AI-Augmented Leadership and Employee Voice Behaviour: Exploring the Roles of Trust and Empowerment

Kingsley Ikwunga Amadi PhD
Department of Business Administration,
Faculty of Administration and Management.
Rivers State University, Nkpolu-Oroworukwo,
PMB 5080, Port Harcourt, Nigeria
Kingsley.amadi@ust.edu.ng
https://doi.org/10.56201/jpslr.vol.12.no8.2026.pg1.12

Abstract
As organisations increasingly embed artificial intelligence (AI) into the routines of leadership, from predictive performance dashboards to AI-assisted feedback and coaching tools, a pressing but underexamined question is how this technological augmentation of leadership shapes employees' willingness to speak up. This paper develops and defends a conceptual model in which AI-augmented leadership predicts employee voice behaviour through two theoretically distinct but interlocking psychological mechanisms: trust in the combined leader-AI system, and psychological empowerment. Drawing on social exchange theory) the integrative model of organisational trust, Spreitzer's theory of psychological empowerment, and the growing empirical literature on trust in artificial intelligence, the paper argues that AI-augmented leadership is neither inherently emancipatory nor inherently silencing; its consequences for voice depend on whether it is experienced by employees as a resource that expands judgement and relational engagement or as a mechanism of algorithmic control that narrows it. The paper further situates trust and empowerment within the longer-standing voice literature on psychological safety and implicit theories of self-censorship, showing how these established constructs sharpen, rather than duplicate, the proposed model. Five research propositions are advanced, a boundary condition involving AI transparency is specified, and a research agenda for empirically testing the model is outlined. The paper contributes to the emerging literature on AI and leadership by reframing the central question from whether AI replaces leaders to how leaders' use of AI is interpreted by employees, and by clarifying the specific psychological pathways through which that interpretation either encourages or forecloses constructive dissent.

Keywords: *AI-augmented leadership, employee voice behaviour, trust, psychological empowerment, psychological safety, algorithmic control, human-AI collaboration*

1. Introduction

Organisations across manufacturing, financial services, platform work, and knowledge-intensive sectors are integrating artificial intelligence into the routines through which leaders direct, evaluate, and communicate with employees. Predictive analytics inform staffing and promotion decisions, sentiment-monitoring tools flag disengagement before it becomes attrition, and AI-enabled chatbots increasingly mediate day-to-day HR interactions that once depended entirely on a human intermediary (Dutta, Mishra, & Tyagi, 2023). This is not simply automation of routine administrative tasks; it is a qualitative change in the informational and relational infrastructure of leadership itself, one that raises a question of both theoretical and

practical urgency: does AI-augmented leadership make employees more, or less, willing to speak up? The stakes of this question are not trivial. Voice, understood as the discretionary expression of suggestions, concerns, and constructive challenges intended to improve organisational functioning, has been shown across decades of research to be a primary channel through which organisations detect errors, surface innovation, and correct course before small problems become large failures (Morrison, 2011, 2014). An organisational context that inadvertently narrows this channel, even while intending to improve decision-making through better data, would trade a visible short-term efficiency gain for a much less visible long-term erosion of organisational learning capacity. The theoretical difficulty is that AI's relationship with voice is not unidirectional in the existing literature, and the inconsistency across studies is instructive rather than merely a gap to be filled. On one hand, Liu, Huang, Cui, Tian, and Li (2025), working from power dependence theory, demonstrate that employees who depend on AI for informational and decision-support functions report a heightened sense of personal power in their dealings with leaders, and that this elevated power translates into more frequent voice behaviour, particularly under coaching-oriented leadership styles that treat AI as a complement to, rather than a substitute for, developmental guidance. On the other hand, Zhang, Yu and Yang (2025), analysing a large corpus of employee reviews from online delivery platforms using structural topic modelling, find that negative experiences of algorithmic management, characterised by opaque, real-time, largely non-negotiable control over scheduling, routing, and performance evaluation, actively suppress voice by eroding employees' identification with organisational culture, an effect that persists even when employees otherwise report reasonable work-life balance. These two findings are not, on close inspection, contradictory; they instead suggest that the label "AI at work" collapses two organisationally and psychologically distinct phenomena that must be theoretically separated before their consequences for voice can be understood. Kellogg, Valentine, and Christin (2020) provide precisely this separation in their influential synthesis of algorithmic control, showing that algorithms enter the workplace through at least six distinguishable mechanisms, which they term the "6 Rs": directing work by recommending, restricting, recording, rating, replacing, and rewarding. Where AI is used chiefly to restrict, record, and rate, with minimal human mediation, it tends to be experienced as disempowering surveillance; where it is used chiefly to recommend, and is filtered through a human leader who retains discretion and relational responsibility, its effects on employee experience are far less determinate and, this paper argues, contingent on the trust and empowerment that leader mediation generates.

