International Journal of Education, Leadership, Artificial Intelligence, Computing, Business, Life Sciences, and Society - https://journalviral.org
DOI: https://doi.org/10.65222/VIRAL.2026.10.55.75

THE LEADERSHIP AUTHENTICITY CRISIS: WHEN AI-GENERATED COMMUNICATION UNDERMINES TRUST IN HUMAN LEADERS

Bogdan COSTACHE
Bucharest University of Economic Studies, Bucharest, Romania

Abstract
Background: *The growing integration of generative artificial intelligence into organizational communication is transforming how managers formulate messages, provide feedback, express recognition, and engage with employees. Although AI-assisted communication may improve efficiency, consistency, and responsiveness, it also raises questions about the perceived sincerity, intentionality, and personal commitment underlying managerial interactions. As technologically mediated messages become increasingly indistinguishable from personally authored communication, employees may struggle to determine whether managerial expressions reflect genuine human judgment or algorithmically constructed responses. This emerging tension introduces a potential crisis of perceived authenticity in contemporary leadership.*
Purpose: *This conceptual paper investigates how AI-generated managerial communication may influence perceived leader authenticity, interpersonal trust, and the quality of leader–employee relationships. It addresses the central research question: Under what conditions does AI-mediated managerial communication weaken or preserve perceived leadership authenticity and employee trust? The study examines the tension between communication efficiency and relational credibility, emphasizing the distinction between technological assistance and the personal responsibility of human leaders.*
Method: *The study adopts a conceptual, theory-building approach grounded in Authentic Leadership Theory, Social Exchange Theory, and Attribution Theory. Through an integrative synthesis of literature on generative AI, organizational communication, interpersonal trust, and technology-mediated relationships, it develops a theoretical framework explaining how employees interpret AI-assisted managerial messages. The analysis focuses on perceived intentionality, emotional congruence, authorship transparency, and relational accountability as mechanisms connecting AI involvement with judgments of managerial authenticity.*
Results: *The conceptual analysis proposes a* ***Leadership Authenticity - AI Mediation Framework****, identifying a potential authenticity gap between the technical quality of managerial messages and employees' perceptions of genuine human involvement. The framework suggests that extensive or undisclosed AI mediation may increase uncertainty regarding communicative intentions and weaken relational credibility, particularly in emotionally sensitive interactions. Conversely, transparent AI assistance, meaningful human personalization, and explicit managerial ownership may help preserve authenticity and interpersonal trust. The study develops theoretical propositions concerning the moderating roles of communication sensitivity, disclosure practices, and the consistency between expressed messages and observable managerial behavior.*
Conclusion: *The paper contributes to leadership and organizational communication research by conceptualizing authenticity as a relational judgment shaped not only by message content but also by perceived authorship, intentionality, and accountability. It argues that effective AI integration requires leaders to preserve personal responsibility for consequential communication rather than equating linguistic sophistication with genuine interpersonal connection. The proposed framework offers a foundation for future empirical research and practical guidance for developing communication governance practices that balance technological efficiency with relational integrity.*

Keywords: Perceived intentionality; relational credibility; authorship transparency; emotional congruence; algorithmic mediation.

1. Introduction

Generative artificial intelligence (AI) is reshaping organizational communication by enabling leaders to draft messages, deliver feedback, and formulate expressions of recognition with unprecedented efficiency. However, the growing delegation of communicative tasks to AI raises a fundamental leadership question: can employees perceive a message as an authentic expression of managerial concern when its language has been substantially generated by a technological system? While AI may improve communicative clarity and responsiveness, these advantages do not necessarily establish the sincerity, intentionality, or personal commitment traditionally associated with credible leadership. The resulting tension concerns the distinction between the technical quality of managerial communication and its perceived relational authenticity.

