International Journal of Education, Leadership, Artificial Intelligence, Computing, Business, Life Sciences, and Society - https://journalviral.org
DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

Beyond Generative AI: Claude Agents and the Emergence of Human-AI Governance in Educational Leadership

Bogdan Costache
Bucharest University of Economic Studies

ABSTRACT: *The emergence of autonomous artificial intelligence agents represents a new phase in the evolution of generative artificial intelligence, extending AI capabilities beyond content generation toward autonomous reasoning, workflow orchestration, institutional coordination, and organizational decision support. Among these developments, Claude Agents exemplify a new generation of agentic AI systems capable of executing complex multi-step tasks, managing institutional information, and interacting continuously with human users across diverse educational environments. Although generative artificial intelligence has attracted substantial scholarly attention, research remains largely centered on pedagogical applications, while the governance and leadership implications of autonomous AI agents continue to be conceptually fragmented and insufficiently theorized.*

This paper develops an integrative conceptual framework that examines how Claude Agents may transform educational leadership, institutional governance, and strategic decision-making in schools and higher education institutions. Drawing upon interdisciplinary scholarship in educational leadership, organizational governance, socio-technical systems theory, organizational information processing theory, human-AI collaboration, and digital transformation, the study adopts a conceptual qualitative methodology based on an integrative literature review and theory-building approach. The analysis argues that Claude Agents should not be understood merely as intelligent assistants but as agentic organizational actors that increasingly participate in institutional information processing, policy implementation, strategic planning, administrative coordination, and evidence-informed decision-making. Their integration creates opportunities for more adaptive, transparent, and data-informed governance while simultaneously introducing new challenges related to accountability, explainability, algorithmic bias, professional autonomy, institutional legitimacy, and ethical oversight.

Building on these insights, the paper proposes the Human-AI Governance Framework for Educational Leadership (HAGF), which conceptualizes leadership as a collaborative governance process in which human judgment and autonomous AI agents jointly contribute to organizational decision-making within clearly defined institutional, ethical, and regulatory boundaries. The study contributes to emerging debates on agentic artificial intelligence by extending existing theories of educational leadership beyond technology adoption toward a governance-oriented perspective that integrates organizational resilience, distributed intelligence, and responsible AI governance. Finally, the paper identifies a future research agenda focused on AI-enabled leadership, institutional trust, governance architectures, and the evolving relationship between educational leaders and autonomous intelligent agents.

KEYWORDS: Claude Agents; agentic artificial intelligence; AI-assisted decision-making; school governance; distributed leadership; organizational resilience; artificial intelligence in education.

1. Introduction

The rapid diffusion of generative artificial intelligence has transformed international debates on educational innovation, institutional governance, and organizational leadership. Within only a few years, educational institutions have moved from experimenting with conversational systems capable of generating text toward engaging with increasingly autonomous artificial intelligence agents that are able to plan, reason, retrieve information,

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

coordinate workflows, and execute complex sequences of tasks with limited human intervention. This transition represents more than a technological refinement. It signals a qualitative shift in the organizational role assigned to artificial intelligence, moving from decision support toward participation in institutional processes that have traditionally been considered exclusively human domains. Consequently, educational leadership is entering a period in which governance structures must accommodate not only digital technologies but also autonomous computational actors capable of influencing administrative coordination, strategic planning, policy implementation, and organizational learning.