This paper positions AI-augmented leadership within that more favourable, human-mediated end of the Kellogg et al. (2020) spectrum, and asks specifically how it relates to voice once psychological mechanisms, rather than the mere presence of AI, are made the object of theoretical attention. Employee voice behaviour has long been understood to be exquisitely sensitive to the perceived social risk of speaking up: Detert and Edmondson (2011) show, across four studies, that employees carry widely shared, largely unexamined "implicit voice theories", tacit beliefs about when and why voicing an opinion is inappropriate or dangerous, that significantly predict silence independent of formal openness to feedback; and Detert and Burris (2007) show that managerial openness, more than transformational rhetoric alone, shapes whether employees judge the door to voice as genuinely open, an effect mediated by employees' sense of psychological safety (Edmondson, 1999). If leadership itself is this sensitive an antecedent of voice, then a change to the informational and evaluative apparatus surrounding leadership, namely its augmentation by AI, is a plausible candidate to shift employees' implicit calculus about whether speaking up is safe, valued, and likely to matter. This paper argues that two mechanisms carry the bulk of this shift: leader-AI trust, extending the integrative trust model of Mayer, Davis, and Schoorman (1995) into the domain of human-AI collaboration via Glikson and Woolley's (2020) review, and psychological empowerment,

following Spreitzer's (1995) four-dimensional formulation and its established links to voice through empowering leadership (Gao & Jiang, 2019) and ethical leadership (Hu et al., 2018). The remainder of the paper is organised as follows. Section 2 develops a precise conceptual boundary for AI-augmented leadership, distinguishing it from algorithmic management and from individual employee dependence on AI. Section 3 revisits employee voice behaviour as the outcome construct, integrating the risk-calculus perspective offered by psychological safety and implicit voice theories. Sections 4 and 5 develop trust and psychological empowerment, respectively, as the paper's central mediating mechanisms. Section 6 integrates these elements into a theoretical framework and five testable propositions, including a boundary condition specified around AI transparency. Section 7 outlines a methodological approach for future empirical testing, and Section 8 discusses the paper's contributions, limitations, and an agenda for subsequent research.

2. AI-Augmented Leadership: A Conceptual Boundary

According to Baran (2025) AI-augmented leadership is defined here as a leadership configuration in which artificial intelligence tools, including predictive analytics, sentiment-monitoring dashboards, and AI-enabled feedback systems, are used by a human leader to inform decisions, structure feedback, and manage informational load, while ultimate relational and ethical judgement remains vested in that leader. This definition is deliberately narrower than popular usage of the term suggests, because it excludes two adjacent phenomena that dominate the empirical literature on AI and employee outcomes and that, if conflated with AI-augmented leadership, would obscure rather than clarify the mechanisms proposed below. The first excluded phenomenon is algorithmic management, in which the algorithm itself directs, schedules, monitors, or evaluates workers with limited human mediation, a pattern most thoroughly documented in gig and platform contexts and shown by X. Zhang et al. (2025) to suppress voice through the erosion of cultural identification, and theorised comprehensively by Kellogg et al. (2020) as a form of "contested terrain" in which algorithmic recording, rating, and restricting mechanisms displace, rather than support, managerial discretion. The second excluded phenomenon is individual AI dependence, in which an employee relies on AI tools for information or decision support independently of, and sometimes despite, their leader's own practices; Liu et al. (2025) show that this form of AI use operates through a power-dependence mechanism that is conceptually distinct from anything a leader does, since it shifts the employee's subjective sense of power relative to the leader by substituting for functions the leader previously monopolised. AI-augmented leadership occupies the conceptual space between these two phenomena: it is AI used by the leader, in service of the leader-employee relationship, rather than AI that replaces the leader's discretion or that employees adopt independently of it. (Marangozoglu, 2025). Framed through the Kellogg *et al.* (2020) taxonomy of algorithmic control mechanisms, AI-augmented leadership corresponds most closely to the "recommending" function, in which algorithmic outputs are offered as inputs to a human decision rather than as the decision itself, and least closely to the "restricting", "recording", and "rating" functions that dominate accounts of algorithmic management's darker consequences. This distinction matters theoretically because it implies that the same underlying technology, a sentiment-analysis dashboard, for instance, may be experienced by employees in psychologically opposite ways depending on whether a leader uses its output to open a coaching conversation or to issue an unexplained directive. It is important to be precise about what kind of construct this is: the recommending-versus-restricting distinction is not a clean dichotomy that sorts real systems into two bins, since Kellogg et al. (2020) themselves note that a single deployed system typically combines several of the six mechanisms at once, for instance a dashboard that both recommends coaching topics to a leader and simultaneously records granular performance data that could later be repurposed for rating. AI-augmented

