Management Decision

The gender gap in AI change: exploring disparities in emotional responses

Daniel Pittino
Jönköping International Business School, Jönköping University, Jönköping, Sweden and
Dipartimento di Scienze Economiche e Statistiche, Universita degli Studi di Udine, Udine, Italy, and
Annika Engström, Nina Edh, Anette Johansson and Alice Mohlin
School of Engineering, Jönköping University, Jönköping, Sweden

Received 31 October 2025
Revised 12 February 2026
12 March 2026
Accepted 13 March 2026

Abstract

Purpose – This paper examines how gender shapes employees' emotional reactions to workplace AI, and whether these differences are associated with epistemic legitimacy, explained by perceived AI knowledge and conditioned by psychological safety.

Design/methodology/approach – We analyse survey data from 104 employees using a conditional process framework. Multiple regression models test direct gender effects on cognitive and emotional reactions to AI, mediation via perceived AI knowledge, and moderation by team-level psychological safety.

Findings – Women report less favourable cognitive and emotional reactions to AI than men. These differences are largely attributable to lower perceived AI knowledge. Psychological safety attenuates the direct gender effect: in high-safety climates, gender gaps in reactions diminish; in low-safety climates, they re-emerge. Overall, the pattern supports a contextualised partial-mediation model in which perceived knowledge is pivotal, but its explanatory power depends on climate.

Research limitations/implications – The cross-sectional design limits causal inference and the generalisability is bounded by the organisational context studied. Future research should use longitudinal or experimental designs, examine additional inequality dimensions (e.g. age, role), and unpack how AI literacy interventions reshape appraisal dynamics over time.

Practical implications – AI initiatives should build employees' perceived understanding (how AI works, limits, and human–AI complementarity) and foster psychological safety so that questions and uncertainty are acceptable. Monitoring gendered participation and confidence during roll-out helps prevent AI from amplifying existing inequalities.

Social implications – Managing AI adoption as an inclusion challenge—rather than solely a technical one—can reduce uneven emotional costs of digital transformation and support fairer access to AI-enabled opportunities.

Originality/value – The study integrates gender, appraisal (perceived AI knowledge), and climate (psychological safety) into a single framework explaining both cognitive and emotional reactions to AI. It reframes gender gaps as contextual, highlighting levers – literacy and climate – that organisations can use to enable more equitable AI adoption.

Keywords Gender, Psychological safety, Emotional reaction, AI adoption, Perceived AI knowledge

Paper type Research article

Introduction

Artificial intelligence (AI) is increasingly embedded as a structural component of contemporary organizations, shaping how work is performed, coordinated, and evaluated. Firms are integrating AI into a wide range of core functions, including operations, product development,

Funding: The authors acknowledge the Knowledge Foundation, Jönköping University and the industrial partners for funding the research and education environment on Knowledge Intensive Product Realization SPARK at Jönköping University, Sweden (Project: AFAIR; No: 20200223).

Management Decision
Emerald Publishing Limited
e-ISSN: 1758-6070
p-ISSN: 0025-1747
DOI 10.1108/MD-10-2025-3371

logistics, accounting, and risk management (Enholm *et al.*, 2021; McKinsey, 2023). Despite this rapid diffusion, organizational outcomes remain uneven. While more than 80% of companies report having implemented AI in at least one core process, only a minority achieve sustained and measurable performance gains (Challapally *et al.*, 2025). This discrepancy has prompted growing scholarly attention to the dark side of AI transformation, namely the (often subtle and unintended) negative organizational consequences that arise as AI systems become embedded in everyday work practices (Tarafdar *et al.*, 2019; Benlian *et al.*, 2022).

Research on the dark side of digital transformation shows that advanced technologies can generate anxiety, alienation, and inequality, even when they function as intended (Tarafdar *et al.*, 2019; Kim *et al.*, 2025). In the context of AI, these dynamics are particularly pronounced. Algorithmic decision-making can erode perceived meaning, agency, and transparency at work, producing what has been described as an "emotional undercurrent" of digitalization (Kim *et al.*, 2025). Studies of algorithmic bias and fairness further demonstrate that AI systems often reproduce or amplify existing social asymmetries rather than neutralizing them (Ntoutsi *et al.*, 2020; Mehrabi *et al.*, 2021). At the same time, research on technostress documents heightened emotional strain associated with continuous learning demands, uncertainty, and reduced perceived autonomy in AI-enabled workplaces (Tarafdar *et al.*, 2019; Cheng *et al.*, 2022b). In particular, although organizations have long experienced waves of information technology (IT) change, AI exhibits features that may amplify dynamics already documented in prior research on technology adoption, in ways that are theoretically and practically relevant. While earlier enterprise systems and digital tools also posed challenges related to complexity and user adaptation, contemporary AI systems are characterized by two interrelated features that are particularly pronounced. First, many AI systems operate through complex, data-driven models whose internal logic is difficult for users to fully observe, interpret, or contest, resulting in persistent epistemic opacity (Afroogh *et al.*, 2024; Gazit, 2026; Mitchell, 2025). Earlier digital technologies also required users to acquire skills and adapt processes, but their logic, once learned, was relatively stable and inspectable. Contemporary AI systems tend to rely on adaptive, probabilistic, and data-intensive models whose functioning can remain only partially transparent even after training and experience (Burrell, 2016; Rai, 2020; Schilke and Reimann, 2025). This does not imply that opacity is absent in other technologies, but rather that it is more persistent and structurally embedded in many AI applications. Second, AI systems are increasingly perceived as agentic in organizational settings, in the sense that they autonomously generate recommendations, classifications, or predictions that shape decisions and actions, rather than merely executing predefined human instructions (Faraj *et al.*, 2018; Glikson and Woolley, 2020). As a result, AI transformation introduces heightened epistemic uncertainty: employees must rely on systems that shape decisions and outcomes while limiting their ability to fully explain, challenge, or correct them (Bennett, 2025). This shift can destabilize established understandings of expertise and control, rendering AI implementation a potentially more emotionally charged organizational process than technology upgrades where uncertainty diminishes with familiarity (Glikson and Woolley, 2020). Taken together, this literature suggests that AI transformation may generate emotional costs that are unevenly distributed across organizational members.

Despite these insights, we still lack an integrated understanding of why emotional reactions to AI transformation vary systematically across individuals. Technology adoption research has traditionally emphasized cognitive evaluations such as perceived usefulness, ease of use, and facilitating conditions (Venkatesh *et al.*, 2003). While these models have been extended to incorporate attitudes and affect, they offer limited leverage for explaining emotional reactions that arise when technologies both resist full transparency and appear to act with a degree of autonomy. As a result, emotions are often treated as secondary outcomes rather than as central mechanisms through which the dark side of AI transformation unfolds.

This limitation becomes particularly salient when considering gender. A substantial body of research documents persistent gender differences in technology-related confidence, self-efficacy, and participation, especially in domains associated with technical expertise

(Eagly and Karau, 2002). Social role theory explains these patterns by pointing to gendered expectations regarding competence, agency, and expertise, which shape both self-evaluations and social judgments (Eagly and Wood, 2012). Recent studies extend these insights to AI, showing that women report lower perceived AI literacy and greater apprehension toward AI adoption at work (Otis *et al*., 2024; Russo *et al*., 2025). However, the emotional implications of these gendered self-evaluations in AI transformation contexts remain conceptually fragmented and empirically underexplored, particularly in settings characterized by persistent uncertainty and shifting boundaries of decision authority.

We argue that gendered dynamics become distinctively impactful in AI transformation, especially in terms of emotions, because AI destabilizes established criteria for knowing, deciding, and exercising authority. When technologies are opaque, individuals' perceived ability to understand and meaningfully engage with them becomes a key psychological resource that substitutes for direct control or explainability. Differences in such self-assessments can translate directly into emotional reactions, such as anxiety, frustration, or disengagement, that shape how AI is experienced and enacted at work (Russo *et al*., 2025).

To examine these dynamics, the present study integrates technology adoption research and social role theory to investigate and theorize on how gender relates to employees' emotional reactions to AI-driven organizational change.

Drawing on social role theory, which highlights how gendered expectations shape perceptions of competence and credibility in organizational settings (Eagly and Karau, 2002; Eagly and Wood, 2012), we focus on two interrelated mechanisms through which AI transformation can generate unequal emotional reactions. These mechanisms can be understood through the lens of epistemic legitimacy, defined as the socially constructed judgment of who is regarded as a credible and entitled knower, whose understanding, questions, or doubts are treated as appropriate in a given context (Suchman, 1995; Fricker, 2007). While epistemic legitimacy is not directly measured in our empirical model, it serves as an interpretive framework for understanding why perceived AI knowledge and psychological safety may carry gendered emotional implications in AI adoption contexts. The first mechanism is perceived AI knowledge, referring to employees' self-assessed understanding of and confidence in interacting with AI technologies (Otis *et al*., 2024; Russo *et al*., 2025). In AI contexts, perceived knowledge reflects not only technical familiarity but also a sense of epistemic control over algorithmic systems whose logic remains only partially accessible (Burrell, 2016; Rai *et al*., 2019). The second mechanism is psychological safety, defined as a shared belief that the work environment permits interpersonal risk-taking, such as admitting uncertainty or questioning AI outputs (Edmondson, 1999).

