan, R., & Morgan, G. (2022). "Alexa, let' s talk about my productivity" : the impact of digital assistants on work productivity. *Journal of Business Research, 142*, 572–584.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709–735.

Mayer, R. C., & Gavin, M. B. (2005). Trust in management and performance: Who minds the shop while the employees watch the boss? *Academy of Management Journal, 48*(5), 874–888.

McAllister, D. J. (1995). Affect-based and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal, 38*(1), 24–59.

McEvily, B. (2011). Reorganizing the boundaries of trust: From discrete alternatives to hybrid forms. *Organization Science, 22*(5), 1266–1276.

McKee, K. R., Bai, X., & Fiske, S. T. (2023). Humans perceive warmth and competence in artificial intelligence. *Iscience, 26*(8), Article 107256.

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy of Management Review, 23*(3), 473–490.

McKnight, D. H., & Chervany, N. L. (2006). Reflections on an initial trust-building model. In R. Bachmann, & A. Zaheer (Eds.), *Handbook of trust research* (pp. 29–51). Edward Elgar.

Möllering, G. (2006). *Trust: Reason, routine, reflexivity*. Elsevier.

Montag, C., Kraus, J., Baumann, M., & Rozgonjuk, D. (2023). The propensity to trust in (automated) technology mediates the links between technology self-efficacy and fear and acceptance of artificial intelligence. *Computers in Human Behavior Reports, 11*, Article 100315.

Moussawi, S., & Benbunan-Fich, R. (2021). The effect of voice and humour on users' perceptions of personal intelligent agents. *Behaviour & Information Technology, 40*(15), 1603–1626.

Natali, C., Marconi, L., Dias Duran, L. D., & Cabitza, F. (2025). AI-induced deskilling in medicine: A mixed-method review and research agenda for healthcare and beyond. *Artificial Intelligence Review, 58*(11), Article 356.

Ng, S. W. T., & Zhang, R. (2025). Trust in AI chatbots: A systematic review. *Telematics and Informatics, 97*, Article 102240.

Norlander, P., Jukic, N., Varma, A., & Nestorov, S. (2021). The effects of technological supervision on gig workers: Organizational control and motivation of Uber, taxi, and limousine drivers. *The International Journal of Human Resource Management, 32*(19), 4053–4077.

Oksanen, A., Savela, N., Latikka, R., & Koivula, A. (2020). Trust toward robots and artificial intelligence: An experimental approach to human-technology interactions online. *Frontiers in Psychology, 11*, Article 568256.

Olan, F., Ogiemwonyi Arakpogun, E., Suklan, J., Nakpodia, F., Damij, N., & Jayawickrama, U. (2022). Artificial intelligence and knowledge sharing: Contributing factors to organizational performance. *Journal of Business Research, 145*, 605–615.

Omrani, N., Rivieccio, G., Fiore, U., Schiavone, F., & Agreda, S. G. (2022). To trust or not to trust? An assessment of trust in AI-based systems: Concerns, ethics and contexts. *Technological Forecasting and Social Change, 181*, Article 121763.

Othman, U., & Yang, E. (2023). Human-robot collaborations in smart manufacturing environments: Review and outlook. *Sensors, 23*(12), Article 5663.

Park, K., & Yoon, H. Y. (2024). Beyond the code: The impact of AI algorithm transparency signaling on user trust and relational satisfaction. *Public Relations Review, 50*(5), Article 102507.

Passini, S. (2016). Concern for close or distant others: The distinction between moral identity and moral inclusion. *Journal of Moral Education, 45*(1), 74–86.

Passini, S., & Morselli, D. (2017). Construction and validation of the moral inclusion/exclusion of other groups (MIEG) scale. *Social Indicators Research, 134*(3), 1195–1213.

Pentina, I., Xie, T., Hancock, T., & Bailey, A. (2023). Consumer-machine relationships in the age of artificial intelligence: Systematic literature review and research directions. *Psychology & Marketing, 40*(8), 1593–1614.

Perry, A. (2023). AI will never convey the essence of human empathy. *Nature Human Behaviour, 7*(11), 1808–1809.

Petrou, P., & Jongerling, J. (2024). Incremental and radical creativity in dealing with a crisis at work. *Creativity Research Journal, 36*(2), 378–394.

Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: The theory, design, and analysis of change. *Journal of Management, 36*(1), 94–120.

Qi, T., Liu, H., & Huang, Z. (2025). An assistant or a friend? The role of parasocial relationship of human-computer interaction. *Computers in Human Behavior, 167*, Article 108625.

Qin, H., Zhu, Y., Jiang, Y., Luo, S., & Huang, C. (2024). Examining the impact of personalization and carefulness in AI-generated health advice: Trust, adoption, and insights in online healthcare consultations experiments. *Technology in Society, 79*, Article 102726.

Qin, X., Zhou, X., Chen, C., Wu, D., Zhou, H., Dong, X.,⋯Lu, J. G. (2025). AI aversion or appreciation? A capability-personalization framework and a meta-analytic review. *Psychological Bulletin, 151*(5), 580–599.