leadership is therefore best treated as a matter of degree along a control continuum, operationalised as the relative weight of recommending functions against restricting, recording, and rating functions within a given deployment, rather than as a discrete category a given tool either belongs to or does not. The construct thus treats AI-augmented leadership not as a fixed technological fact but as an organisationally and relationally variable practice, one whose position along this continuum is precisely what should determine its downstream consequences for trust, empowerment, and, ultimately, voice.

3. Employee Voice Behaviour and the Risk Calculus of Speaking Up

Employee voice behaviour is defined, following Morrison (2014), as the voluntary, upward communication of suggestions, concerns, or work-related opinions intended to improve organisational functioning, as distinct from silence, which withholds equivalent input. A substantial body of research, synthesised across three successive reviews by Morrison (2011, 2014, 2023), establishes voice as a form of extra-role, proactive behaviour that carries genuine benefits for organisational learning, innovation, and early problem detection, while simultaneously exposing the employee who speaks to social and, in some contexts, career risk. This dual character, valuable to the organisation yet risky to the individual, is what makes voice theoretically interesting and empirically fragile: unlike task performance, which is often incentivised directly, voice is discretionary precisely because its costs are borne locally while its benefits accrue collectively, a structure that maps naturally onto the logic of social exchange (Blau, 1964), in which employees calibrate discretionary contribution against the trustworthiness and perceived goodwill of the party to whom they are contributing.

Two refinements of the voice construct are especially relevant to the model developed here. First, voice is not unidimensional: Bai and He (2025) distinguish promotive voice, which offers new ideas or suggestions for improvement, from prohibitive voice, which raises concerns about existing or anticipated problems, and show that shared leadership arrangements can affect these two forms of voice in opposite directions simultaneously, a "double-edged sword" pattern that cautions against treating voice as a single outcome that always moves in lockstep. Prohibitive voice, because it more directly threatens the status quo and the parties responsible for it, is consistently found to be the riskier and rarer of the two, which suggests that any technological change to the leadership context should be examined for asymmetric effects on promotive versus prohibitive voice rather than assumed to affect both uniformly. Second, and more foundationally, whether an employee perceives voice as safe at all depends on psychological states that precede and condition the trust and empowerment mechanisms this paper foregrounds. Edmondson's (1999) construct of psychological safety, the shared belief that a given interpersonal context is safe for risk-taking, has been shown in a systematic review by Newman, Donohue, and Eva (2017) to be one of the most consistently replicated antecedents of voice across levels of analysis, from dyads to teams to whole organisations, while Detert and Edmondson (2011) demonstrate that employees additionally carry durable, often unexamined implicit theories about when speaking up is inappropriate, theories that operate largely below conscious awareness and that significantly predict silence even after psychological safety and other situational variables are controlled. Detert and Burris (2007) add a leadership-specific refinement to this picture, showing in a large field study that managerial openness, rather than transformational rhetoric per se, is what most consistently predicts whether subordinates judge the door to voice as genuinely open, and that this relationship operates substantially through subordinates' perceptions of psychological safety. Taken together, this literature implies that any change to the leadership context, including its augmentation by AI, must be evaluated not merely for its stated intent but for how it is read against employees' pre-existing implicit theories about when it is safe to speak.