Using survey data collected in 2025 from 104 employees in two Swedish manufacturing firms undergoing early-stage AI integration, we test a moderated mediation model linking gender to emotional and cognitive reactions through perceived AI knowledge, with psychological safety as a contextual moderator. The results show that women report lower perceived AI knowledge and less positive emotional and cognitive reactions than men. Perceived AI knowledge is associated with this relationship in a pattern consistent with mediation, suggesting that gender differences in self-assessed competence may account for part of the observed emotional disparity. Psychological safety moderates the direct effect of gender on reactions: gender differences are strongest in low-safety environments and attenuated in climates characterized by trust and openness.

This study makes two main contributions. First, it advances research on the dark side of AI and digital transformation by identifying an emotional and relational mechanism through which AI implementation can reproduce inequality inside organizations, even in the absence of overt bias or technical failure (Tarafdar *et al*., 2019; Benlian *et al*., 2022). By explicitly distinguishing AI transformation from earlier IT change, the study argues how epistemic opacity and persistent uncertainty may make perceived legitimacy central to emotional reactions and therefore to adoption outcomes (Burrell, 2016; Rai *et al*., 2019; Suchman, 1995). This argument is supported indirectly by the empirical pattern of results, as epistemic opacity and legitimacy are theorized rather than directly operationalized in the present design.

Second, it contributes to social role theory (Eagly and Karau, 2002; Eagly and Wood, 2012) by providing evidence consistent with the proposition that gendered self-evaluations translate into systematically different emotional experiences in AI-enabled workplaces. By identifying psychological safety as a contextual buffer, the study highlights how greater gender-based emotional inequalities during AI transformation can either be associated to certain features of organizational climate (Edmondson, 1999; Detert and Edmondson, 2011).

Gender and technology adoption: the theoretically relevant features of AI

Research on technology acceptance, shaped by the TAM and UTAUT frameworks, has established that adoption is influenced by perceived usefulness and ease of use, and that gender systematically moderates these evaluations (Venkatesh *et al*., 2003, 2012). Men's adoption intentions tend to be driven by performance expectancy, while women's are more closely associated with ease of use, social support, and self-efficacy (Venkatesh and Davis, 2000; Ong and Lai, 2006; Im *et al*., 2011). Social role theory provides a complementary explanation: men are socialized toward agentic roles emphasizing autonomy and mastery, while women are oriented toward communal roles valuing connectedness and caution (Eagly, 1987; Eagly and Karau, 2002; Eagly and Wood, 2012). In technology contexts, this translates into men being more likely to view new systems as opportunities for demonstrating competence, while women are more attentive to disruption and the social conditions surrounding adoption (Brough *et al*., 2016). This theoretical lens is central to the present study, as it explains not only different evaluation patterns but also why uncertainty in technology contexts may carry different emotional weight depending on gendered expectations about expertise and authority.

Relatedly, it is important to note that the traditional technology adoption frameworks have focused largely on technologies whose uncertainty tends to decline with experience and whose outputs are transparent and interpretable. Recent organizational research underscores that AI may not function simply as another "next generation" IT tool, given its persistent epistemic opacity and its influence on decision authority. While acknowledging that some degree of opacity and decision influence characterizes other complex technologies as well, we argue that these features are sufficiently pronounced in AI contexts to warrant dedicated theoretical attention,

AI systems more often than other technologies operate as black boxes whose internal reasoning cannot be fully inspected or explained, making it difficult for users to form stable judgments even with experience (Mitchell, 2025; Afroogh *et al*., 2024). In some cases, greater transparency disclosures paradoxically reduce perceived legitimacy and trust (Schilke and Reimann, 2025). At the same time, AI increasingly produces recommendations that shape organizational outcomes autonomously, altering how decision authority is perceived (Zeiser, 2024). Rather than claiming categorical distinctiveness, we argue that these features are sufficiently pronounced in AI contexts to warrant dedicated theoretical attention, while acknowledging that some degree of opacity and decision influence characterizes other complex technologies as well.

These characteristics have implications for individual well-being and perceptions of fairness (Sadeghi, 2024; Ruckenstein, 2023), and, as we argue below, they create conditions under which gendered expectations about competence and authority become particularly consequential. Because opaque, decision-influencing systems do not readily reveal how or why they produce outputs, individuals must continuously interpret and judge those outputs as part of their work. This ongoing interpretation is not merely a cognitive task but a legitimacy challenge: employees must navigate who is entitled to understand, question, and potentially override system recommendations. Drawing on social role theory, such legitimacy challenges can be gendered; women, shaped by prevailing expectations around expertise and authority, may be more likely than men to experience uncertainty as a threat to competence under conditions of opacity and perceived agency. In expertise-driven contexts, displays of epistemic uncertainty are often more heavily sanctioned for women than men, reinforcing emotional strain and self-doubt (Ridgeway, 2011; Cheryan *et al*., 2015). In this way, the features that are particularly pronounced in AI contexts connect directly to gendered emotional experiences because they elevate the stakes of interpretation, judgment, and social legitimacy.

These considerations prepare the theoretical ground for understanding the mechanisms of perceived AI knowledge and psychological safety in the hypotheses that follow.

Research hypotheses

Gender and reactions to AI organizational adoption

AI organizational adoption introduces systems that are simultaneously epistemically opaque and agentic. Consequently, as AI becomes embedded in everyday work, employees are required to interact with technologies that, other than support tasks but also participate in judgment, ranking, and recommendation, thereby redistributing authority and accountability (Benlian *et al.*, 2022; Cheng *et al.*, 2022a). These features make AI-driven change an inherently affective experience, as uncertainty is no longer a temporary learning challenge but a persistent condition of working with systems that act upon organizational processes.

Emotional reactions to AI adoption are shaped by how employees experience uncertainty, exposure, and responsibility in relation to systems that influence outcomes yet resist full explanation. When AI operates in ways that are difficult to interpret or challenge, employees may experience heightened anxiety, frustration, or emotional withdrawal, particularly when accountability for outcomes remains human while control is partially delegated to the technology (Tarafdar *et al.*, 2019; Benlian *et al.*, 2022). Such emotional responses reflect concerns about how one's role, judgment, and standing are affected by working alongside opaque and decision-influencing systems.

These emotional experiences are not evenly distributed across employees. Gender differences are likely to emerge in how AI-related uncertainty is experienced and evaluated within organizational contexts. Prior research consistently shows that women report lower self-assessed AI knowledge and lower confidence in their ability to engage with AI tools (Chiu *et al.*, 2021; Otis *et al.*, 2024). Beyond differences in familiarity, research on decision-making under uncertainty suggests that women tend to be more sensitive to the social and evaluative consequences of ambiguity, particularly in contexts where responsibility for outcomes is unclear and errors are visible (Miller, 2012; Wester *et al.*, 2024). Recent evidence in the AI domain further indicates that gender differences in attitudes toward AI are shaped by differential exposure to and consequences of uncertainty in AI-enabled workplaces (Borwein and Loewen, 2026).

In AI implementation settings, the combination of opacity and perceived agency is likely to intensify these dynamics. Because AI systems can generate outputs whose rationale is only partially accessible while simultaneously shaping organizational decisions, uncertainty may become both more persistent and more directly relevant for outcomes than in settings where system logic is more transparent and decision authority more clearly retained by users. Under such conditions, employees are likely to deal with how uncertainty about AI is interpreted by others. For women, who more often operate under heightened scrutiny regarding technical competence and accountability in expertise-laden domains, AI adoption may therefore involve greater exposure to evaluation, potential deskilling, or relational disruption, leading to less positive emotional reactions. Men, by contrast, are more likely to experience AI adoption as an opportunity to align with technologically valued roles, reinforcing more positive emotional responses (Brough *et al.*, 2016; Young *et al.*, 2021).

Accordingly, we hypothesize that:

H1. Women report less positive emotional reactions to AI adoption than men.

The mediating role of perceived AI knowledge

Consistent with our theoretical framing, we argue that gendered differences in emotional reactions to AI adoption may be associated with mechanisms that structure how employees experience their relationship with AI systems during organizational implementation.

Specifically, we propose that perceived AI knowledge represents a plausible pathway through which gender relates to emotional reactions, as it may be seen as a way through which gendered expectations about competence and authority become embedded in employees' emotional responses to AI-driven change. Perceived AI knowledge refers to employees' self-assessed understanding of and confidence in engaging with AI systems (Ng *et al.*, 2024; Boob-Engel, 2025). In AI implementation contexts, this perception takes on particular significance because employees must relate to technologies whose functioning is only partially accessible while their outputs increasingly influence work outcomes. Under such conditions, perceived AI knowledge does not simply reflect familiarity with a tool, but shapes how uncertainty is subjectively experienced. When individuals perceive themselves as lacking sufficient AI knowledge, uncertainty is more likely to be internalized and emotionally charged, increasing feelings of vulnerability, apprehension, or withdrawal (Benlian *et al.*, 2022; Tarafdar *et al.*, 2019). By contrast, higher perceived AI knowledge is associated with a greater sense of orientation and emotional reassurance when engaging with AI-enabled work.