Most scholarly discussions surrounding artificial intelligence in education have concentrated on instructional innovation, personalized learning, assessment practices, academic integrity, and student engagement. The widespread adoption of large language models following the public release of ChatGPT stimulated an unprecedented body of research examining opportunities and risks associated with generative AI in teaching and learning environments. Reviews published during the past two years reveal considerable progress in understanding pedagogical applications while simultaneously acknowledging persistent concerns regarding hallucinations, ethical use, privacy, transparency, and assessment integrity (Papa & Jackson, 2021; Richardson et al., 2025). Nevertheless, the literature remains predominantly classroom-oriented. Comparatively less attention has been devoted to understanding how autonomous AI agents may reshape the organizational architecture of educational institutions and redefine the nature of leadership itself. This imbalance has become increasingly evident as recent developments extend beyond conversational assistance toward systems capable of autonomous planning, persistent memory, multi-agent collaboration, and workflow execution. Claude Agents exemplify this emerging generation of agentic artificial intelligence. Unlike conventional generative AI applications designed primarily to respond to isolated prompts, agentic systems are increasingly capable of maintaining contextual continuity across extended interactions, coordinating multiple information sources, decomposing complex objectives into executable subtasks, and adapting their behavior in response to evolving organizational requirements. Rather than functioning solely as sophisticated information retrieval systems, these agents begin to assume responsibilities traditionally associated with managerial support, organizational coordination, and institutional decision-making. Such developments require educational researchers to reconsider prevailing assumptions regarding leadership, authority, accountability, and professional judgment within digitally mediated educational organizations. As Presti et al. (2025) argue in their discussion of collaborative multi-agent systems, the educational relevance of advanced AI increasingly lies not only in individual interactions with learners but also in coordinated organizational processes supported by networks of intelligent agents.

These developments resonate with broader transformations occurring across management and organizational research. Contemporary leadership scholarship increasingly acknowledges that decision-making has become distributed across complex socio-technical systems in which human expertise is complemented by algorithmic reasoning, predictive analytics, and autonomous computational support (Appelo, 2025; Becker et al., 2026; Woods, 2026). Rather than replacing human leaders, emerging AI agents contribute to information processing, pattern recognition, scenario simulation, and administrative coordination, thereby modifying how organizations interpret uncertainty and formulate strategic responses. Dai et al. (2025) describe this evolution as a movement toward symbiotic collaboration in which artificial

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

intelligence augments rather than substitutes educational leadership. Similarly, Sollazzo et al. (2026) argue that the increasing autonomy of agentic AI necessitates governance arrangements capable of balancing organizational efficiency with appropriate mechanisms of human oversight and institutional accountability. These arguments suggest that leadership can no longer be examined independently of the intelligent technological infrastructures through which organizational decisions are increasingly produced.

From an educational perspective, this transition raises questions extending well beyond technology adoption. School leaders have traditionally been understood as individuals responsible for articulating institutional vision, supporting professional development, cultivating organizational culture, allocating resources, and ensuring accountability toward multiple stakeholders. Digital leadership research has expanded this perspective by demonstrating that technological competence, organizational trust, and innovation capacity increasingly influence leadership effectiveness (Agustina et al., 2020). However, agentic AI introduces an additional dimension by creating organizational environments in which artificial intelligence systems become active participants in institutional information flows rather than passive technological instruments. Richardson et al. (2025) note that school leaders already perceive artificial intelligence as influencing communication practices, administrative routines, and strategic decision-making, yet conceptual frameworks explaining these transformations remain underdeveloped. Existing leadership theories therefore provide only partial explanations for governance arrangements in which autonomous AI agents contribute continuously to organizational intelligence.

Theoretical developments in adjacent disciplines provide valuable foundations for addressing this emerging phenomenon. Socio-technical systems theory emphasizes that organizational performance depends upon the reciprocal adaptation of technological infrastructures and human social systems rather than the optimization of either component independently. Organizational Information Processing Theory similarly suggests that increasingly uncertain organizational environments require mechanisms capable of expanding information processing capacity and reducing decision ambiguity. Within educational institutions characterized by rapidly changing regulatory environments, increasing data complexity, and heightened stakeholder expectations, Claude Agents may significantly extend organizational information-processing capabilities while simultaneously generating new governance challenges associated with explainability, responsibility, transparency, and institutional legitimacy. Human-AI collaboration literature likewise suggests that effective organizational outcomes emerge not through technological substitution but through complementary configurations combining computational efficiency with human judgment, ethical reasoning, contextual interpretation, and professional experience (Cheong, 2025; Walther, 2024).