4. Trust in the Leader-AI System as a Mediating Mechanism

Trust has long occupied a central place in explanations of why employees take the social risk that voice entails, because voice is, in the terms used by Mayer, Davis, and Schoorman (1995), an act of vulnerability: the employee cedes control over how a disclosed concern or suggestion will be received, interpreted, and acted upon. Mayer et al.'s (1995) integrative model specifies that a trustor's willingness to accept this vulnerability depends on perceptions of the trustee's ability, benevolence, and integrity, three dimensions that have proven remarkably durable across three decades of organisational trust research and that this paper argues apply not only to the human leader but, increasingly, to the AI systems that leader deploys. Glikson and Woolley (2020), synthesising the empirical literature on human trust in artificial intelligence, distinguish cognitive trust, built through an AI system's perceived tangibility, transparency, and reliability, from emotional trust, built through the AI's perceived humanness or anthropomorphism, and show that these two forms of trust are shaped by substantially different antecedents and are not interchangeable in their downstream effects. Applied to AI-augmented leadership, this distinction implies that employees form two partially independent trust judgements, one about their leader and one about the AI tools that leader visibly relies upon, and that these judgements are unlikely to remain fully separable in practice: an employee who distrusts an opaque analytics dashboard may extend that distrust, rightly or wrongly, to the leader who defers to it, while an employee who trusts a leader's judgement may extend provisional trust to an AI tool on the strength of that leader's endorsement.

Recent empirical work in leadership contexts lends support to this cross-contamination hypothesis. Q. Zhang, Wang, Liao, and Li (2025), studying manufacturing employees in AI-intensive firms, find that paternalistic leadership, a style characterised by authority, benevolence, and moral integrity, can activate and channel employees' trust in AI toward innovative performance, and that this effect is strengthened further when leaders are perceived to have a clear grasp of where AI genuinely creates opportunity rather than mere novelty. Their finding that leadership style conditions the translation of AI trust into constructive workplace behaviour offers indirect but meaningful support for treating leader-AI trust as a joint, rather than separable, construct: it is not merely that employees trust AI in the abstract, but that the leadership context surrounding AI's introduction substantially determines whether that trust, once formed, is channelled into behaviours that benefit the organisation. Building on this literature, this paper proposes a composite construct, leader-AI trust, defined as the degree to which employees trust the combined system of their leader's judgement and the AI tools that visibly inform it. This is a genuine conceptual extension rather than a straightforward application of existing measurement: Mayer et al.'s (1995) original model was built for a single human trustee, and no validated instrument yet exists for a hybrid human-AI trustee of the kind AI-augmented leadership implies, which means leader-AI trust as specified here is, at present, a theoretical proposal rather than an established construct, and its unidimensionality cannot simply be assumed. An equally defensible modelling choice, and one this paper regards as an open empirical question rather than a settled matter, would treat trust in the leader and trust in the AI system as two correlated but separable variables and test whether a composite or a two-factor structure fits the data better; the composite formulation is retained here chiefly because it generates the more theoretically interesting prediction, namely that distrust of either component can depress voice by contaminating perceptions of the other, a cross-contamination dynamic that a purely additive two-factor model would not capture, but this prediction, and the composite's psychometric viability, should be treated as contestable until tested. Where AI-augmented leadership is experienced as reliable, transparent in its inputs and limitations, and deployed with evident benevolent intent toward employee development rather than surveillance, leader-AI trust should rise; and, following the logic of social exchange (Blau, 1964), this rise in trust should obligate a reciprocal, discretionary contribution on the

employee's part, of which voice is a paradigmatic example, since it is precisely the kind of contribution that cannot be extracted through formal contract or command but must instead be freely offered.

5. Psychological Empowerment as a Mediating Mechanism

Psychological empowerment, as formalised by Spreitzer (1995) building on the earlier conceptual work of Thomas and Velthouse (1990), is a motivational construct comprising four interrelated cognitions: meaning, the value an employee attaches to their work goal relative to their own standards; competence, or self-efficacy specific to the work role; self-determination, a sense of autonomy over how tasks are initiated and carried out; and impact, the degree to which an employee believes they can influence outcomes within their unit. Spreitzer's (1995) original validation work demonstrated that these four dimensions, while empirically distinct, jointly constitute a coherent higher-order construct with predictable antecedents in the structural and social characteristics of the work environment, a finding that has been replicated and extended across dozens of subsequent studies. The relevance of empowerment to voice is well established independently of any consideration of AI: Gao and Jiang (2019), studying 674 supervisor-subordinate dyads, show that empowering leadership predicts voice both directly and indirectly through employees' harmonious passion for their work, with job autonomy strengthening this indirect path, while Hu et al. (2018) show, in a sample of 308 employees at a state-owned telecommunications company, that ethical leadership promotes voice indirectly through the quality of leader-member exchange, which in turn cultivates both psychological safety and psychological empowerment simultaneously, suggesting that these two proximal states, safety and empowerment, though conceptually distinct, tend to be jointly produced by the same high-quality leadership relationships.The theoretically interesting question this paper raises is whether AI-augmented leadership strengthens or weakens this empowerment pathway, and the honest answer, on the basis of existing evidence, is that the effect is plausibly bidirectional and contingent on implementation. On one hand, when AI absorbs routine information-processing and administrative burden that would otherwise consume a leader's attention, it may free leaders to engage more substantively with employees on decisions that carry real discretion, thereby strengthening the impact and self-determination dimensions of empowerment; this reading is broadly consistent with Liu et al.'s (2025) finding that AI-facilitated informational access can increase employees' felt power in their dealings with leaders. On the other hand, if AI-generated outputs are presented to employees as directives rather than as inputs to a shared discussion, employees may reasonably infer that their own competence and self-determination have been narrowed rather than expanded, a reading consistent with Kellogg et al.'s (2020) observation that algorithmic "restricting" and "rating" mechanisms are experienced by workers as disempowerment even when they nominally improve decision accuracy. This paper resolves the tension not by assuming a single direction of effect but by proposing that the direction depends substantially on the same leader-mediation variable that distinguishes AI-augmented leadership from algorithmic management in the first place, which is precisely why leader-AI trust and psychological empowerment are modelled here as parallel, mutually reinforcing mechanisms rather than as a single undifferentiated pathway.