Importantly, perceived AI knowledge is systematically affected by gender. Prior research shows that women report lower perceived AI knowledge and greater apprehension toward AI-enabled work practices (Chiu *et al.*, 2021; Otis *et al.*, 2024). These differences reflect broader dynamics in how expertise and credibility are socially recognized in technical and decision-oriented domains. Research on epistemic injustice and epistemic exclusion indicates that women are more likely to encounter credibility deficits in knowledge interactions, even when they possess relevant experience, because judgments of competence are shaped by socially structured expectations about who counts as a legitimate contributor (Fricker, 2007; Settles *et al.*, 2022; Lobo, 2025). Such dynamics extend into organizational settings, where recognition of understanding and access to interpretive authority are unevenly distributed, reinforcing gendered differences in how uncertainty is emotionally experienced (Rossiter, 1993; Lobo, 2025)

Social role theory further clarifies why perceived AI knowledge operates as a mediating mechanism in AI implementation contexts. Gendered role expectations shape how individuals evaluate their own competence and how uncertainty is emotionally processed in environments where authority and accountability are salient (Eagly and Karau, 2002; Eagly and Wood, 2012).

In domains associated with technical mastery and decision authority, women are more likely to interpret lower perceived AI knowledge as a signal of personal inadequacy rather than as a normal feature of working with complex systems. This dynamic is accentuated in AI contexts, where uncertainty cannot always be resolved through explanation (opacity) and where system outputs carry authoritative weight (agency), thereby intensifying the emotional consequences of perceived knowledge gaps. Research on decision-making under uncertainty further suggests that women are more sensitive to the social and reputational implications of uncertainty, particularly when responsibility for outcomes is ambiguous and evaluative pressures are present (Miller, 2012; Wester *et al.*, 2024).

Taken together, these arguments suggest that perceived AI knowledge may function as a mediating pathway in the relationship between gender and emotional reactions to AI adoption, linking gendered expectations about legitimate competence and credibility to differentiated emotional experiences.

Accordingly, we hypothesize that:

H2. The relationship between gender and employees' reactions to AI adoption is mediated by perceived AI knowledge, such that women report lower perceived AI knowledge, which is associated to less positive emotional reactions.

The moderating role of psychological safety

Individual patterns of reaction to technology occur within specific organizational contexts that can be associated with stronger or weaker underlying differences, as they may regulate how uncertainty, competence, and emotional expression are interpreted in everyday work.

In AI adoption, employees are asked to work alongside systems that increasingly participate in judgment and decision processes, while the reasoning behind outputs may remain difficult to interrogate in practice (Afoogh *et al.*, 2024; Mitchell, 2025). This combination makes uncertainty a recurring feature of work that is not easily mitigated through training and that can be experienced as exposure rather than as neutral "not-knowing" (Sadeghi, 2024).

Psychological safety captures a central part of this context because it concerns whether interpersonal risk-taking, such as asking for clarification, admitting uncertainty, or raising concerns, can occur without fear of negative interpersonal consequences (Edmondson, 1999, 2018). Framed in relation to AI adoption, psychological safety is key, mainly because it governs whether uncertainty can be expressed and processed collectively or whether it is individualized and silently borne. Where uncertainty is normalized as an acceptable part of work, it can be metabolized through dialogue, mutual adjustment, and help-seeking; where it is stigmatized, it becomes emotionally charged and potentially reputationally damaging (Edmondson, 1999, 2018; Carmeli and Gittell, 2009; Frazier *et al.*, 2017).

This "governance" function becomes particularly salient when gendered expectations about competence and authority are taken into account. Social role theory emphasizes that workplaces carry persistent normative expectations about how men and women "should" enact competence, mastery, and emotional display (Eagly and Karau, 2002; Eagly and Wood, 2012). In technical and decision-relevant domains, these expectations interact with status beliefs about expertise and authority (Ridgeway, 2011), and they shape how uncertainty is interpreted when it is voiced. The same reaction in the form of: "I'm not sure how this works" or "I'm not convinced by this output" can be heard as caution or as lack of competence depending on who speaks, because credibility and authority are not evenly distributed in many organizational interactions (Ridgeway, 2011; Eagly and Wood, 2012).

Work on epistemic legitimacy related to injustice and epistemic exclusion sharpens this point by foregrounding how knowledge interactions themselves can be structured by credibility deficits and uneven access to interpretive authority (Fricker, 2007; Settles *et al.*, 2022; Rossiter, 1993). In organizations, this means that "having a question" is not only an informational state; it can also be a social act that invites evaluation. When interpretive authority is unevenly distributed, uncertainty becomes more than a private feeling—it becomes an interactional moment where legitimacy is granted or withheld. This is especially relevant in AI implementation, where employees may be asked to align with outputs that shape outcomes while the rationale is not fully available, making questioning and interpretation unavoidable parts of competent participation (Benlian *et al.*, 2022; Cheng *et al.*, 2022a). Under such conditions, climates differ in whether they treat questioning as contribution or as disruption.

The moderation findings indicate that gender differences in emotional reactions diminish under conditions of high psychological safety. This pattern is consistent with the idea that psychologically safe climates make it more feasible for employees to externalize uncertainty without immediate reputational cost (Edmondson, 1999, 2018; Carmeli and Gittell, 2009; Frazier *et al.*, 2017), and that when uncertainty can be voiced without sanction, gendered expectations about mastery and competence display have less opportunity to convert uncertainty into emotional strain (Eagly and Karau, 2002; Eagly and Wood, 2012).

By contrast, in low psychological safety climates, uncertainty is more likely to be individualized and internalized. When asking questions, admitting doubt, or challenging outputs is socially risky, employees may engage in self-censorship and emotion management rather than open sensemaking (Edmondson, 1999, 2018; Frazier *et al.*, 2017). In such environments, existing status and role expectations can become more consequential: those who already face tighter standards of competence display may experience a higher emotional burden from the same underlying uncertainty, because the cost of being seen as unsure is higher (Ridgeway, 2011; Eagly and Wood, 2012). This connects naturally to the AI context: when AI outputs are influential and difficult to interpret, low safety can make uncertainty feel like a competence threat rather than a normal feature of implementation, intensifying negative emotional reactions (Tarafdar *et al.*, 2019; Benlian *et al.*, 2022).

This contextual logic also aligns with how perceived AI knowledge operates as a mediating mechanism (Chiu *et al.*, 2021; Otis *et al.*, 2024). Psychological safety affects whether perceived knowledge gaps become “socially relevant”, whether they are treated as a legitimate basis for questions and learning, or as a stigmatizing marker of incompetence. In environments where uncertainty is sanctioned, the emotional meaning of “not knowing enough” becomes heavier, and gendered expectations about whose competence is presumed can make that weight uneven (Fricker, 2007; Settles *et al.*, 2022; Ridgeway, 2011). In environments where uncertainty is normalized and voiced, the same perceived gap is less likely to be translated into shame, anxiety, or withdrawal (Edmondson, 1999, 2018; Carmeli and Gittell, 2009).

Accordingly, we theorize that psychological safety shapes the extent to which gendered expectations about competence and authority are activated in affecting the experience of AI adoption, and we hypothesize that:

H3a. Psychological safety moderates the relationship between gender and employees’ emotional reactions to AI adoption, such that gender differences are weaker in contexts characterized by high psychological safety.

H3b. Psychological safety moderates the relationship between perceived AI knowledge and employees’ emotional and cognitive reactions to AI adoption, such that higher psychological safety strengthens the positive association between perceived AI knowledge and reactions.

A summary of our hypotheses structure is illustrated in Figure 1.

Method

Sample and data collection

Data for this study were collected in February 2025 as part of a broader research project examining organizational readiness for AI transformation in the manufacturing sector. The study involved two units belonging to industrial firms operating in medium- and medium–high technology manufacturing and manufacturing support in Sweden.

The data were gathered through a web-based survey platform, distributed internally to all employees via a link posted in the companies’ communication channels. The questionnaire was developed in collaboration with the participating organizations to ensure contextual relevance and was pretested with a small group of employees for clarity and comprehension before full-scale distribution. Participation was voluntary and anonymous, and no personal identifiers were collected. Respondents were informed about the academic purpose of the

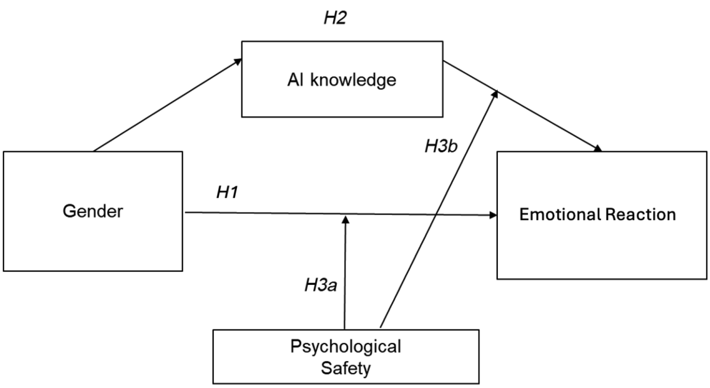

Figure 1. Structure of the hypotheses

study, the confidential handling of responses, and the fact that participation would not affect their employment in any way.