Authentic Leadership Theory provides an essential foundation for examining this distinction. Avolio and Gardner [1] emphasized self-awareness, value congruence, and relational processes in authentic

DOI: https://doi.org/10.65222/VIRAL.2026.10.55.75

leadership development, while Walumbwa et al. [2] identified relational transparency, internalized moral perspective, balanced processing, and self-awareness as its principal dimensions. These perspectives suggest that authenticity extends beyond persuasive language to the perceived consistency between leaders' intentions, values, and behavior. Critical contributions further indicate that authenticity is socially interpreted rather than simply demonstrated through positive communication [3,4]. Consequently, AI-generated messages may appear empathetic and professionally appropriate without necessarily being interpreted as evidence of genuine managerial engagement.

The implications extend to interpersonal trust. Mayer et al. [5] conceptualized organizational trust through perceptions of ability, benevolence, and integrity, while Social Exchange Theory emphasizes the reciprocal obligations through which workplace relationships develop [6]. From these perspectives, employees may distinguish between a leader's ability to communicate effectively and the perceived willingness to invest personally in a relationship. Importantly, trust in AI should not be conflated with trust in the human leader using it. Although research has examined the determinants of human trust in artificial intelligence [7], the mechanisms through which AI-assisted communication influences judgments about managerial authenticity require separate theoretical consideration.

This problem reflects a broader tension between technological automation and human augmentation. Raisch and Krakowski [8] demonstrated that these processes are interdependent, suggesting that improvements in managerial efficiency may coexist with new organizational challenges. In communication-intensive settings, AI can support message production without assuming the leader's interpersonal responsibilities. Whether employees interpret this arrangement as legitimate assistance or diminished personal involvement may depend on authorship transparency, perceived intentionality, emotional congruence, and the sensitivity of the interaction. These considerations are particularly relevant to educational leadership, where communication concerning professional feedback, staff well-being, and institutional change often carries substantial relational significance.

Against this background, the present conceptual study addresses the research question: Under what conditions does AI-mediated managerial communication weaken or preserve perceived leadership authenticity and employee trust? Drawing on Authentic Leadership Theory, Social Exchange Theory, and attribution-based reasoning, the paper develops a Leadership Authenticity - AI Mediation Framework explaining how employees may interpret technological involvement in managerial communication. Its proposed contribution is to distinguish message quality from perceived human intentionality, identify conditions influencing authenticity judgments, and clarify how leaders can retain relational accountability while using generative AI. Rather than presuming that AI necessarily undermines trust, the framework examines the circumstances under which technological assistance may weaken or preserve credible human leadership.

2. Literature Review

Authentic leadership provides a theoretical foundation for examining how generative AI may alter employees' interpretations of managerial communication. Avolio and Gardner [1] associated authentic leadership with self-awareness, value congruence, and relational processes, while Walumbwa et al. [2] conceptualized authenticity through self-awareness, relational transparency, balanced processing, and an internalized moral perspective. These dimensions suggest that authenticity depends not merely on what leaders communicate but also on whether their expressions correspond to genuine intentions and observable conduct. However, critical perspectives challenge the assumption that authenticity constitutes a universally recognizable leadership attribute. Liu et al. [3] emphasized its socially constructed character, while Alvesson and Einola [4] questioned overly affirmative interpretations of authentic leadership. Accordingly, the authenticity of AI-assisted communication cannot be established solely through its linguistic sophistication or emotional tone; it must also be understood through employees' interpretations of managerial involvement.

Organizational trust theory further clarifies the relational consequences of technological mediation. Mayer et al. [5] identified ability, benevolence, and integrity as central determinants of perceived trustworthiness, whereas Cropanzano and Mitchell [6] explained how reciprocal obligations shape workplace relationships. Applied to managerial communication, these perspectives suggest that

DOI: https://doi.org/10.65222/VIRAL.2026.10.55.75

employees may distinguish between a leader's capacity to deliver effective messages and the perceived sincerity of the intentions underlying them. AI assistance may enhance communicative competence while introducing uncertainty about personal commitment. Nevertheless, technological involvement does not necessarily diminish trust when leaders demonstrate meaningful engagement and consistency between communication and behavior. Importantly, trust in an AI system differs conceptually from trust in the leader employing it. Glikson and Woolley [7] examined the determinants of human trust in artificial intelligence, but their findings do not establish that confidence in technological performance translates directly into interpersonal trust.