6. Theoretical Framework and Research Propositions

The model developed in this paper is anchored primarily in social exchange theory (Blau, 1964), which holds that favourable treatment by an organisation or its representatives generates a felt obligation to reciprocate through discretionary, organisation-benefiting behaviour, voice being a canonical instance of such behaviour precisely because it cannot be commanded into existence through formal authority. This exchange logic is given psychological specificity by

the integrative model of organisational trust (Mayer et al., 1995), extended into the human-AI domain by Glikson and Woolley (2020), and by Spreitzer's (1995) theory of psychological empowerment, while power dependence theory, as operationalised by Liu et al. (2025), offers a complementary account of how AI specifically alters the relative power positions from which trust and empowerment are experienced. The voice-specific literature on psychological safety (Edmondson, 1999; Newman et al., 2017), implicit voice theories (Detert & Edmondson, 2011), and leadership openness (Detert & Burris, 2007) is treated in this framework not as a competing explanation but as the proximal risk-calculus through which the more distal mechanisms of trust and empowerment are translated into an employee's moment-to-moment decision to speak or remain silent. Five propositions follow from this integration. Proposition 1 holds that AI-augmented leadership, defined as leader-mediated, recommendation-oriented use of AI as distinguished from algorithmic management in Section 2, is positively associated with employee voice behaviour, an association expected to be considerably weaker or reversed when the same underlying technology is deployed in a restricting, recording, or rating configuration consistent with algorithmic management (Kellogg et al., 2020; X. Zhang et al., 2025). Proposition 2 holds that leader-AI trust mediates the relationship between AI-augmented leadership and employee voice behaviour, consistent with the exchange logic of Blau (1964) and the trust model of Mayer et al. (1995) as extended to AI by Glikson and Woolley (2020). Proposition 3 holds that psychological empowerment independently mediates the same relationship, consistent with the empowerment-voice linkages documented by Gao and Jiang (2019) and Hu et al. (2018). Proposition 4 specifies that leader-AI trust and psychological empowerment operate as parallel, partially independent mechanisms rather than a single sequential pathway, such that the indirect effect through trust remains significant when empowerment is statistically controlled, and vice versa. This proposition carries a genuine risk that must be stated rather than assumed away: Hu et al. (2018) found that leader-member exchange cultivates psychological safety and psychological empowerment jointly, which raises the possibility that leader-AI trust and empowerment share enough common antecedent variance, both being downstream of the same high-quality leadership relationship, to be empirically difficult to separate, threatening the discriminant validity Proposition 4 requires. The proposition remains defensible only because trust and empowerment are measured as distinct psychological states with different referents, ability, benevolence, and integrity judgements about a trustee in the case of trust (Mayer et al., 1995) versus meaning, competence, self-determination, and impact judgements about one's own role in the case of empowerment (Spreitzer, 1995), rather than because the two are expected to be uncorrelated; Proposition 4 should therefore be read as a claim about incremental predictive validity net of shared variance, not a claim of zero correlation, and its empirical test should explicitly model the covariance between the two mediators rather than treat them as orthogonal. Proposition 5 introduces a boundary condition: the transparency of the AI system used by the leader, meaning the degree to which its inputs, limitations, and role in a given decision are clearly explained to employees, moderates the strength of the relationship between AI-augmented leadership and leader-AI trust specifically, a prediction grounded directly in Glikson and Woolley's (2020) identification of transparency as a primary antecedent of cognitive trust in AI.