Despite these emerging theoretical perspectives, the concept of **Human-AI Governance** remains insufficiently articulated within educational leadership research. Existing scholarship often discusses artificial intelligence either as an instructional technology or as an administrative tool without fully considering how autonomous AI agents alter governance structures themselves. Governance encompasses more than operational efficiency; it involves the distribution of authority, mechanisms of accountability, institutional legitimacy, ethical oversight, regulatory compliance, and collective responsibility for organizational decisions. As artificial intelligence acquires greater autonomy in organizational processes, governance becomes an essential analytical lens through which educational leadership can be

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

reconsidered. The challenge is therefore not merely technological adoption but the construction of governance architectures capable of integrating autonomous AI agents while preserving human accountability, professional autonomy, institutional trust, and democratic educational values.

Against this background, the present paper addresses the following research question: **How do Claude Agents reshape educational leadership and contribute to the emergence of Human-AI Governance within contemporary educational institutions?** To address this overarching question, the analysis further examines three interrelated issues. First, it explores how agentic artificial intelligence differs conceptually from earlier forms of generative AI within educational organizations. Second, it investigates how autonomous AI agents may transform institutional leadership, organizational coordination, and strategic decision-making. Third, it develops a conceptual governance framework capable of explaining collaborative relationships between educational leaders and autonomous intelligent agents under conditions of increasing organizational complexity.

Rather than evaluating the technical performance of Claude Agents or assessing specific educational implementations, this article adopts a theory-building perspective that synthesizes contemporary scholarship across educational leadership, organizational governance, socio-technical systems, human-AI collaboration, and agentic artificial intelligence. The central argument advanced throughout the paper is that educational leadership is gradually evolving toward a governance model in which human expertise and autonomous AI agents jointly participate in organizational intelligence while remaining embedded within clearly defined ethical, institutional, and regulatory structures. By conceptualizing this transition through the notion of **Human-AI Governance**, the paper seeks to extend current debates beyond technology adoption and contribute a governance-oriented framework capable of informing future empirical research on educational leadership in the era of autonomous artificial intelligence.

2. Literature Review

2.1. From Generative Artificial Intelligence to Agentic Systems

The educational literature on artificial intelligence has developed through several partially overlapping phases. Earlier work concentrated on intelligent tutoring systems, automated assessment, learning analytics, and recommendation technologies whose functions were relatively bounded and whose operation depended on predefined instructional or administrative objectives. The diffusion of large language models altered this landscape by making natural-language generation, summarization, conversational assistance, and knowledge synthesis available through general-purpose interfaces. Yet the conceptual vocabulary used to describe these systems has often remained tied to the image of AI as a tool activated by a human user. Such a framing is increasingly inadequate for systems that can decompose goals, select tools, maintain contextual memory, coordinate sequences of actions, and revise their conduct in response to intermediate results.

Agentic artificial intelligence differs from conventional generative AI principally through the organization of action around goals rather than isolated prompts. A generative model may

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

produce a policy summary or draft a communication when requested. An agentic system may retrieve the relevant policy corpus, compare institutional requirements, identify inconsistencies, prepare alternative responses, route documents for review, monitor subsequent actions, and update the workflow as circumstances change. The distinction is therefore not simply one of performance. It concerns temporal continuity, delegated initiative, interaction with external systems, and the capacity to organize several computational operations around an institutional objective. Russell and Norvig (2021) define intelligent agency in terms of systems that perceive an environment and select actions in pursuit of goals. Contemporary large language model agents extend this principle by combining linguistic reasoning with memory, planning, retrieval, tool use, and workflow orchestration.