The sample includes 104 complete responses from employees in two organizations undergoing AI transformation. The sample size is modest, which places constraints on statistical power and the precision of estimates. To aid interpretation, we complement statistical significance testing with standardized coefficients and incremental variance explained (f^2). The findings should accordingly be understood as theory-consistent and exploratory rather than as definitive tests of the proposed model.

Although survey participation rates varied across departments, the final sample reflected the diversity of functions and hierarchical levels within the two firms. The respondents represented a cross-section of the workforce, including production, logistics, engineering, maintenance, administration, and management roles. The gender distribution was relatively balanced, with 42% identifying as women and 58% as men. The age composition indicated a predominantly mid-career population, with most employees between 36 and 55 years of age, approximately 20% younger than 35, and about 8% older than 60. In terms of organizational tenure, a majority (52%) had worked in their current firm for more than ten years, another 19% between five and ten years, and the remaining portion had shorter tenure, reflecting a workforce with deep organizational experience. About one-third of respondents held university-level degrees (bachelor’s or higher), while the majority had completed upper secondary or vocational education, consistent with the educational profile typical of industrial and technical occupations in Sweden.

Regarding organizational position, roughly two-thirds of respondents occupied operational or professional roles, such as technicians, operators, administrative, or support staff. Around 10% held first-line managerial roles, and about one-quarter reported middle or senior managerial responsibilities, including department heads and specialists.

Study variables

The *dependent variable* captured employees’ attitudinal responses to artificial intelligence at work in the form of *emotional reaction.* The construct is adapted from Chiu *et al*. (2021) and measured on five-point Likert-type scale ranging from 1 = strongly disagree to 5 = strongly agree. The emotional reaction scale measured the affective dimension of employees’ responses to the introduction of artificial intelligence, including feelings such as enthusiasm, curiosity, and comfort toward its integration into their work. The measure demonstrates high internal consistency and convergent validity (emotional reaction: $\alpha = 0.91$, CR = 0.92, AVE = 0.74)

The *focal independent variable* was *gender,* coded 0 = woman and 1 = man. This coding allows positive coefficients to reflect relatively more favorable reactions among men.

The *mediating variable was perceived AI knowledge*, which captured respondents’ self-assessed familiarity and understanding of artificial intelligence concepts and applications. The construct consisted of three items adapted from Chiu *et al*. (2021), such as “I know quite a lot about artificial intelligence.” Reliability and validity indicators were satisfactory ($\alpha = 0.86$, CR = 0.86, AVE = 0.67).

The *moderating variable* was psychological safety, measured with three items adapted from Edmondson (1999) and Baer and Frese (2003). This construct reflects the extent to which individuals perceive their workplace as a safe environment for expressing opinions, taking interpersonal risks, and discussing problems. One negatively worded item was excluded following reliability testing. The final three-item scale showed acceptable reliability ($\alpha = 0.73$, CR = 0.72, AVE = 0.47).

Several demographic and individual characteristics were included as *control variables* to account for alternative explanations. These comprised age, organizational tenure, educational level, and organizational level. All these variables are captured on ordinal scales with ranges varying from 1 to 10 for age (1 = less or equal to 25; 10 = equal or higher than 65); from 1 to 5 for organizational tenure (1 = less or equal to 5; 5 = equal or more than 20); from 1 to 10 for educational level (1 = no level completed; 10 = PhD level); from 1 to 6 to organizational level (1 = shop floor employees; 6 = Top management). Two conceptually relevant individual factors have been also controlled for: technical competence and readiness for change. Technical competence measured self-reported proficiency in digital and data-related skills ($\alpha = 0.88$, CR = 0.87, AVE = 0.59), while readiness for change assessed emotional attitude, openness and adaptability toward organizational change ($\alpha = 0.89$, CR = 0.89, AVE = 0.62). The latter variable is particularly important as a control in the present study, to account for general affective predispositions toward change that might otherwise confound the analysis of emotional reactions to AI. This variable captures individuals' emotional openness and responsiveness to organizational transformation, irrespective of the specific technology involved.

Including this control is particularly important in a gendered context, given prior evidence that women are more likely than men to express emotional responses in self-report instruments (e.g. Brody and Hall, 2008). By adjusting for general emotional readiness to change, we aim to isolate the effect of gender on AI-specific emotional reactions, ruling out the alternative explanation that gender differences in AI-related emotional responses reflect baseline differences in emotional expressiveness or reporting style, rather than reactions specific to AI technology.

All continuous variables were mean-centered prior to analysis, and categorical variables were treated as continuous to maintain consistency in the regression-based models.

Descriptive statistics and correlations among study variables are reported in Table 1.

Estimation technique and checks

The analysis followed a sequential approach designed to assess both direct and indirect relationships between gender and employees' reactions to artificial intelligence. All analyses were performed in SPSS using the PROCESS macro (Hayes, 2022) and relied on ordinary least squares regression with bootstrapped confidence intervals. The first step examined the direct effect of gender on both emotional and cognitive reactions to artificial intelligence while controlling for demographic and contextual variables. This initial test established whether gender differences existed in employees' responses before accounting for potential explanatory mechanisms.

In the second step, mediation analyses were conducted to determine whether perceived knowledge of artificial intelligence explained the relationship between gender and the two outcome variables. This approach allowed the identification of the indirect effect through employees' self-assessed understanding of artificial intelligence.

In the third step, moderation analyses were performed to assess whether psychological safety influenced the strength of the direct or mediated relationships. This step examined whether the effect of gender on emotional and cognitive reactions, either directly or indirectly via perceived knowledge, depended on the level of psychological safety in the work environment.

Indirect and conditional effects were estimated using 5,000 bootstrap samples with 95% bias-corrected confidence intervals. Control variables were entered simultaneously with the focal predictors in all models to account for demographic and contextual heterogeneity.

Prior studies have routinely adopted cross-sectional designs to model how gendered outcomes are *transmitted through* psychological or relational variables, also with a cross-sectional design, while explicitly refraining from strong causal claims. For example, research on gender and leadership aspirations has modeled gender differences as operating through mediating mechanisms such as perceived support, supervisory behavior, or job control using cross-sectional survey data and mediation analysis, while adopting cautious, associational language (e.g. Fritz and Van Knippenberg, 2020). Similarly, studies examining gender differences in work–life conflict and identity processes have relied on mediation frameworks

Table 1. Descriptives and correlations

	Mean	SD	1	2	3	4	5	6	7	8	9	10
Gender	0.61	0.51	1.00									
Emotional reaction to AI	3.96	2.07	0.10	1.00								
Perceived AI knowledge	3.31	0.91	0.31	0.35	1.00							
Psychological safety	3.65	0.83	−0.14	0.19	0.10	1.00						
Age	3.96	2.07	0.04	−0.12	−0.25	0.00	1.00					
Organizational tenure	5.42	2.68	0.24	−0.02	−0.12	−0.03	0.40	1.00				
Educational level	5.70	2.83	0.09	−0.06	0.00	−0.17	0.00	0.03	1.00			
Organizational level	5.14	3.22	−0.08	−0.05	0.00	0.12	0.10	0.08	0.14	1.00		
Readiness for change	4.46	0.57	−0.25	0.38	0.09	0.39	−0.06	−0.11	0.15	0.16	1.00	
Technical competence	2.81	0.73	0.21	0.25	0.55	−0.03	0.00	−0.06	0.06	0.01	0.16	1.00

Note(s): Correlations with value ∣0.20∣ or higher are significant at $p < 0.05$

linking gender to outcomes through perceptual or identity-related mechanisms, without implying causal determinism (Morgenroth *et al.*, 2021).

Because the study relies on self-reported survey data collected at a single point in time, we assessed the potential for common method bias using established procedural and statistical remedies (Podsakoff *et al.*, 2003, 2012). Procedurally, respondents were assured of anonymity and confidentiality, reducing evaluation apprehension. Statistically, we conducted a Harman-type single-factor test by submitting all measurement items to an unrotated principal component analysis. The first factor accounted for approximately 20% of the total variance, well below commonly cited thresholds for serious common method concerns, suggesting that no single factor dominates the covariance among measures.

Additional features of the study further mitigate concerns about common method bias. The focal constructs are conceptually distinct in their referents and levels of analysis, and the primary independent variable, gender, is a demographic characteristic rather than a perceptual measure, reducing the likelihood of shared method inflation. Moreover, the observed pattern of results is selective rather than uniformly strong, with partial mediation and a targeted moderation effect, which is difficult to reconcile with a general common method explanation (Siemsen *et al.*, 2010; Aguinis *et al.*, 2005). While common method bias cannot be fully ruled out, as is typical in cross-sectional survey research, these checks suggest it is unlikely to be the primary driver of the observed relationships.

Results

The analyses followed a sequential procedure designed to test the hypothesized relationships between gender and employees' reactions to artificial intelligence, first examining direct effects, then mediation, and finally moderated mediation involving psychological safety (Table 2)

Model 1 tests the direct effects of gender on emotional reaction to artificial intelligence. Results showed that women reported less favorable emotional ($b = 0.31$, $p < 0.05$) reactions than men. These findings provide support for H1, indicating that gender differences exist in affective responses toward artificial intelligence in the workplace.