The automation–augmentation paradox offers a complementary explanation of this tension. Raisch and Krakowski [8] argued that automation and augmentation are interdependent processes rather than mutually exclusive alternatives. In managerial communication, generative AI may automate linguistic production while augmenting leaders' ability to communicate clearly and efficiently. However, these advantages may coexist with ambiguity concerning authorship and responsibility. Jarrahi [9] emphasized the complementarity of human and artificial intelligence in organizational decision-making, while Shrestha et al. [10] examined the redistribution of decision-making activities between human and algorithmic systems. Although these studies primarily concern organizational decisions, they support a distinction between delegating operational tasks and retaining human accountability. A leader may delegate message drafting without relinquishing responsibility for the intentions and commitments communicated.

Attribution-based reasoning is particularly relevant to explaining how employees interpret this arrangement. When managerial messages are AI-assisted, recipients may face uncertainty about whether expressions of concern, appreciation, or empathy originate from the leader's personal judgment or from algorithmically generated language. Gambino et al. [11], extending the Computers Are Social Actors paradigm, demonstrated the theoretical importance of social interpretations in human–machine communication. However, leadership communication differs from direct interaction with AI because employees evaluate a human authority figure whose message has been technologically mediated. Consequently, perceived intentionality and authorship attribution may influence whether employees interpret a message as an expression of genuine leadership engagement.

Authorship transparency introduces an additional complication. Disclosure of AI involvement may support procedural honesty while simultaneously drawing attention to reduced human participation in message production. Its relational implications may therefore depend on the sensitivity of the communication, the leader's explanation for using AI, and the degree of meaningful human personalization. Similarly, emotional congruence may influence whether AI-assisted messages appear consistent with the leader's established communication style and subsequent behavior. These mechanisms suggest that the relationship between AI mediation and authenticity is conditional rather than uniformly negative.

Taken together, the literature establishes the importance of authenticity, interpersonal trust, technological mediation, and attribution processes, but their integration remains theoretically consequential. The proposed **Leadership Authenticity - AI Mediation Framework** addresses this intersection by distinguishing communication quality from perceived human intentionality and by examining how authorship transparency, emotional congruence, and relational accountability may influence trust in human leaders. This perspective is particularly relevant to educational management, where professional feedback, institutional communication, and staff support frequently involve expectations of personal attention and responsible leadership.

3. Methodology

This study adopts a conceptual, theory-building design combining integrative literature synthesis with abductive theoretical reasoning to examine how generative AI influences perceived leadership authenticity and interpersonal trust. This approach is appropriate for an emerging organizational phenomenon in which technological developments raise questions that established leadership models do not fully explain. Rather than testing causal effects through primary data, the study integrates established theories and relevant research on AI-mediated communication to clarify constructs, identify explanatory mechanisms, and formulate empirically testable propositions.

The theoretical architecture integrates Authentic Leadership Theory, Social Exchange Theory, and Attribution Theory. Authentic Leadership Theory establishes the conceptual foundation for examining self-awareness, relational transparency, balanced processing, and value congruence. Social Exchange Theory explains how perceived managerial engagement and reciprocal expectations

DOI: https://doi.org/10.65222/VIRAL.2026.10.55.75

contribute to interpersonal trust. Attribution Theory provides the interpretive mechanism through which employees infer sincerity, intentionality, and responsibility when managerial messages are partially or substantially generated by AI. Research on human - AI interaction and the automation–augmentation paradox complements these perspectives by distinguishing technological assistance from human relational accountability.