Presti et al. (2025) place this development within an educational trajectory leading from individual large language model applications toward collaborative multi-agent systems. Their account is significant because it shifts the unit of analysis from a single user-model exchange to an ecology of specialized agents capable of coordinating educational functions. One agent might retrieve regulations, another analyze institutional data, while a third prepares options for human deliberation. Although such configurations remain emergent, they make visible a form of computational division of labor that differs materially from conventional chatbot use. The organizational consequences arise not merely because AI produces better answers, but because interconnected agents may mediate how tasks are assigned, sequenced, verified, and recorded.

This movement toward goal-directed autonomy is also apparent in management research. Appelo (2025) interprets intelligent agents as participants in algorithmically mediated organizations, where managerial work becomes distributed across human actors, digital infrastructures, and automated coordination mechanisms. Woods (2026) similarly presents AI-enabled leadership as dependent on faster information processing and decision support, although speed alone cannot constitute a defensible leadership model. What becomes analytically important is the redistribution of cognitive and coordinative work. Becker et al. (2026) extend this argument by showing how customized conversational agents can be incorporated into leadership and management research as interactive collaborators. Their methodological focus also carries an organizational implication: once an agent is configured around a role, corpus, or decision process, it ceases to operate as a generic interface and begins to participate in a structured practice.

The emerging agentic paradigm should nevertheless be approached cautiously. Claims of autonomy may obscure the extent to which agents remain dependent on human-defined objectives, institutional data, system permissions, interface design, and model-generated inferences that may be unreliable. Agentic AI does not possess organizational authority in a legal or moral sense merely because it can execute a workflow. Nor does technical capacity establish legitimate agency. The term “agentic organizational actor” is therefore used in this study functionally rather than anthropomorphically. It denotes a system that participates consequentially in organizational processes, not an entity possessing human intention, responsibility, or institutional standing. This distinction is crucial because imprecise language may encourage inappropriate delegation and blur the location of accountability.

To further clarify the conceptual evolution from conventional generative artificial intelligence to agentic AI, Table 1 summarizes the principal technological, organizational, and governance-related distinctions identified in the literature. Rather than representing

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

incremental improvements in language generation, agentic systems introduce a qualitatively different organizational paradigm characterized by autonomous reasoning, multi-step planning, workflow orchestration, and continuous collaboration with human decision-makers.

Table 1. Evolution from Generative AI to Agentic AI in Educational Organizations

Dimension	Generative AI (LLMs)	Agentic AI (Claude Agents)	Implications for Educational Leadership
Primary function	Content generation and question answering	Goal-oriented planning and autonomous task execution	Supports organizational decision-making rather than isolated information retrieval
User interaction	Prompt-response interaction	Continuous multi-step collaboration	Enables sustained leadership support throughout administrative processes
Decision support	Provides recommendations upon request	Proactively organizes workflows and coordinates tasks	Assists leaders in complex institutional decision-making
Memory	Limited conversational context	Persistent contextual memory across tasks	Improves continuity in strategic planning and organizational learning
Reasoning capability	Generates responses from prompts	Plans, evaluates, revises, and adapts actions dynamically	Facilitates evidence-informed leadership decisions
Workflow management	Performs individual tasks	Coordinates multiple interconnected activities	Enhances organizational efficiency and administrative coordination
Tool integration	Limited external functionality	Integrates databases, documents, calendars, APIs, and digital tools	Supports institution-wide digital ecosystems
Organizational role	Intelligent assistant	Agentic organizational actor	Participates in governance processes under human oversight
Human involvement	Human initiates every interaction	Human supervises autonomous execution	Shifts leadership from direct control toward governance and oversight
Governance requirement	Ethical AI use	Human-AI Governance with accountability, transparency, and institutional oversight	Requires formal governance mechanisms to regulate delegated organizational activities

Source: Authors' synthesis based on Russell and Norvig (2021), Jarrahi (2018), Appelo (2025), Presti et al. (2025), Becker et al. (2026), Woods (2026), Dai et al. (2025), and Sollazzo et al. (2026).