Model 2 examine whether these gender differences were mediated by perceived knowledge of artificial intelligence. Gender is positively associated with perceived knowledge ($b = 0.70$, $p < 0.001$), and perceived knowledge in turn predicted more positive emotional reactions ($b = 0.26$, $p < 0.05$). When perceived knowledge was included in the models, the direct effects of gender on emotional reaction became nonsignificant ($b = 0.19$, $p = 0.27$), while the indirect effect through perceived knowledge was significant (95% CI [0.03, 0.43]). These results are consistent with H2, suggesting that women's less favorable reactions toward artificial intelligence are statistically associated with their lower perceived knowledge of the technology in a pattern consistent with the proposed mediating pathway.

Model 3 introduces psychological safety as a moderator of both the direct and indirect effects of gender. Gender remained significant ($b = 1.72$, $p < 0.05$), but the interaction between perceived knowledge and psychological safety was not ($b = 0.01$, $p = 0.93$). These findings indicate that the indirect effect of gender through perceived knowledge was not conditional on psychological safety, thus, H3b is not supported.

Finally, Models 4 focused on a model in which psychological safety moderated only the direct relationship between gender and emotional and cognitive reactions, while the mediation through perceived knowledge remained constant. The interaction between gender and psychological safety was significant ($b = -0.42$, $p < 0.05$) supporting H3a. This indicates that gender differences in reactions to artificial intelligence are stronger at lower levels of psychological safety and diminish when psychological safety is high. Specifically, conditional effects analyses showed that when psychological safety was low (one SD below the mean), the effect of gender was significant for change in emotional reactions ($b = 0.59$, $p < 0.05$) while these effects became non-significant at average or high levels of psychological safety.

Table 2. Regression models. Dependent variable: emotional reaction to AI adoption

	Model 1 (direct)	Beta (Std. Coeff.)	Model 2 (mediation)	Beta (Std. Coeff.)	Model 3(Full moderated mediation)	Beta (Std. Coeff.)	Model 4 (mediation + direct moderation)	Beta (Std. Coeff.)
Gender (women vs men)	0.31*	0.19*	0.19	0.12	1.72**	0.34**	1.71**	0.33**
Perceived AI knowledge	–		0.26**	0.18***	0.25	0.17***	0.28***	0.2***
Psychological safety	0.05	0.04	0.19	0.13	0.17	0.12	0.19	0.13
Gender × psychological safety	–		–		−0.42**	−0.22***	−0.42**	−0.22***
Perceived AI knowledge × psychological safety	–		–		0.01	0.01	–	
Age	−0.04	−0.03	−0.02	−0.01	−0.01	−0.01	−0.01	−0.01
Organizational tenure	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01
Organizational level	−0.02	−0.01	−0.03	−0.02	−0.02	−0.01	−0.02	−0.01
Education level	−0.04	−0.07	−0.04	−0.07	−0.04	−0.07	−0.04	−0.07
Readiness for change	0.63***	0.45***	0.64***	0.46***	0.66***	0.47***	0.66***	0.47***
Technical competence	0.15	0.11	−0.03	−0.02	−0.03	−0.02	0.01	0.01
Model fit	$R^2 = 0.31$		$R^2 = 0.28$		$R^2 = 0.32$		$R^2 = 0.32$	
	F 4.43***		$F = 5.37$***		$F = 4.43$ ***		$F = 4.98$ ***	

Note(s): *Stars indicate two-tailed significance*: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$*

The inclusion of psychological safety as moderator of the direct path also modified the nature of the mediation. The indirect effects of gender through perceived knowledge remained significant across levels of psychological safety (95% CI [0.05, 0.44]), but the direct effects re-emerged under low safety conditions. This pattern is consistent with partial mediation, suggesting that perceived knowledge is associated with gender differences in emotional reactions particularly when employees feel safe, whereas under low psychological safety, women's less favorable emotional and cognitive reactions appear to reflect both their lower perceived knowledge and additional gender-related differences in how they evaluate or emotionally respond to artificial intelligence.

In addition to statistical significance, we assessed the magnitude of key effects using standardized coefficients (β) and incremental variance explained (f^2). The standardized coefficient for gender in the direct model (Model 1; $\beta = 0.31$) indicates a small-to-moderate effect, suggesting that gender differences in emotional reactions to AI adoption are substantively meaningful even when controlling for individual and contextual factors.

The mediating path through perceived AI knowledge shows a standardized effect of $\beta = 0.18$–0.20 across models, corresponding to a small but non-trivial association. Importantly, the indirect effect accounts for a meaningful portion of the total association between gender and emotional reactions, consistent with expectations for perceptual and affective mechanisms in organizational change research.

The interaction between gender and psychological safety in Models 3 and 4 yields a standardized coefficient of $\beta = -0.22$, indicating a moderate moderation effect. This suggests that psychological safety meaningfully alters the extent to which gender differences translate into emotional reactions under AI adoption.

To further assess practical significance, we examined incremental variance explained (f^2). The inclusion of perceived AI knowledge as a mediator resulted in a small-to-moderate increase in explained variance (0.07), while the addition of the gender $\times$ psychological safety interaction produced an f^2 of 0.05. These values indicate modest but non-trivial contributions, consistent with what might be expected in an early-stage organizational context where AI implementation is still unfolding and where the constructs under study capture perceptual and affective processes. According to conventional benchmarks in organization and management studies (e.g. Aguinis *et al*., 2017), these effect sizes indicate substantively meaningful contributions given the early-stage nature of AI implementation and the organizational context under study.

Supplementary analyses

To assess the robustness of the findings, we conducted additional analyses using cognitive reactions to AI adoption as the dependent variable. The overall pattern of results was consistent with the main analyses, with perceived AI knowledge mediating the association between gender and reactions, and psychological safety moderating the direct gender effect, although effect sizes were generally smaller. We also tested alternative model specifications in which perceived AI knowledge was positioned as a moderator rather than a mediator of the gender–reaction relationship, with psychological safety as a mediator. These models did not yield significant effects. Taken together, these supplementary analyses provide additional confidence in the robustness and internal coherence of the proposed theoretical model.

Discussion

This study aims at examining gendered emotional reactions to AI adoption in AI-driven organizational change and to understand how such reactions are structured by individual and contextual mechanisms. Recent work documents gender gaps in AI usage and attitudes and highlights self-assessed AI knowledge, anxiety, and perceived risk as key correlates (Aldasoro *et al*., 2024; Russo *et al*., 2025; Borwein and Loewen, 2026). Parallel research on the dark side of AI at work emphasizes technostress, alienation, and well-being consequences (Lițan, 2025; Hai *et al.*, 2025; Giuntella *et al.*, 2025). Our study's findings extend these conversations by showing that gendered emotional responses (and, by extension, well-being consequences) during organizational AI rollout are associated with perceived AI knowledge and conditioned by psychological safety, in a pattern consistent with a relational mechanism through which AI transformation could reproduce inequality inside organizations even without overt bias or technical failure. Specifically, the study findings show a consistent pattern: women report less positive emotional reactions to AI adoption than men; this difference is partially transmitted through perceived AI knowledge; and psychological safety conditions whether gender differences persist beyond perceived knowledge. We believe that, taken together, these results offer a more ambitious account of the dark side of AI transformation, one that moves beyond individual anxiety or technostress and toward the reproduction of inequality through emotional and epistemic dynamics embedded in organizational change (Fricker, 2007; Settles *et al.*, 2022).

Contributions to theory

A first contribution concerns how the dark side of AI adoption is conceptualized. Prior research has emphasized that advanced digital technologies can generate negative affective responses such as stress, anxiety, alienation, and disengagement, even when systems function as intended (Tarafdar *et al.*, 2019; Benlian *et al.*, 2022; Cheng *et al.*, 2022a; Kim *et al.*, 2025). This stream has been crucial in shifting attention away from purely instrumental accounts of technology adoption. However, much of this work treats negative affect as a broadly shared individual-level outcome. The present findings suggest that the emotional consequences of AI adoption are not only widespread but also unevenly distributed across groups, reflecting deeper social structures inside organizations. In this sense, the dark side of AI transformation other than including heightened emotional strain points also to a patterned allocation of emotional exposure, referring to who bears the emotional costs of uncertainty and under what conditions.

This interpretation gains strength when AI is considered in light of features that may be particularly pronounced in AI-driven organizational change. While earlier information systems also posed interpretive challenges, AI technologies are often experienced as especially opaque in their functioning and as participating in judgment and decision-making processes that carry material consequences for work outcomes (Burrell, 2016; Faraj *et al.*, 2018; Glikson and Woolley, 2020). Although our study design does not permit direct comparison across technology types, and therefore does not allow us to directly theorize a “difference in kind” between AI-driven change and earlier IT-driven change, we can argue that in AI contexts, uncertainty is not just a transitional feature of adoption but can remain a persistent feature of work, even as systems become routinized. Prior research has shown that persistent uncertainty increases emotional strain and heightens the salience of control and accountability concerns (Beaudry and Pinsonneault, 2010; Maier *et al.*, 2019; Cheng *et al.*, 2022b). Our study suggests that this uncertainty is also socially evaluative: uncertainty becomes emotionally consequential to the extent that it intersects with expectations about competence, credibility, and authority.