The analytical procedure comprises three interconnected stages. First, a targeted interdisciplinary synthesis examines foundational and contemporary literature on leadership authenticity, organizational trust, social exchange, generative AI, and technology-mediated communication. Particular attention is given to differentiating trust in AI as a technological system from trust in the human leader responsible for managerial communication. Second, thematic cross-mapping identifies conceptual tensions between communication efficiency and perceived relational authenticity. This stage establishes perceived intentionality and emotional congruence as potential explanatory mechanisms, while distinguishing authorship transparency and relational accountability as conditions influencing employees' interpretations. Third, the identified constructs are integrated into the **Leadership Authenticity - AI Mediation Framework**, from which theoretical propositions are derived concerning the relationships between AI involvement, perceived authenticity, and interpersonal trust.

The resulting framework incorporates communication sensitivity, authorship disclosure, meaningful human personalization, and behavioral consistency as potential boundary conditions. The proposed relationships are assessed for conceptual coherence, explanatory relevance, and empirical testability. As the study does not involve participant recruitment, primary data collection, or statistical estimation, its findings represent theoretically derived propositions rather than validated causal effects. Future research may examine these relationships through experimental vignettes, longitudinal organizational studies, and leader–employee dyadic surveys.

4. Results

The conceptual synthesis yields the **Leadership Authenticity - AI Mediation Framework**, which explains how employees may evaluate the authenticity of human leaders when managerial messages are produced with generative AI. The framework distinguishes the extent of technological involvement from employees' interpretations of the leader's intentions and responsibilities. Its central theoretical result is that communication efficiency and perceived authenticity are not necessarily aligned: a technically effective message may generate uncertainty about personal commitment, whereas transparently assisted communication may remain credible when the leader demonstrates meaningful involvement.

Table 1 synthesizes the principal constructs of the proposed Leadership Authenticity - AI Mediation Framework, highlighting their theoretical roles and expected implications for perceived leadership authenticity and interpersonal trust.

Table 1. Leadership Authenticity - AI Mediation Framework: Constructs, Mechanisms, and Expected Relational Outcomes

Construct	Role in the framework	Expected implications for authenticity and trust
Degree of AI mediation	Antecedent	Greater AI involvement may weaken perceived personal engagement, particularly in sensitive communication.
Perceived intentionality	Mediating mechanism	Stronger perceptions of genuine managerial intention may support leadership authenticity.
Emotional congruence	Interpretive mechanism	Alignment between message tone and established leader behavior may reinforce perceived sincerity.
Communication sensitivity	Contextual moderator	Emotionally consequential messages may amplify concerns about AI-generated communication.

DOI: https://doi.org/10.65222/VIRAL.2026.10.55.75

Authorship transparency	Contextual moderator	Disclosure may support honesty but also increase awareness of limited human involvement.
Human personalization and relational accountability	Protective conditions	Meaningful leader involvement and responsibility may preserve trust despite AI assistance.
Behavioral consistency	Contextual moderator	Alignment between managerial messages and subsequent actions may strengthen credibility.
Perceived leadership authenticity	Relational outcome	Greater perceived authenticity may contribute to interpersonal trust in the leader.
Interpersonal trust	Final relational outcome	Trust may be preserved or weakened depending on employees' interpretations of AI-mediated communication.