Claude Agents illustrate this transition from episodic generation to sustained organizational assistance. Their analytical importance lies less in brand-specific features than in the broader architecture they represent: language-model-based systems capable of pursuing multi-step objectives, consulting institutional information, using external tools, and maintaining continuity across tasks. Within educational institutions, such capabilities could support policy

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

comparison, meeting preparation, regulatory monitoring, resource planning, communication workflows, and institutional reporting. They also create a new research problem. Once AI systems move from drafting documents to coordinating activities and recommending action, their implications can no longer be contained within conventional discussions of educational technology adoption.

2.2. Educational Leadership, Digital Mediation, and Organizational Judgment

Educational leadership has long involved more than formal authority. It combines sensemaking, institutional representation, resource allocation, professional support, conflict mediation, instructional direction, and accountability to multiple constituencies. These responsibilities are exercised in organizations where objectives are contested, outcomes are difficult to measure, and decisions carry social as well as operational consequences. Mintzberg's (1973) account of managerial work remains relevant because it depicts leadership as fragmented, relational, and information-intensive rather than as a sequence of detached rational choices. Weick (1979) similarly frames organizing as an interpretive accomplishment through which actors construct workable meanings under conditions of ambiguity. These perspectives challenge the assumption that leadership can be reduced to computational optimization.

The development of digital leadership scholarship has nevertheless shown that leaders' capacity to shape technological change matters materially for institutional practice. Agustina et al. (2020) found that principals' digital leadership influenced teachers' reflective practices through trust, self-efficacy, and work engagement. The importance of this finding lies not only in the association between leadership and technology use. It indicates that digital change is socially mediated. Infrastructure does not produce professional adaptation independently of organizational relationships, perceived competence, and trust. The same principle is likely to apply more strongly to agentic AI because these systems enter decision processes that affect professional roles and institutional authority.

Papa and Jackson (2021) place human agency at the center of educational leadership in environments increasingly shaped by artificial intelligence. Their edited volume resists a simple opposition between human leaders and machines, emphasizing instead the need to examine how technological systems reconfigure professional action. This orientation remains useful, although the spread of agentic systems sharpens the issue. Educational leaders are no longer deciding only whether to adopt AI-enabled applications. They may also need to determine which tasks can be delegated, what evidence an agent may access, when recommendations require verification, how affected stakeholders can contest outputs, and where responsibility resides when automated coordination produces adverse consequences.

Richardson et al. (2025) show that school leaders' discussions of AI encompass both practical interest and uncertainty. Administrative efficiency, communication, instructional support, and data use appear alongside concerns about policy, ethics, reliability, and readiness. The coexistence of enthusiasm and hesitation is theoretically important. It suggests that AI adoption in schools is not governed by a stable professional consensus. Leaders interpret the technology through institutional pressures, local capacities, perceived risks, and changing expectations from teachers, students, families, and policymakers. Agentic AI intensifies this ambiguity because the systems do not merely inform decisions but may increasingly structure

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

the process through which information reaches decision-makers.

Dai et al. (2025) address this problem through a framework for symbiotic collaboration in AI-assisted administrative decision-making. Their account rejects both full automation and purely human decision control as analytically incomplete. AI can contribute speed, pattern detection, data synthesis, and consistency, while educational leaders retain contextual understanding, normative judgment, political sensitivity, and responsibility for consequences. Such complementarity is persuasive, but it should not be treated as automatic. Symbiosis depends on task design, data quality, institutional safeguards, and the capacity of leaders to interrogate machine-generated outputs. Without these conditions, complementarity may become asymmetrical dependence, with human actors formally retaining authority while practically following recommendations they cannot adequately inspect.

Research on leadership development reinforces this concern. Aboumoussa and Pfister (2024) argue that the age of artificial intelligence requires reconsideration of leadership capabilities, particularly where judgment, adaptation, and human relationships intersect with machine-supported analysis. Drożdżyńska and Kucharski (2025), through scenario-based assessments of large language models on managerial soft-skills tasks, draw attention to the limits of equating linguistically plausible responses with situated managerial competence. Leadership problems often involve incomplete information, conflicting values, tacit knowledge, and emotionally charged relationships. Performance on a textual scenario may reveal useful analytical capacity, but it does not establish that an AI system can assume responsibility for the interpersonal and moral dimensions of leadership.