Transparency is not proposed to moderate the empowerment pathway in the same way, because its theorised mechanism, reducing ambiguity about what an AI system is and is not doing, maps onto the informational, ability-and-tangibility basis of cognitive trust far more directly than onto the four empowerment cognitions, which are proposed here to depend chiefly on how much genuine discretion a leader delegates once AI has absorbed routine tasks, a resourcing effect rather than an informational one; this asymmetry is itself a falsifiable sub-claim of the model rather than an oversight, and a finding that transparency also moderates the empowerment pathway would meaningfully qualify Proposition 5 as stated. Collectively, these

five propositions describe a moderated dual-mediation model in which AI-augmented leadership shapes voice through trust and empowerment as parallel, correlated conduits, with AI transparency conditioning the strength of the trust pathway specifically.

7. Toward Empirical Testing

Because this is a conceptual paper, no primary data were collected, and the propositions above are offered as a structured basis for future empirical work rather than as established findings. A time-lagged, multi-source survey design would be well suited to testing the model: employees could report their perceptions of AI-augmented leadership, leader-AI trust, psychological empowerment, and psychological safety at an initial wave, with supervisor-rated or peer-rated voice behaviour collected at a subsequent wave, an approach consistent with designs already used successfully in adjacent voice research (Hu et al., 2018; Gao & Jiang, 2019) and one that would help mitigate the common-method bias that single-source, single-wave designs are especially prone to when both predictor and outcome are self-reported. Existing validated instruments could be adapted rather than newly constructed for most constructs in the model: psychological empowerment can be measured using Spreitzer's (1995) four-dimensional scale without modification, psychological safety using items derived from Edmondson (1999), and trust using items derived from the ability-benevolence-integrity framework of Mayer et al. (1995), extended, following the logic of Glikson and Woolley (2020), to reference explicitly both the leader and the AI tools the leader visibly uses. Organisations that have already adopted AI-based leadership-support tools, such as performance-analytics dashboards, AI-assisted feedback platforms, or AI-enabled coaching systems, would provide an appropriate sampling frame, and a design that explicitly samples across variation in AI transparency, whether through natural organisational variation or through vignette-based experimental manipulation, would allow a direct test of Proposition 5.

Comparative sampling across sectors and, ideally, across regions with differing levels of AI maturity and differing leadership-culture norms would additionally help establish how far the model's Chinese-context empirical foundations (Liu et al., 2025; Bai & He, 2025; X. Zhang et al., 2025) generalise to other institutional settings.

Two further design features would strengthen a test of this model beyond what the basic time-lagged survey described above can deliver on its own. First, given the discriminant-validity risk flagged in Section 6, any empirical test should run a confirmatory factor analysis comparing the fit of a two-factor structure, separating leader-AI trust from psychological empowerment, against both a one-factor structure and, for the trust construct specifically, a two-factor split between leader trust and AI trust, before proceeding to test the structural propositions; a model that fails this discriminant check would falsify Proposition 4 as stated rather than merely weaken it. Second, the correlational, time-lagged design proposed above cannot rule out reverse causality or selection: it is equally plausible that leaders extend AI-augmented practices to teams that already voice concerns readily, or that employees predisposed to trust authority self-select into organisations that adopt AI-augmented leadership, either of which would produce the predicted associations without the causal structure the model proposes. Quasi-experimental designs that exploit the staggered rollout of AI leadership-support tools across otherwise comparable units, using rollout timing as a source of exogenous variation, or vignette-based experiments that manipulate AI transparency directly, would offer stronger causal identification than the observational design alone and are recommended as a complement to it.