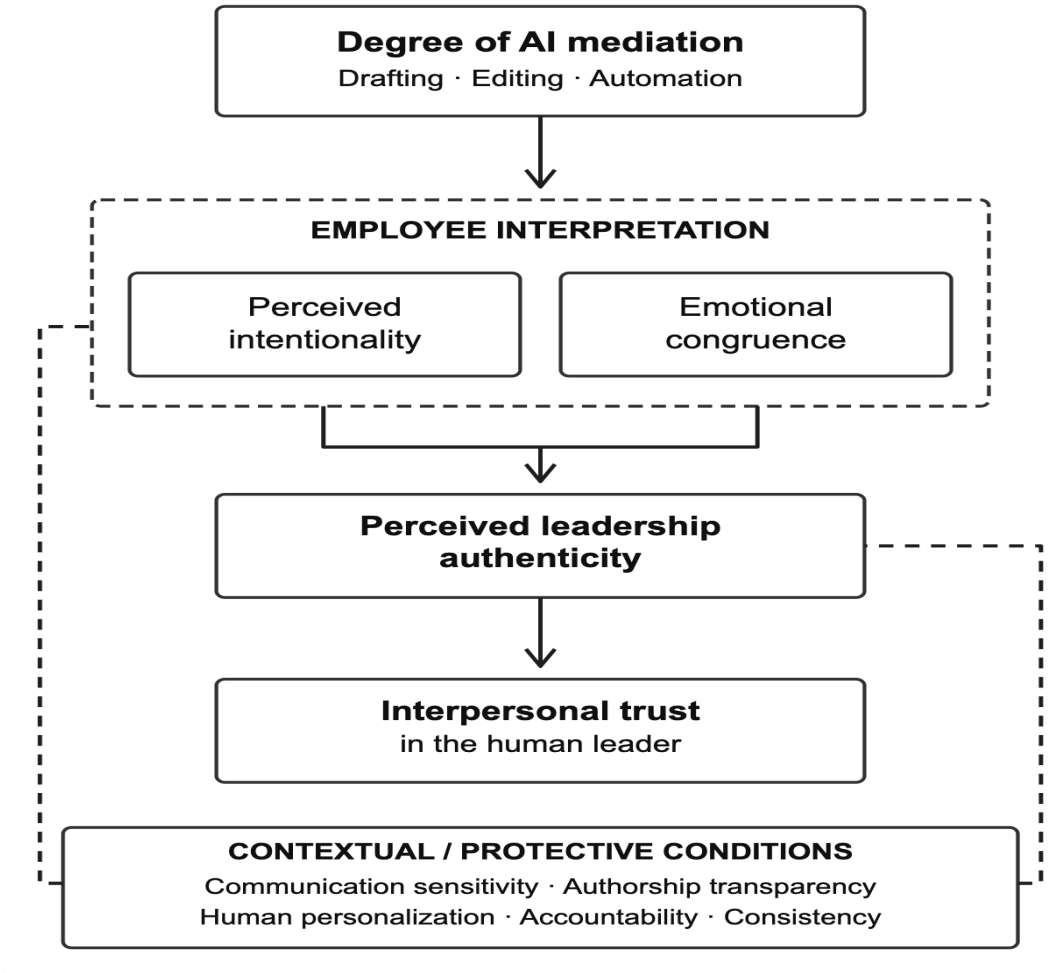
Note: The relationships presented are theoretically proposed rather than empirically validated. The table summarizes the conceptual framework and does not report observed effects.

Source: Authors' own conceptual contribution.

Figure 1 illustrates the proposed Leadership Authenticity - AI Mediation Framework, showing how the degree of AI involvement may influence interpersonal trust through employees' perceptions of managerial intentionality and leadership authenticity, under different relational and contextual conditions

Figure 1. Leadership Authenticity - AI Mediation Framework: Theoretical Pathways from AI-Mediated Communication to Interpersonal Trust

DOI: https://doi.org/10.65222/VIRAL.2026.10.55.75



Source: Authors' own conceptual contribution.

The framework identifies three interconnected analytical components. The first concerns **AI mediation**, ranging from limited drafting assistance to substantial algorithmic production of managerial messages. The second concerns **relational interpretation**, particularly perceived intentionality and emotional congruence. Employees may evaluate whether the message reflects the leader's genuine intentions and whether its emotional expression corresponds to established interpersonal behavior. The third concerns **relational outcomes**, represented by perceived leadership authenticity and trust. Consistent with authentic leadership and organizational trust theories [2,5], these outcomes are conceptualized as judgments about the human leader rather than evaluations of the technological system.

The synthesis generates four theoretical propositions.

P1: Greater perceived AI involvement in managerial message production is associated with lower perceived leader intentionality when employees expect direct personal engagement.

P2: Perceived intentionality mediates the relationship between AI involvement and perceived leadership authenticity. These propositions distinguish technological authorship from the human intentions attributed to the resulting communication and are consistent with the broader distinction between automation and augmentation [8].

The framework further identifies contextual conditions that may alter these relationships.