Kanbakan et al. (2026) offer a related perspective through an AI-based resuscitation simulation designed for team leadership training. Although situated in medical education rather than school administration, the study illustrates a defensible use of AI in leadership contexts: the system supports simulation, feedback, and practice while human participants remain the subjects of professional development. This differs categorically from delegating actual high-stakes leadership decisions to an agent. The comparison clarifies a continuum of institutional uses, extending from training and rehearsal to advisory support, coordinated execution, and potentially autonomous action. Governance requirements become more stringent as systems move along that continuum.

Walther (2024) supplies an ethical counterweight to accounts centered on efficiency and decision speed. Human leadership for humane technology implies that institutional uses of AI should remain accountable to human dignity, relational responsibility, and social purpose. In education, these commitments cannot be treated as supplementary values added after implementation. They form part of the legitimacy of educational leadership itself. A technically accurate decision may still be educationally inappropriate when it disregards vulnerability, procedural fairness, professional identity, or community trust. Agentic AI therefore creates a dual demand: leaders must understand what the systems can do while preserving the capacity to judge what institutions ought to do.

2.3. Decision-Making, Information Processing, and the Limits of Delegation

The attraction of agentic AI in educational administration is closely connected to the informational burden placed on contemporary institutions. Schools and universities process student records, staffing data, curriculum requirements, financial constraints, quality

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

indicators, safeguarding obligations, accreditation standards, and rapidly changing regulations. Galbraith's (1974) Organizational Information Processing Theory proposes that organizations facing uncertainty must either reduce the need for information or increase their capacity to process it. Agentic AI appears to offer the latter by synthesizing dispersed data, monitoring events, preparing scenarios, and coordinating responses across systems.

This perspective helps explain why educational organizations may adopt agents even in the absence of strong evidence that they improve leadership. The systems promise relief from information overload and administrative fragmentation. Yet increased processing capacity does not necessarily produce better judgment. Simon's (1977) account of bounded rationality distinguishes computational availability from substantive rationality. Decision-makers operate with limited attention, incomplete knowledge, and constrained time. AI may extend search and comparison, but it can also introduce new limits through opaque model behavior, uncertain provenance, automation bias, and an inability to represent tacit institutional knowledge. In this sense, agentic systems may alter rather than eliminate bounded rationality.

Doumanas et al. (2026), examining causal reasoning and large language models in military decision-making, provide a useful comparison from another high-stakes domain. Their concern with command structures illustrates how AI-supported causal analysis may redistribute information and challenge established decision hierarchies. Educational leadership differs from military command in purpose, institutional culture, and acceptable forms of authority. Even so, both contexts involve consequential decisions under uncertainty, heterogeneous information, and the need to preserve accountable command. The comparison suggests that the governance of AI-supported reasoning cannot be separated from the design of authority structures. A recommendation's influence depends not only on its content but on where it enters a decision process, who can challenge it, and whether alternatives remain visible.

March's (1991) distinction between exploration and exploitation further clarifies the possible contribution of Claude Agents. Agents may support exploitation by standardizing recurring processes, retrieving precedents, and improving the consistency of established routines. They may also facilitate exploration by generating scenarios, identifying unfamiliar patterns, and comparing strategic options. These functions create different risks. Excessive exploitation may encode historical routines and institutional biases, while unconstrained exploration may generate attractive but weakly grounded alternatives. Educational leaders must therefore govern not only the accuracy of AI outputs but the organizational learning logic that agents reinforce.