Ren, F., & Song, Z. (2024). Employee radical and incremental creativity: A systematic review. *The Journal of*

Creative Behavior, 58(2), 297–308.

Riedl, R. (2022). Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions. *Electronic Markets, 32*(4), 2021–2051.

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.

Saffarizadeh, K., Keil, M., & Maruping, L. (2024). Relationship between trust in the AI creator and trust in AI systems: The crucial role of AI alignment and steerability. *Journal of Management Information Systems, 41*(3), 645–681.

Salah, M., Alhalbusi, H., Ismail, M. M., & Abdelfattah, F. (2024). Chatting with ChatGPT: Decoding the mind of chatbot users and unveiling the intricate connections between user perception, trust and stereotype perception on self-esteem and psychological well-being. *Current Psychology, 43*(9), 7843–7858.

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors, 58*(3), 377–400.

Sedlakova, J., & Trachsel, M. (2023). Conversational artificial intelligence in psychotherapy: A new therapeutic tool or agent? *American Journal of Bioethics, 23*(5), 4–13.

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G., Elkins, A.,⋯ Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management, 57*(2), Article 103174.

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review, 61*(4), 66–83.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99–118.

Singh, N., & Sinha, N. (2020). How perceived trust mediates merchant' s intention to use a mobile wallet technology. *Journal of Retailing and Consumer Services, 52*, Article 101894.

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzæg, P. B. (2021). My chatbot companion—a study of human-chatbot relationships. *International Journal of Human-Computer Studies, 149*, Article 102601.

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzæg, P. B. (2022). A longitudinal study of human-chatbot relationships. *International Journal of Human-Computer Studies, 168*, Article 102903.

Song, J., & Lin, H. (2024). Exploring the effect of artificial intelligence intellect on consumer decision delegation: The role of trust, task objectivity, and anthropomorphism. *Journal of Consumer Behaviour, 23*(2), 727–747.

Sowa, K., Przegalinska, A., & Ciechanowski, L. (2021). Cobots in knowledge work: Human-AI collaboration in

managerial professions. *Journal of Business Research, 125*, 135–142.

Suseno, Y., Chang, C., Hudik, M., & Fang, E. S. (2022). Beliefs, anxiety and change readiness for artificial intelligence adoption among human resource managers: The moderating role of high-performance work systems. *International Journal of Human Resource Management, 33*(6), 1209–1236.

Taylor, A., & Greve, H. R. (2006). Superman or the fantastic four? Knowledge combination and experience in innovative teams. *Academy of Management Journal, 49*(4), 723–740.

Telkamp, J. B., & Anderson, M. H. (2022). The implications of diverse human moral foundations for assessing the ethicality of artificial intelligence. *Journal*

of Business Ethics, 178(4), 961–976.

Tomlinson, E. C., Schnackenberg, A. K., Dawley, D., & Ash, S. R. (2020). Revisiting the trustworthiness-trust relationship: Exploring the differential predictors of cognition- and affect-based trust. *Journal of Organizational Behavior, 41*(6), 535–550.

Topsakal, Y. (2025). How familiarity, ease of use, usefulness, and trust influence the acceptance of generative artificial intelligence (AI)-assisted travel planning. *International Journal of Human-Computer Interaction, 41*(15), 9478–9491.

Townsend, D. M., Hunt, R. A., Rady, J., Manocha, P., & Jin, J. H. (2025). Are the futures computable? Knightian uncertainty and artificial intelligence. *Academy of Management Review, 50*(2), 415–440.

Tu, Y., Li, J., Chen, J., Li, C., & He, W. (in press). When AI becomes my teammate: Unpacking how employees perceive and collaborate with gendered AI teammates. *Journal of Organizational Behavior*.

Ulfert, A., Georganta, E., Centeio Jorge, C., Mehrotra, S., & Tielman, M. (2024). Shaping a multidisciplinary understanding of team trust in human-AI teams: A theoretical framework. *European Journal of Work and Organizational Psychology, 33*(2), 158–171.

Ullrich, D., Butz, A., & Diefenbach, S. (2021). The development of overtrust: An empirical simulation and psychological analysis in the context of human-robot interaction. *Frontiers in Robotics and AI, 8*, Article 554578.

Valori, I., Kraus, J., & Fairhurst, M. T. (2026). Interdisciplinary perspectives and current findings on the role of trust as a psychological mediator in human interaction with artificial intelligence: Editorial overview. *Computers in Human Behavior, 180*, Article 108957.

van der Werff, L., Legood, A., Buckley, F., Weibel, A., & de Cremer, D. (2019). Trust motivation: The self-regulatory processes underlying trust decisions. *Organizational Psychology Review, 9*(2-3), 99–123.

van der Werff, L., & Buckley, F. (2017). Getting to know you: A longitudinal examination of trust cues and trust

development during socialization. *Journal of Management, 43*(3), 742–770.

van Knippenberg, D. (2018). Reconsidering affect-based trust: A new research agenda. In R. H. Searle, A. I. Nienaber, & S. B. Sitkin (Eds.), *The Routledge companion to trust* (pp.3–13). Taylor & Francis.