P3: The negative association between perceived AI involvement and leadership authenticity is stronger in emotionally sensitive communication than in routine administrative communication. This proposition reflects the expectation that feedback, recognition, and expressions of concern may carry greater relational significance than standardized informational exchanges.

P4: Authorship transparency and meaningful human personalization jointly moderate the relationship between AI involvement and interpersonal trust, such that transparent assistance accompanied by demonstrable managerial engagement is more likely to preserve trust than extensive AI involvement

DOI: https://doi.org/10.65222/VIRAL.2026.10.55.75

without meaningful human oversight.

These propositions reveal a potential **authenticity gap** between the apparent interpersonal quality of a message and employees' perceptions of genuine human involvement. The gap is not assumed to arise from AI use alone; rather, it depends on how technological mediation is interpreted within an existing leadership relationship. Accordingly, the framework suggests that managerial responsibility remains relevant even when linguistic production is delegated to AI. These findings are theoretical propositions derived from conceptual integration and require empirical validation before their direction, magnitude, or generalizability can be established.

5. Discussion

The Leadership Authenticity - AI Mediation Framework advances the understanding of AI-mediated managerial communication by suggesting that perceived authenticity depends less on technological involvement itself than on how employees interpret the leader's intentions, engagement, and accountability. This distinction extends Authentic Leadership Theory beyond its conventional emphasis on interpersonal behavior by considering situations in which managerial expressions are technologically produced or modified. While AI may improve clarity, consistency, and responsiveness, these functional benefits do not necessarily communicate the benevolence and integrity underlying interpersonal trust. The framework therefore challenges the assumption that greater communicative sophistication automatically strengthens leadership relationships.

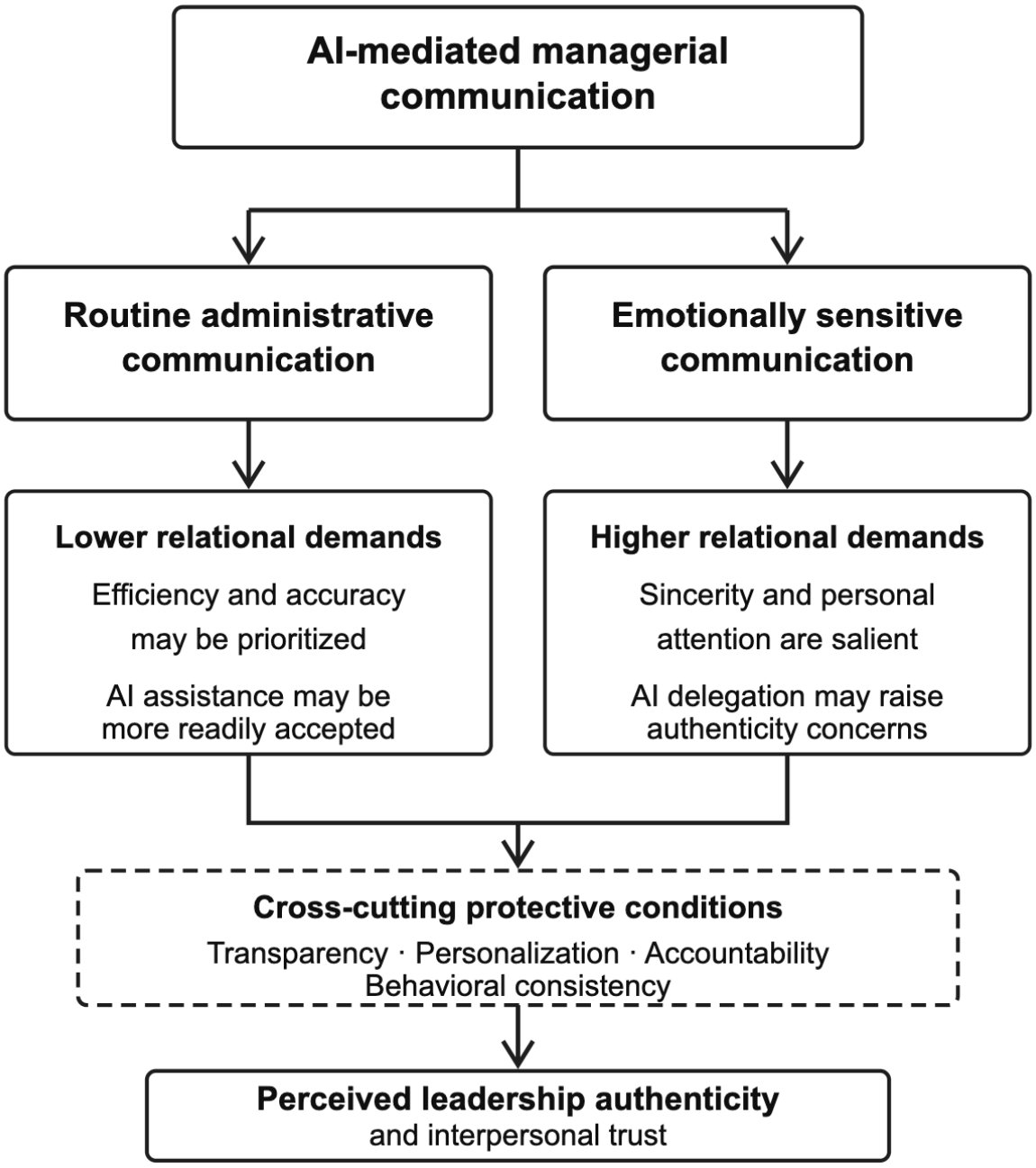
The proposed relationships also extend Social Exchange Theory by treating managerial communication as a potential signal of relational investment. Employees may interpret personally composed messages as evidence that a leader has devoted attention and effort to their concerns. Conversely, extensive AI involvement may weaken this signal when employees believe that the leader has delegated not only linguistic production but also the interpersonal responsibility associated with the message. Nevertheless, AI assistance may preserve or enhance relational exchange when it enables more thoughtful communication and is accompanied by demonstrable human engagement. The theoretical distinction is therefore between delegating communicative tasks and relinquishing relational responsibility.