The delegation of tasks to AI should consequently be differentiated from the delegation of authority. An agent may prepare a staffing analysis, summarize stakeholder feedback, or monitor compliance deadlines without receiving the institutional authority to determine appointments, interpret community priorities, or decide sanctions. This distinction appears straightforward but may become blurred in practice. When an agent controls the sequence of information, defines the available options, or triggers automated workflows, it exercises practical influence even when formal authority remains human. Human approval can become ceremonial if leaders lack the time, expertise, or system access needed to conduct meaningful review.

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

A defensible delegation architecture would therefore require graduated levels of autonomy. Low-risk, reversible, and well-specified tasks may permit greater automated execution. Decisions affecting rights, opportunities, professional status, resource distribution, or student welfare require stronger human control and documented justification. The relevant criterion is not merely technical confidence but institutional consequence. This reasoning supports a move away from binary debates about whether AI should make decisions. The more precise concern is how different forms of participation should be configured across different classes of decisions.

2.4. Human-AI Collaboration, Co-Agency, and Professional Authority

Human-AI collaboration is frequently presented as a middle path between technological replacement and human exclusivity. Jarrahi (2018) describes human-AI symbiosis as a configuration in which machine capabilities complement human intuition, contextual understanding, and judgment. The concept is attractive because it acknowledges differentiated strengths without presuming that one form of intelligence should dominate the other. In educational leadership, however, collaboration must be interpreted institutionally. Leaders do not act only as individual decision-makers. They represent legal responsibilities, professional norms, organizational histories, and public values that an AI system cannot independently bear.

Cheong (2025) exposes some of the tensions hidden by harmonious accounts of collaboration. Human-machine communication may facilitate coordination while simultaneously disturbing established understandings of authority, identity, and relational practice. Her analysis of leadership tensions is particularly relevant because agentic systems can appear socially present and communicatively competent even when their outputs are generated without experience, commitment, or moral accountability. Fluency may therefore produce an impression of partnership that exceeds the system's actual epistemic or ethical standing.

The concept of co-agency needs similar qualification. In a descriptive sense, human and artificial actors may jointly shape an outcome. A school leader defines a problem, an agent retrieves information and proposes options, and a leadership team interprets the result. Agency is distributed across the process. In a normative sense, however, the actors are not equivalent. The human participants remain accountable for institutional purposes, legal compliance, and the effects of decisions on persons. Artificial agents possess delegated operational capacity, not moral or democratic authority. Human-AI Governance must preserve this asymmetry rather than conceal it beneath the language of partnership.

Appelo's (2025) treatment of human-robot-agent configurations is useful in highlighting that leadership increasingly occurs within hybrid organizational arrangements. Yet algorithmic management also presents risks of deprofessionalization when standardized metrics or automated workflows narrow the scope of judgment. Educational institutions are especially vulnerable to this problem because important outcomes are not always immediately measurable. Student development, inclusion, trust, professional growth, and community legitimacy resist complete translation into operational indicators. An agent optimized around available data may privilege measurable proxies while marginalizing less visible educational values.

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

Woods (2026) emphasizes the capacity of AI to support faster and more informed leadership decisions. Such capacity can be valuable, particularly during periods of disruption or administrative pressure. Speed, however, may reduce opportunities for deliberation. Educational governance often requires explanation, consultation, and procedural fairness, even where a technically plausible answer is readily available. A human-AI arrangement should therefore be judged not only by decision accuracy or efficiency but by whether it preserves the institutional conditions through which decisions become legitimate.

The resulting conception of collaborative intelligence is deliberately restrained. Claude Agents may extend leaders' analytical reach, reduce routine workload, identify inconsistencies, and support scenario development. Educational leaders contribute professional judgment, ethical interpretation, relational knowledge, and responsibility. Neither contribution is sufficient in every context, yet the relationship is not one of equal standing. The central design problem is to combine computational participation with unambiguous human accountability.