8. Discussion, Contributions, and Limitations

This paper makes three contributions to the leadership and employee-voice literatures. First, it draws a conceptual boundary, grounded in the Kellogg et al. (2020) taxonomy of algorithmic

control mechanisms, between AI-augmented leadership and the two adjacent phenomena, algorithmic management and individual AI dependence, that currently dominate empirical discussion of AI and voice (X. Zhang et al., 2025; Liu et al., 2025), a boundary that clarifies why existing findings in this literature appear to point in opposite directions and specifies the leader-mediation variable that plausibly reconciles them. Second, it proposes leader-AI trust and psychological empowerment as parallel, theoretically distinct mechanisms rather than folding either into the other, building an explicit bridge between the organisational trust literature (Mayer et al., 1995), the still-nascent human-AI trust literature (Glikson & Woolley, 2020), and the well-established empowerment-voice literature (Spreitzer, 1995; Gao & Jiang, 2019; Hu et al., 2018), while situating both mechanisms against the longer-standing psychological safety and implicit-voice-theory literatures (Edmondson, 1999; Newman et al., 2017; Detert & Edmondson, 2011) that describe the proximal risk-calculus through which any distal mechanism must ultimately operate. Third, it identifies AI transparency as an actionable boundary condition, translating an abstract theoretical distinction into a practical lever, namely how leaders explain, disclose, and contextualise their use of AI, that organisations can influence directly, independent of the underlying AI technology itself.

The paper's central limitation is its conceptual status: the five propositions require empirical testing before any claim can be made about their direction, magnitude, or causal structure, and it remains entirely possible that empirical testing will reveal the trust and empowerment pathways to be less separable, or the transparency moderation less consequential, than theorised here. A second limitation concerns cultural and institutional context; a striking proportion of the empirical literature this paper draws upon to motivate its propositions, including Liu et al. (2025), Bai and He (2025), Hu et al. (2018), and both X. Zhang et al. (2025) and Q. Zhang et al. (2025), is drawn from Chinese organisational settings characterised by particular norms around hierarchy, paternalism, and collective identification that may not travel unchanged to other institutional contexts, including the Sub-Saharan African settings in which this paper's readership may be most directly interested; testing the model in such settings would be a natural and valuable extension. A third limitation concerns the voice construct itself: because Bai and He (2025) demonstrate that promotive and prohibitive voice can respond to the same leadership antecedent in opposite directions, future work should test whether AI-augmented leadership's proposed positive effect on voice holds uniformly across both forms, or whether, as this paper's discussion of the Kellogg et al. (2020) "rating" mechanism would predict, prohibitive voice specifically, the riskier act of flagging a problem the AI system itself may have helped create or obscure, proves more sensitive to variation in AI transparency than promotive voice does. A fourth limitation is endogeneity: as noted in Section 7, the model as specified is consistent with reverse causal or selection-based explanations, in which pre-existing voice climates shape which teams receive AI-augmented leadership, or trust-prone employees select into AI-adopting organisations, rather than AI-augmented leadership causing the trust, empowerment, and voice outcomes proposed here, a possibility the observational designs discussed in Section 7 only partially address and that the quasi-experimental alternatives proposed there are specifically intended to rule out. A fifth limitation concerns the sectoral base of the supporting evidence: beyond the national-context concentration already noted, the empirical studies motivating this model are drawn overwhelmingly from manufacturing (Liu et al., 2025; Q. Zhang et al., 2025), gig-platform delivery work (X. Zhang et al., 2025), and a state-owned telecommunications firm (Hu et al., 2018), sectors in which task structure, supervisory ratios, and the visibility of AI systems to frontline employees may differ substantially from professional, knowledge-intensive, or public-sector contexts in which AI-augmented leadership is also emerging; the model's applicability to these under-sampled sectors should be regarded as an open question rather than an assumption. Finally, longitudinal designs extending beyond the two-wave structure proposed in Section 7 would be valuable for

testing whether repeated negative experiences of AI-augmented leadership produce a durable shift in employees' implicit voice theories (Detert & Edmondson, 2011) that persists even after specific AI systems are improved or replaced, a possibility the present cross-sectional framing cannot itself adjudicate.

9. Conclusion

As artificial intelligence becomes further embedded in the practice of leadership, understanding its consequences for employee voice is essential to preserving the upward flow of information on which organisational learning, error correction, and innovation depend. This paper has argued that AI-augmented leadership is unlikely to bear on voice directly or uniformly; rather, its effects are proposed to run through the trust employees place in the combined leader-AI system and through the psychological empowerment they experience in their roles, with the transparency of the AI system conditioning how strongly the trust pathway in particular unfolds, and with the longer-established literatures on psychological safety and implicit voice theories describing the proximal risk-calculus through which these more distal mechanisms are ultimately expressed or suppressed. Distinguishing AI-augmented leadership carefully from algorithmic management, following the taxonomy offered by Kellogg, Valentine, and Christin (2020), is not a merely definitional exercise but a theoretical necessity, since the same technology can plausibly produce opposite consequences for voice depending on whether its use is mediated by a trusted, transparent, empowering leader or imposed as an unmediated mechanism of restriction, recording, and rating. Testing the model proposed here would help organisations design AI-augmented leadership practices deliberately, rather than incidentally, toward strengthening, rather than silencing, the voice of the employees they depend on to notice what leaders, human or algorithmic, cannot see alone.