Attribution-based reasoning further explains why identical AI-assisted messages may generate different interpretations. In routine administrative communication, employees may prioritize accuracy and efficiency, making technological assistance relatively unproblematic. In emotionally consequential exchanges, including performance feedback, conflict resolution, or expressions of support, recipients may place greater importance on personal intentionality and emotional congruence. Authorship transparency introduces an additional tension: disclosure may reinforce honesty while simultaneously highlighting reduced human participation. Accordingly, transparency should not be understood as a sufficient condition for authenticity. Its potential benefits depend on whether leaders retain meaningful oversight, personalize sensitive messages, and demonstrate behavioral consistency.

Figure 2 illustrates the proposed contextual conditions under which AI-mediated managerial communication may either preserve or weaken perceived leadership authenticity and interpersonal trust

Figure 2. Contextual Conditions Shaping the Relational Impact of AI-Mediated Leadership Communication

DOI: https://doi.org/10.65222/VIRAL.2026.10.55.75



Note: The figure illustrates theoretically proposed contextual differences, not empirically established effects. Protective conditions may influence outcomes across both communication contexts.

Source: Authors' own conceptual contribution.

These implications are particularly relevant to educational leadership, where communication often concerns professional development, institutional change, and staff well-being. School and university leaders may reasonably use generative AI to improve routine announcements, summarize complex information, or structure administrative correspondence. However, communication involving professional recognition, difficult feedback, or individual distress may require greater direct human involvement. The framework therefore supports a differentiated approach to AI adoption based on the relational significance of the message rather than a universal preference for either automation or exclusively human authorship.