Vanneste, B. S., & Puranam, P. (2025). Artificial intelligence, trust, and perceptions of agency. *Academy of Management Review, 50*(4), 726–744.

Vuori, N., Burkhard, B., & Pitkäranta, L. (2026). It' s amazing—but terrifying!: Unveiling the combined effect of emotional and cognitive trust on organizational

member' behaviours, AI performance, and adoption. *Journal of Management Studies, 63*(2), 473–514.

Wang, P., & Ding, H. (2024). The rationality of explanation or human capacity? Understanding the impact of explainable artificial intelligence on human-AI trust and decision performance. *Information Processing & Management, 61*(4), Article 103732.

Wang, L., Gao, R., Váncza, J., Krüger, J., Wang, X. V., Makris, S., & Chryssolouris, G. (2019). Symbiotic human-robot collaborative assembly. *CIRP Annals - Manufacturing Technology, 68*(2), 701–726.

Wang, W., Gao, G. G., & Agarwal, R. (2024). Friend or foe? Teaming between artificial intelligence and workers with variation in experience. *Management Science, 70*(9), 5753–5775.

Wang, W., Pei, S., & Sun, T. (in press). Unraveling generative AI from a human intelligence perspective: A battery of experiments. *Information Systems Research.*

Wang, W., Qiu, L., Kim, D., & Benbasat, I. (2016). Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. *Decision Support Systems, 86*, 48–60.

Weber, J. M., Malhotra, D., & Murnighan, J. K. (2004). Normal acts of irrational trust: Motivated attributions and the trust development process. *Research in Organizational Behavior, 26*(4), 75–101.

Weisz, E., Herold, D. M., Ostern, N. K., Payne, R., & Kummer, S. (2025). Artificial intelligence (AI) for supply chain collaboration: Implications on information sharing and trust. *Online Information Review, 49*(1), 164–181.

Williams, M. (2001). In whom we trust: Group membership as an affective context for trust development. *Academy of Management Review, 26*(3), 377–396.

Wirz, C. D., Demuth, J. L., Bostrom, A., Cains, M. G., Ebert-Uphoff, I., Gagne, D. J.,···Madlambayan, D. (2025). (Re)conceptualizing trustworthy AI: A foundation for change. *Artificial Intelligence, 342*, Article 104309.

Wykowska, A. (2021). Robots as mirrors of the human mind. *Current Directions in Psychological Science, 30*(1), 34–40.

Xu, S., & Li, W. (2022). A tool or a social being? A dynamic longitudinal investigation of functional use and

relational use of AI voice assistants. *New Media & Society, 26*(7), 3912–3930.

Yam, K. C., Bigman, Y. E., Tang, P. M., Ilies, R., De Cremer, D., Soh, H., & Gray, K. (2021). Robots at work: People prefer—and forgive—service robots with perceived feelings. *Journal of Applied Psychology, 106*(10), 1557–1572.

Yin, Z., Kong, H., Baruch, Y., L' Espoir Decosta, P., & Yuan, Y. (2024). Interactive effects of AI awareness and change-oriented leadership on employee-AI collaboration: The role of approach and avoidance motivation. *Tourism Management, 105*, Article 104966.

Yu, M., & Choi, J. N. (2022). How do feedback seekers think? Disparate cognitive pathways towards incremental and radical creativity. *European Journal of Work and Organizational Psychology, 31*(3), 470–483.

Yue, B., & Li, H. (2023). The impact of human-AI collaboration types on consumer evaluation and usage intention: A perspective of responsibility attribution. *Frontiers in Psychology, 14*, Article 1277861.

The dynamics of human trust in AI from the instrumental and value perspectives

SONG Yu1, HU Xiaoran2

(1 School of Economics and Management, Southeast University, Nanjing 211189, China)

(2 Department of Management, The London School of Economics and Political Science, London WC2A 2AE, UK)

Abstract: With the rapid development of artificial intelligence (AI) technology, human-AI relationships have become increasingly prevalent and consequential in organizations. Human trust in AI lies at the core of human-AI relationships and is critical to the effectiveness of human-AI interactions. Key challenges in research on human trust in AI include how to conceptualize trust, understand dynamic patterns of human-AI relationships, and achieve complementary advantages through human-AI interactions. This study addresses these issues by focusing on the dyadic interaction between humans and AI to explore the dynamic processes of human trust in AI over time. First, drawing on the perspective of technological ethics, this study conceptualizes human trust in AI as a two-dimensional construct comprising instrumental trust and value trust, and further develops a corresponding measurement scale. Second, adopting a dynamic development perspective, the study explores the temporal characteristics and dynamic patterns of human trust in AI, thereby opening the "black box" of trust dynamics in human-AI relationships. Finally, from the perspective of human-AI collaboration, the study investigates the effect of different forms of human trust in AI on employee creativity, offering a nuanced understanding of human-AI relationship development and providing insights into how trust in AI shapes employees' core competencies in the digital intelligence era.

Keywords: human trust in AI, instrumental trust, value trust, human–AI relationship, dynamics

Note: Figure translations are in progress. See original paper for figures.