The mediating role of perceived AI knowledge is central to this argument. Consistent with recent work treating perceived AI knowledge or AI literacy as a meaningful psychological construct distinct from objective skill (Chiu *et al.*, 2021; Ng *et al.*, 2024; Boob-Engel, 2025), our findings show that gendered emotional reactions are associated with how employees assess their own understanding and confidence in engaging with AI systems, in a manner consistent with the theorized mediating role of perceived AI knowledge. Importantly, this mechanism should not be read as a trivial confirmation that “confidence matters.” Rather, perceived AI knowledge functions as a subjective resource that substitutes for full transparency in contexts where systems are difficult to interrogate and where outputs matter for decisions. When perceived AI knowledge is low, uncertainty is more likely to be internalized and emotionally charged, increasing vulnerability, apprehension, and withdrawal (Tarafdar *et al.*, 2019; Benlian *et al.*, 2022). When perceived knowledge is higher, uncertainty may still exist, but it is less likely to be experienced as personally exposing.

Crucially, perceived AI knowledge is not socially neutral. A substantial body of research shows that women report lower self-assessed technological knowledge and higher apprehension in technology-intensive domains, including AI (Otis *et al.*, 2024; Russo *et al.*, 2025). The present study extends this literature by interpreting these patterns through a broader account of epistemic legitimacy (Fricker, 2007). Epistemic legitimacy is employed here as a theoretical lens to make sense of why perceived AI knowledge carries gendered emotional weight, and does not appear as a construct directly operationalized in our measurement model, but results are consistent with an explanation anchored in this construct. Research on epistemic injustice and epistemic exclusion has long argued that women are more likely to experience credibility deficits in knowledge interactions, not because of lower competence, but because of socially structured assumptions about who counts as a legitimate knower (Rossiter, 1993; Fricker, 2007). More recent organizational scholarship has shown how such dynamics persist in contemporary workplaces, shaping access to interpretive authority, recognition of understanding, and participation in sensemaking (Settles *et al.*, 2022; Lobo, 2025). From this perspective, lower perceived AI knowledge reflects also the internalization of social expectations about competence and authority under conditions of uncertainty.

Social role theory further helps clarify why this mechanism is emotionally relevant in AI contexts. Gendered role expectations shape how individuals evaluate their own competence and how uncertainty is emotionally processed in environments where authority and accountability are salient (Eagly and Karau, 2002; Eagly and Wood, 2012). In domains associated with technical mastery and decision authority, women are more likely to interpret uncertainty as a signal of personal inadequacy rather than as a normal feature of complex work. This tendency is reinforced in AI implementation, where uncertainty cannot always be resolved through explanation and where system outputs may appear authoritative despite limited transparency. Research on decision-making under uncertainty further suggests that women are more sensitive to the social and reputational consequences of ambiguous responsibility and evaluative pressure (Miller, 2012; Wester *et al.*, 2024). Together, these dynamics help explain why perceived AI knowledge mediates the relationship between gender and emotional reactions: it channels gendered expectations about legitimate competence into differentiated emotional experiences during AI-driven change.

The moderating role of psychological safety further deepens the theoretical contribution by showing how organizational context governs whether these gendered emotional dynamics remain salient. Psychological safety has often been framed as a supportive climate that reduces anxiety and facilitates learning (Edmondson, 1999). Our findings suggest a more critical interpretation. Psychological safety conditions whether uncertainty can be voiced without reputational cost and whether questioning is treated as legitimate participation or as a sign of incompetence (Detert and Edmondson, 2011; Carmeli and Gittell, 2009; Frazier *et al.*, 2017). The moderation results indicate that when psychological safety is low, gender remains directly associated with emotional reactions even after accounting for perceived AI knowledge. When psychological safety is high, this direct association diminishes.

Importantly, psychological safety does not strengthen the relationship between perceived AI knowledge and emotional reactions. While caution is warranted in interpreting null effects, this pattern is consistent with the possibility that psychological safety operates less by enhancing the emotional returns of perceived knowledge and more by shaping whether uncertainty becomes emotionally costly in the first place. If confirmed in future research, this would suggest that organizational climates regulate the social evaluation of uncertainty rather than simply amplifying the benefits of confidence (Ridgeway, 2011; Edmondson, 2018).

In AI contexts, where uncertainty is persistent and system outputs carry weight, psychological safety may function as a form of "epistemic governance": it shapes whose doubts are tolerated, whose questions are heard, and whose emotional reactions are legitimized. We note that this interpretation extends beyond what the empirical model directly tests; our measure of psychological safety captures perceived interpersonal safety rather than epistemic governance as such. However, the observed moderation pattern, in which gender differences diminish under high psychological safety, is consistent with this theoretical reading.

Taken together, these findings support a more nuanced view of the dark side of AI transformation. Rather than being simply “stronger” versions of familiar technology-induced stress reactions, AI-related emotional disparities may prove more persistent if, as theorized here, the conditions that generate uncertainty are structurally embedded in the technology itself. The present findings are consistent with this possibility, but confirming it would require longitudinal and comparative designs that directly assess the distinctive features of AI systems, as we also highlight in the limitations section.

The study contributes more broadly also to social role theory by identifying AI implementation as a context that activates gendered expectations about competence and authority in emotionally significant ways. Rather than merely confirming that gender differences in technology-related confidence exist, the findings show how such differences are transmitted through perceived epistemic standing and conditioned by organizational climates that regulate the legitimacy of uncertainty (Detert and Edmondson, 2011; Rossiter, 1993; Lobo, 2025). This extends social role theory into the domain of AI-enabled organizational change and highlights how new technologies can reshape the situations in which gendered role expectations become salient. At the same time, the study does not warrant the claim that AI produces gendered emotional reactions that are fundamentally different in kind from those observed in earlier technological change. The data do not allow for direct comparison across technologies. Accordingly, the contribution lies not in demonstrating AI’s categorical distinctiveness but in showing that the pattern observed here is consistent with the possibility that AI may intensify exclusionary dynamics by sustaining uncertainty and making its social evaluation more important. In this sense, AI transformation may heighten the persistence of inequality-producing processes that have long been documented in studies of gender, expertise, and organizational authority (Eagly and Wood, 2012; Ridgeway, 2011; Fricker, 2007).

Implications for practice

For practice, the results suggest that addressing the dark side of AI transformation requires more than technical training. While increasing AI-related skills and knowledge is important, the findings indicate that it may be insufficient if organizations do not also attend to how uncertainty is socially handled. Leaders play a crucial role in signaling whether not-knowing is compatible with competence, whether questioning AI outputs is legitimate, and whether emotional reactions to AI-driven change are acceptable. To the extent that the patterns observed here hold across contexts, early interventions that legitimize uncertainty and foster psychological safety may help attenuate gendered emotional disparities during AI implementation.

Based on the observed associations, managers and project leaders may benefit from approaching AI implementation with explicit attention to gendered experiences. Training and communication aimed at building confidence and understanding across the workforce, particularly among those who perceive themselves as less knowledgeable, could help address the emotional disparities identified here. Explaining how AI works, opening space for questions, and normalizing uncertainty are practical steps consistent with our finding that perceived AI knowledge and psychological safety are associated with more favorable emotional reactions.

Creating psychologically safe environments is equally important. When employees feel free to express doubts or ask for help, gender-based differences in reactions fade. Leaders play a key role in setting this tone, by showing openness, encouraging dialogue, and recognizing that learning about AI is a shared process rather than an individual test of expertise.

Finally, the findings suggest that AI initiatives would benefit from monitoring emotional and epistemic experiences across demographic groups. While the present study focused on gender, the underlying logic that perceived knowledge and psychological safety shape emotional reactions to AI likely extends to other dimensions of inequality. Paying attention to who participates, who learns, and who feels confident may help organizations identify and address emerging disparities during AI implementation. The moderation findings suggest that creating psychologically safe environments is associated with reduced gender differences in emotional reactions to AI. When employees feel free to express doubts or ask for help, the observed gender gap in reactions appears to diminish. Leaders may play a key role in setting this tone, by showing openness, encouraging dialogue, and recognizing that learning about AI is a shared process rather than an individual test of expertise.

Conclusions, limitations and future research directions

This study examined how gender is associated with employees' emotional reactions to AI adoption during organizational transformation and identified perceived AI knowledge and psychological safety as key mechanisms structuring these experiences. Women report less positive emotional reactions to AI adoption than men, and this difference is statistically associated with lower perceived AI knowledge in a pattern consistent with mediation. Psychological safety conditions whether gender differences persist beyond perceived knowledge, with the gender gap in emotional reactions emerging most clearly in low-safety environments and diminishing when psychological safety is high.