The study also has limitations. Its propositions are theoretically derived and do not establish the magnitude or direction of actual employee responses. Perceptions of authenticity may vary according

DOI: https://doi.org/10.65222/VIRAL.2026.10.55.75

to organizational culture, prior trust, employees' familiarity with AI, and individual expectations concerning managerial communication. Moreover, the framework does not assume that human-authored messages are inherently sincere or that AI-assisted messages are necessarily deceptive. Future experimental studies could manipulate the degree of AI involvement, disclosure, and communication sensitivity, while longitudinal research could investigate whether repeated AI-mediated interactions influence trust over time. Such investigations would help determine when technological augmentation supports, rather than compromises, authentic leadership.

6. Conclusion

The growing use of generative artificial intelligence in managerial communication raises important questions about the relationship between technological efficiency, perceived leadership authenticity, and interpersonal trust. This conceptual study addresses these questions through the Leadership Authenticity–AI Mediation Framework, integrating Authentic Leadership Theory, Social Exchange Theory, and Attribution Theory to explain how employees may interpret AI-assisted messages attributed to human leaders.

The framework suggests that AI involvement does not inherently undermine authentic leadership. Rather, its potential relational consequences depend on whether employees perceive genuine managerial intentionality, emotional congruence, and continued human accountability. The proposed relationships further indicate that communication sensitivity and authorship transparency may shape these interpretations. While AI can enhance the clarity and efficiency of routine communication, emotionally consequential exchanges may require more visible personal engagement to sustain perceptions of sincerity and trust.

The study contributes to leadership scholarship by distinguishing the technological production of communication from the human responsibility associated with its meaning and consequences. For educational leaders, this distinction suggests that responsible AI integration should prioritize relational judgment alongside operational efficiency, particularly in communication involving staff development, recognition, and well-being.

As a theory-building investigation, the study does not establish empirical causal effects. Its propositions provide a foundation for future research examining how AI involvement, disclosure practices, and behavioral consistency influence employee perceptions across organizational and educational contexts. Ultimately, the framework proposes that the central challenge of AI-mediated leadership is not whether technology participates in communication, but whether leaders remain recognizably responsible for the intentions, commitments, and relationships their messages represent.

7. References

[1] Avolio BJ, Gardner WL. Authentic leadership development: Getting to the root of positive forms of leadership. *The Leadership Quarterly*. 2005;16(3):315–338. https://doi.org/10.1016/j.leaqua.2005.03.001 .

[2] Walumbwa FO, Avolio BJ, Gardner WL, Wernsing TS, Peterson SJ. Authentic leadership: Development and validation of a theory-based measure. *Journal of Management*. 2008;34(1):89–126. https://doi.org/10.1177/0149206307308913 .

[3] Liu H, Cutcher L, Grant D. Doing authenticity: The gendered construction of authentic leadership. *Gender, Work & Organization*. 2015;22(3):237–255.

[4] Alvesson M, Einola K. Warning for excessive positivity: Authentic leadership and other traps in leadership studies. *The Leadership Quarterly*. 2019;30(4):383–395.

[5] Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. *Academy of Management Review*. 1995;20(3):709–734. https://doi.org/10.5465/amr.1995.9508080335 .

[6] Cropanzano R, Mitchell MS. Social exchange theory: An interdisciplinary review. *Journal of Management*. 2005;31(6):874–900. https://doi.org/10.1177/0149206305279602 .

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

[8] Raisch S, Krakowski S. Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*. 2021;46(1):192–210. https://doi.org/10.5465/amr.2018.0072 .

[9] Jarrahi MH. Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making. *Business Horizons*. 2018;61(4):577–586. https://doi.org/10.1016/j.bushor.2018.03.007 .

DOI: https://doi.org/10.65222/VIRAL.2026.10.55.75

[10] Shrestha YR, Ben-Menahem SM, von Krogh G. Organizational decision-making structures in the age of artificial intelligence. *California Management Review*. 2019;61(4):66–83. https://doi.org/10.1177/0008125619862257 .
[11] Gambino A, Fox J, Ratan RA. Building a stronger CASA: Extending the Computers Are Social Actors paradigm. *Human-Machine Communication*. 2020;1:71–86. https://doi.org/10.30658/hmc.1.5 .