References

Bai, Y., & He, H. (2025). The "double-edged sword" effect of shared leadership on employee voice behavior. *Frontiers in Psychology,* 16, Article 1560554. https://doi.org/10.3389/fpsyg.2025.1560554

Baran, M. (2025). AI-augmented leadership: Redefining leaders' style and competencies. In *AI and Humanistic Management* (pp. 147-169). Routledge

Blau, P. M. (1964). *Exchange and power in social life*. Wiley.

Detert, J. R., & Burris, E. R. (2007). Leadership behavior and employee voice: Is the door really open? *Academy of Management Journal,* 50(4), 869–884. https://doi.org/10.5465/amj.2007.26279183

Detert, J. R., & Edmondson, A. C. (2011). Implicit voice theories: Taken-for-granted rules of self-censorship at work. *Academy of Management Journal*, 54(3), 461–488. https://doi.org/10.5465/amj.2011.61967925

Dutta, D., Mishra, S. K., & 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. https://doi.org/10.1080/09585192.2022.2085525

Edmondson, A. C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383. https://doi.org/10.2307/2666999

Gao, A., & Jiang, J. (2019). Perceived empowering leadership, harmonious passion, and employee voice: The moderating role of job autonomy. *Frontiers in Psychology*, 10, Article 1484. https://doi.org/10.3389/fpsyg.2019.01484

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals,* 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Hu, Y., Zhu, L., Zhou, M., Li, J., Maguire, P., Sun, H., & Wang, D. (2018). Exploring the influence of ethical leadership on voice behavior: How leader-member exchange, psychological safety and psychological empowerment influence employees' willingness to speak out. *Frontiers in Psychology*, 9, Article 1718. https://doi.org/10.3389/fpsyg.2018.01718

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals,* 14(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Liu, J., Huang, M., Cui, M., Tian, G., & Li, X. (2025). The positive effects of employee AI dependence on voice behavior—Based on power dependence theory. *Behavioral Sciences,* 15(12), Article 1709. https://doi.org/10.3390/bs15121709

Marangozoglu, S. (2025). *Reconstructing Leadership: The Convergence of Leadership Theories and AI-Driven Decision-Making* (Doctoral dissertation, Technische Universität Wien).

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review,* 20(3), 709–734. https://doi.org/10.2307/258792

Morrison, E. W. (2011). Employee voice behavior: Integration and directions for future research. *Academy of Management Annals,* 5(1), 373–412. https://doi.org/10.5465/19416520.2011.574506

Morrison, E. W. (2014). Employee voice and silence. *Annual Review of Organizational Psychology and Organizational Behavior,* 1, 173–197. https://doi.org/10.1146/annurev-orgpsych-031413-091328

Morrison, E. W. (2023). Employee voice and silence: Taking stock a decade later. *Annual Review of Organizational Psychology and Organizational Behavior,* 10, 79–107. https://doi.org/10.1146/annurev-orgpsych-120920-054654

Newman, A., Donohue, R., & Eva, N. (2017). Psychological safety: A systematic review of the literature. *Human Resource Management Review,* 27(3), 521–535. https://doi.org/10.1016/j.hrmr.2017.01.001

Spreitzer, G. M. (1995). Psychological empowerment in the workplace: Dimensions, measurement, and validation. *Academy of Management Journal*, 38(5), 1442–1465. https://doi.org/10.2307/256865

Thomas, K. W., & Velthouse, B. A. (1990). Cognitive elements of empowerment: An "interpretive" model of intrinsic task motivation. *Academy of Management Review,* 15(4), 666–681. https://doi.org/10.5465/amr.1990.4310926

Zhang, Q., Wang, F., Liao, G., & Li, M. (2025). How does AI trust foster innovative performance under paternalistic leadership? The roles of AI crafting and leader's AI opportunity perception. *Behavioral Sciences,* 15(8), Article 1064. https://doi.org/10.3390/bs15081064

Zhang, X., Yu, D., & Yang, J. (2025). Discovering the dark side of AI and algorithmic HRM: Evidence from negative experiences of employees on online delivery platforms. SAGE Open, 15. https://doi.org/10.1177/21582440251405329