These findings contribute to research on the dark side of AI transformation by suggesting that the emotional consequences of AI adoption are not only widespread but unevenly distributed across groups, and that this uneven distribution is associated with gendered self-assessments of competence and with organizational climates that shape how uncertainty is socially handled. Rather than representing a routine instance of technology-induced stress, the pattern observed here is consistent with the possibility that AI's epistemic opacity and influence on decision authority create conditions under which gendered expectations about competence and authority become particularly consequential for emotional experience. At the same time, the study does not establish that these dynamics are unique to AI, nor does it directly measure the theoretical constructs — epistemic legitimacy, opacity, perceived agency — that are used to interpret the findings. The contribution lies in offering a theoretically grounded and empirically supported account of how gender, perceived knowledge, and organizational climate intersect during AI-driven change, while using epistemic legitimacy as an interpretive lens rather than an operationalized construct.

Several limitations should be acknowledged. First, the data are cross-sectional, which precludes causal inference. Although the hypothesized relationships are theoretically grounded and the statistical pattern is consistent with the proposed mediating pathway, alternative interpretations remain plausible. In particular, employees with less positive emotional reactions may report lower perceived AI knowledge, rather than the reverse, and both variables may be jointly shaped by unobserved features of organizational experience. The findings should therefore be understood as identifying patterns of association consistent with the proposed theoretical model rather than as confirming causal sequences.

Second, the sample size is modest, which constrains statistical power and the precision of estimates. The consistency of effects across models and robustness checks suggests that the observed patterns are meaningful within the studied context, but the findings are best understood as theory-consistent and exploratory. Replication with larger and more diverse samples is needed to assess the stability and generalizability of these associations.

Third, the study examines AI adoption in isolation and does not include a comparison condition involving non-AI technologies. As a result, it cannot establish whether the observed gendered emotional dynamics are specific to AI or would also emerge in other complex technology implementations. The theoretical argument that AI's features may intensify these dynamics is plausible, but assessing the degree to which AI is distinctive in this regard requires comparative designs across technology types.

Fourth, all variables rely on self-reported measures, raising the possibility of common method variance. We mitigated this risk through the use of theoretically distinct constructs, established scales, and moderation analyses, but future studies should consider incorporating multi-source data, behavioral indicators, or experimental designs where feasible.

Fifth, several constructs that are central to the theoretical argument — epistemic legitimacy, epistemic opacity, and perceived agency — are not directly operationalized in the empirical model. Perceived AI knowledge and psychological safety serve as proxies that are theoretically linked to these broader constructs, and the pattern of results is consistent with the proposed interpretation. However, future research that directly measures epistemic legitimacy, perceptions of system opacity, and perceived algorithmic agency would provide stronger empirical grounding for the theoretical claims advanced here.

Sixth, gender was operationalized as a binary variable, which does not capture the full complexity of gender identities. Future research should adopt more inclusive measures and examine how gender intersects with other dimensions of inequality.

These limitations open several avenues for future research. Longitudinal studies could examine whether gendered emotional reactions to AI persist, attenuate, or intensify as AI systems move from pilot phases to routinized use, helping to distinguish short-term adjustment effects from more durable inequality dynamics. Comparative research across technology types would clarify whether the patterns observed here reflect features specific to AI or broader dynamics of contemporary digital transformation. Studies that directly operationalize epistemic legitimacy and perceived algorithmic opacity would strengthen the theoretical bridge between gendered self-assessments and emotional reactions to AI. Qualitative and mixed-methods approaches could illuminate the micro-processes through which employees interpret AI outputs, negotiate uncertainty, and experience legitimacy in everyday work. Finally, extending the framework to other forms of epistemic inequality, such as professional status, occupational expertise, or organizational tenure, would help assess whether the mechanisms identified here generalize beyond gender to other groups whose knowledge claims may be differentially valued in AI-enabled organizations.

References

Afroogh, S., Akbari, A., Malone, E., Kargar, M. and Alambeigi, H. (2024), “Trust in AI: progress, challenges, and future directions”, *Humanities and Social Sciences Communications*, Vol. 11 No. 1, pp. 1-30, doi: 10.1057/s41599-024-04044-8.

Aguinis, H., Beaty, J.C., Boik, R.J. and Pierce, C.A. (2005), “Effect size and power in assessing moderating effects of categorical variables using multiple regression”, *Journal of Applied Psychology*, Vol. 90 No. 1, pp. 94-107, doi: 10.1037/0021-9010.90.1.94.

Aguinis, H., Edwards, J.R. and Bradley, K.J. (2017), “Improving our understanding of moderation and mediation in strategic management research”, *Organizational Research Methods*, Vol. 20 No. 4, pp. 665-685, doi: 10.1177/1094428115627498.

Aldasoro, I., Armantier, O., Doerr, S., Gambacorta, L. and Oliviero, T. (2024), “The gen AI gender gap”, *Economics Letters*, Vol. 241, 111814, doi: 10.1016/j.econlet.2024.111814.

Baer, M. and Frese, M. (2003), “Innovation is not enough: climates for initiative and psychological safety, process innovations, and firm performance”, *Journal of Organizational Behavior*, Vol. 24 No. 1, pp. 45-68, doi: 10.1002/job.179.

Beaudry, A. and Pinsonneault, A. (2010), “The other side of acceptance: studying the direct and indirect effects of emotions on information technology use”, *MIS Quarterly*, Vol. 34 No. 4, pp. 689-710, doi: 10.2307/25750701.

Benlian, A., Klumpe, J. and Hinz, O. (2022), “Mitigating the dark side of algorithmic decision making: differential effects of explanation and user control on emotions and trust”, *Information Systems Research*, Vol. 33 No. 2, pp. 588-607.

Bennett, S.J. (2025), “Artificial intelligence and the ethics of navigating ambiguity”, *Big Data and Society*, Vol. 12 No. 2, doi: 10.1177/20539517251347594.

Boob-Engel, K. (2025), “Is AI as trustworthy as human experts? The role of advice source, technology knowledge, and perceptions of AI’s impact in organizational decision-making”, *Journal of Decision Systems*, Vol. 34 No. 1, 2594620, doi: 10.1080/12460125.2025.2594620.

Borwein, S. and Loewen, P.J. (2026), “The role of risk orientation and risk exposure in the gender gap in attitudes toward AI in the workplace”, *PNAS Nexus*.

Brody, L.R. and Hall, J.A. (2008), “Gender and emotion in context”, in *Handbook of Emotions*, Vol. 3, pp. 395-408.

Brough, A.R., Wilkie, J.E.B., Ma, J., Isaac, M.S. and Gal, D. (2016), “Is eco-friendly unmanly? The green-feminine stereotype and its effect on sustainable consumption”, *Journal of Consumer Research*, Vol. 43 No. 4, pp. 567-582, doi: 10.1093/jcr/ucw044.

Burrell, J. (2016), “How the machine ‘thinks’: understanding opacity in machine learning algorithms”, *Big Data and Society*, Vol. 3 No. 1, pp. 1-12, doi: 10.1177/2053951715622512.

Carmeli, A. and Gittell, J.H. (2009), “High-quality relationships, psychological safety, and learning from failures in work organizations”, *Journal of Organizational Behavior*, Vol. 30 No. 6, pp. 709-729, doi: 10.1002/job.565.

Challapally, A., Pease, C., Raskar, R. and Chari, P. (2025), “The GenAI divide: state of AI in business 2025”, Project NANDA, Massachusetts Institute of Technology.

Cheng, X., Gu, J. and Turel, O. (2022a), “Algorithmic management, technostress, and employee outcomes: a review and research agenda”, *Internet Research*, Vol. 32 No. 6, pp. 1877-1906.

Cheng, X., Gu, J. and Turel, O. (2022b), “Algorithmic decision-making and employee well-being: the dark side of digital transformation”, *Information Systems Frontiers*, Vol. 24 No. 4, pp. 1323-1336.

Cheryan, S., Master, A. and Meltzoff, A.N. (2015), “Cultural stereotypes as gatekeepers”, *Frontiers in Psychology*, Vol. 6, 49, doi: 10.3389/fpsyg.2015.00049.

Chiu, Y.-T., Zhu, Y.-Q. and Corbett, J. (2021), “In the hearts and minds of employees: pre-adoptive appraisal toward AI”, *International Journal of Information Management*, Vol. 60, 102379, doi: 10.1016/j.ijinfomgt.2021.102379.

Detert, J.R. and Edmondson, A.C. (2011), “Implicit voice theories”, *Academy of Management Journal*, Vol. 54 No. 3, pp. 461-488.

Eagly, A.H. (1987), *Sex Differences in Social Behavior*, Lawrence Erlbaum.

Eagly, A.H. and Karau, S.J. (2002), “Role congruity theory of prejudice toward female leaders”, *Psychological Review*, Vol. 109 No. 3, pp. 573-598, doi: 10.1037/0033-295x.109.3.573.

Eagly, A.H. and Wood, W. (2012), “Social role theory”, in Van Lange, P.A.M., Kruglanski, A.W. and Higgins, E.T. (Eds), *Handbook of Theories of Social Psychology*, Sage, Vol. 2, pp. 458-476.

Edmondson, A.C. (1999), “Psychological safety and learning behavior in work teams”, *Administrative Science Quarterly*, Vol. 44 No. 2, pp. 350-383, doi: 10.2307/2666999.

Edmondson, A.C. (2018), *The Fearless Organization: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth*, John Wiley & Sons.

Enholm, I.M., Papagiannidis, S., Mikalef, P. and Krogstie, J. (2021), “Artificial intelligence and business value”, *Information Systems Frontiers*, Vol. 23 No. 5, pp. 1383-1401.

Faraj, S., Pachidi, S. and Sayegh, K. (2018), “Working and organizing in the age of the learning algorithm”, *Information and Organization*, Vol. 28 No. 1, pp. 62-70, doi: 10.1016/j.infoandorg.2018.02.005.

Frazier, M.L., Fainshmidt, S., Klinger, R.L., Pezeshkan, A. and Vracheva, V. (2017), “Psychological safety: a meta-analytic review”, *Journal of Applied Psychology*, Vol. 102 No. 7, pp. 1136-1153.

Fricker, M. (2007), *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford University Press.

Fritz, C. and Van Knippenberg, D. (2020), "Gender and leadership aspiration: supervisor gender, support, and job control", *Applied Psychology*, Vol. 69 No. 3, pp. 741-768, doi: 10.1111/apps.12197.

Gazit, L. (2026), "Constitutive knowledge sources: an institutional approach to epistemic trust in opaque AI systems", *AI and Ethics*, Vol. 6 No. 1, p. 58, doi: 10.1007/s43681-025-00930-2.

Giuntella, O., Konig, J. and Stella, L. (2025), "Artificial intelligence and the wellbeing of workers", *Scientific Reports*, Vol. 15 No. 1, 20087, doi: 10.1038/s41598-025-98241-3.

Glikson, E. and Woolley, A.W. (2020), "Human trust in artificial intelligence: review of empirical research", *Academy of management annals*, Vol. 14 No. 2, pp. 627-660, doi: 10.5465/annals.2018.0057.

Hai, S., Long, T., Honora, A., Japutra, A. and Guo, T. (2025), "The dark side of employee-generative AI collaboration in the workplace: an investigation on work alienation and employee expediency", *International Journal of Information Management*, Vol. 83, 102905, doi: 10.1016/j.ijinfomgt.2025.102905.

Hayes, A.F. (2022), *Introduction to Mediation, Moderation, and Conditional Process Analysis*, 3rd ed., Guilford Press.

Im, I., Hong, S. and Kang, M.S. (2011), "An international comparison of technology adoption: testing the UTAUT model", *Information and management*, Vol. 48 No. 1, pp. 1-8, doi: 10.1016/j.im.2010.09.001.

Kim, D.G., Park, S. and Choi, Y. (2025), "The dark side of artificial intelligence adoption", *Humanities and Social Sciences Communications*, Vol. 12, 123.

Lițan, D.E. (2025), "Mental health in the 'era' of artificial intelligence: technostress and the perceived impact on anxiety and depressive disorders—an SEM analysis", *Frontiers in Psychology*, Vol. 16, 1600013, doi: 10.3389/fpsyg.2025.1600013.

Lobo, M. (2025), "A new direction for epistemic injustice in business", *Business Ethics Quarterly*, pp. 1-26, doi: 10.1017/beq.2025.10084.

Maier, C., Laumer, S., Wirth, J. and Weitzel, T. (2019), "Technostress and the hierarchical levels of information systems use", *Information Systems Journal*, Vol. 29 No. 6, pp. 1220-1240.

McKinsey Global Institute (2023), "The state of AI in 2023", McKinsey & Company.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2021), "A survey on bias and fairness in machine learning", *ACM Computing Surveys*, Vol. 54 No. 6, pp. 1-35, doi: 10.1145/3457607.

Miller, L. and Ubeda, P. (2012), "Are women more sensitive to the decision-making context?", *Journal of Economic Behavior and Organization*, Vol. 83 No. 1, pp. 98-115, doi: 10.1016/j.jebo.2011.06.014.

Mitchell, T. (2025), "Trust and transparency in artificial intelligence", *Philosophy and Technology*, Vol. 38 No. 3, p. 87, doi: 10.1007/s13347-025-00916-2.

Morgenroth, T., Ryan, M.K., Rink, F. and Begeny, C. (2021), "The (in) compatibility of identities: understanding gender differences in work–life conflict through the fit with leaders", *British Journal of Social Psychology*, Vol. 60 No. 2, pp. 448-469, doi: 10.1111/bjso.12411.

Ng, D.T.K., Su, J., Leung, J.K.L. and Chu, S.K.W. (2024), "Artificial intelligence (AI) literacy education in secondary schools: a review", *Interactive Learning Environments*, Vol. 32 No. 10, pp. 6204-6224, doi: 10.1080/10494820.2023.2255228.

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., Broelemann, K., Kasneci, G., Tiropanis, T. and Staab, S. (2020), "Bias in data-driven AI systems", *WIREs Data Mining and Knowledge Discovery*, Vol. 10 No. 3, e1356, doi: 10.1002/widm.1356.

Ong, C.S. and Lai, J.Y. (2006), "Gender differences in perceptions and relationships among dominants of e-learning acceptance", *Computers in Human Behavior*, Vol. 22 No. 5, pp. 816-829, doi: 10.1016/j.chb.2004.03.006.

Otis, N.G., Athey, S., Eckles, D. and Imbrie-Moore, A. (2024), "The uneven impact of generative AI", *PNAS Nexus*, Vol. 3 No. 2, p. 591.

Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y. and Podsakoff, N.P. (2003), "Common method biases in behavioral research", *Journal of Applied Psychology*, Vol. 88 No. 5, pp. 879-903, doi: 10.1037/0021-9010.88.5.879.

Podsakoff, P.M., MacKenzie, S.B. and Podsakoff, N.P. (2012), "Sources of method bias", *Annual Review of Psychology*, Vol. 63, pp. 539-569, doi: 10.1146/annurev-psych-120710-100452.

Rai, A. (2020), "Explainable AI: from black box to glass box", *Journal of the Academy of Marketing Science*, Vol. 48 No. 1, pp. 137-141, doi: 10.1007/s11747-019-00710-5.

Ridgeway, C.L. (2011), *Framed by Gender*, Oxford University Press.

Rossiter, M.W. (1993), "The matthew matilda effect in science", *Social Studies of Science*, Vol. 23 No. 2, pp. 325-341, doi: 10.1177/030631293023002004.

Ruckenstein, M. (2023), *The Feel of Algorithms*, Univ of California Press.

Russo, C., Messina, M. and Meringolo, P. (2025), "Gender differences in AI", *Frontiers in Psychology*, Vol. 16, 1559457.

Sadeghi, S. (2024), "Employee well-being in the age of AI: perceptions, concerns, behaviors, and outcomes", *arXiv preprint* arXiv:2412.04796.

Schilke, O. and Reimann, M. (2025), "The transparency dilemma: how AI disclosure erodes trust", *Organizational Behavior and Human Decision Processes*, Vol. 188, 104405, doi: 10.1016/j.obhdp.2025.104405.

Settles, I.H., Jones, M.K., Buchanan, N.T. and Dotson, K. (2022), "Epistemic exclusion", *Journal of Social Issues*, Vol. 78 No. 2, pp. 378-403.

Siemsen, E., Roth, A. and Oliveira, P. (2010), "Common method bias in regression models with interaction effects", *Organizational Research Methods*, Vol. 13 No. 3, pp. 456-476, doi: 10.1177/1094428109351241.

Suchman, M.C. (1995), "Managing legitimacy", *Academy of Management Review*, Vol. 20 No. 3, pp. 571-610, doi: 10.2307/258788.

Tarafdar, M., Cooper, C.L. and Stich, J.-F. (2019), "The technostress trifecta", *Information Systems Journal*, Vol. 29 No. 1, pp. 6-42, doi: 10.1111/isj.12169.

Venkatesh, V. and Davis, F.D. (2000), "A theoretical extension of the technology acceptance model", *Management Science*, Vol. 46 No. 2, pp. 186-204, doi: 10.1287/mnsc.46.2.186.11926.

Venkatesh, V., Morris, M.G., Davis, G.B. and Davis, F.D. (2003), "User acceptance of information technology: toward a unified view", *MIS Quarterly*, Vol. 27 No. 3, pp. 425-478, doi: 10.2307/30036540.

Venkatesh, V., Thong, J.Y.L. and Xu, X. (2012), "Consumer acceptance and use of information technology", *MIS Quarterly*, Vol. 36 No. 1, pp. 157-178, doi: 10.2307/41410412.

Wester, M., Alfaro, E. and Lama, P. (2024), "Gender differences in risk perception", *Oxford Research Encyclopedia of Natural Hazard Science*.

Young, S.G., McCormick-Huhn, K. and Lammers, J. (2021), "Gendered perceptions of AI", *Computers in Human Behavior*, Vol. 122, 106841.

Zeiser, J. (2024), "Owning decisions: AI decision-support and the attributability-gap", *Science and Engineering Ethics*, Vol. 30 No. 4, p. 27, doi: 10.1007/s11948-024-00485-1.

Corresponding author
Daniel Pittino can be contacted at: daniel.pittino@ju.se