2.5. From AI Governance to Human-AI Governance

AI governance literature generally addresses principles, controls, and institutions intended to regulate the development and use of artificial intelligence. Floridi and Cowls (2019) organize this normative field around beneficence, non-maleficence, autonomy, justice, and explicability. These principles remain highly relevant to education, where AI systems may influence access, evaluation, resource allocation, and professional opportunities. Their application becomes more difficult, however, when AI is embedded in recurring workflows rather than used for isolated outputs. Governance must then extend from model-level properties to the organizational arrangements through which agents receive data, permissions, objectives, and opportunities to act.

Sollazzo et al. (2026) frame the governance challenge of agentic AI as a balance between autonomy and oversight. This formulation captures a central tension but should not imply that autonomy and oversight are simply opposing quantities. Excessive autonomy may create unaccountable action, while excessive control may render the agent ineffective or conceal informal reliance on ungoverned tools. The more productive question concerns how oversight is designed. Human review must be timely, informed, and consequential. Audit logs must allow reconstruction of actions. Escalation rules must identify circumstances in which an agent cannot proceed. Institutional policies must specify who owns the data, who configures the system, who can contest an output, and who bears responsibility for harm.

Dai et al. (2025) provide a basis for moving from governance of AI toward governance with AI. Their model of symbiotic collaboration implies that AI participates in administrative decision processes while human leaders retain evaluative and contextual roles. Human-AI Governance develops this insight at the institutional level. It concerns the structures through which human and artificial contributions are combined, constrained, documented, and rendered accountable. The object of governance is therefore not only the technology. It is the hybrid decision system.

To consolidate the interdisciplinary perspectives reviewed in this section, Table 2 compares the principal theoretical approaches that inform Human-AI Governance in educational

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

leadership. The comparison demonstrates that existing studies emphasize complementary dimensions of governance, leadership, organizational design, and human-AI collaboration. Building upon these perspectives, the present study integrates these fragmented contributions into the proposed Human-AI Governance Framework for Educational Leadership.

Table 2. Comparative Perspectives on Human-AI Governance in Educational Leadership

Theoretical Perspective	Primary Focus	Contribution to Human-AI Governance	Key References
Organizational Information Processing Theory	Organizational decision-making under uncertainty	Explains how AI agents increase institutional information-processing capacity and support complex decision-making	Galbraith (1974); Dai et al. (2025)
Socio-Technical Systems Theory	Interaction between technology and organizational structures	Positions AI agents as components of integrated socio-technical systems rather than isolated technologies	Trist & Bamforth (1951); Papa & Jackson (2021)
Human-AI Collaboration Theory	Complementarity between human and artificial intelligence	Defines collaborative decision-making based on the complementary strengths of humans and AI agents	Jarrahi (2018); Dai et al. (2025); Cheong (2025)
AI Governance	Ethical, legal, and institutional oversight of AI	Establishes principles of transparency, accountability, fairness, explainability, and responsible AI deployment	Floridi & Cowls (2019); Sollazzo et al. (2026)
Educational Leadership Theory	Leadership, organizational effectiveness, and institutional decision-making	Explains how leadership responsibilities evolve within AI-supported educational organizations	Richardson et al. (2025); Agustina et al. (2020); Walther (2024)
Agentic AI	Autonomous reasoning, planning, workflow orchestration, and multi-agent collaboration	Introduces AI agents as organizational actors capable of supporting institutional governance	Appelo (2025); Becker et al. (2026); Woods (2026); Presti et al. (2025)
Human-AI Governance (This Study)	Collaborative governance between educational leaders and autonomous AI agents	Integrates leadership, governance, socio-technical systems, and agentic AI into a unified conceptual framework for educational organizations	Authors' conceptualization

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

Source: Authors' synthesis based on Galbraith (1974), Trist and Bamforth (1951), Jarrahi (2018), Floridi and Cowls (2019), Papa and Jackson (2021), Agustina et al. (2020), Richardson et al. (2025), Appelo (2025), Dai et al. (2025), Presti et al. (2025), Becker et al. (2026), Sollazzo et al. (2026), Walther (2024), and Woods (2026).

This distinction matters because educational institutions may satisfy formal AI policies while leaving hybrid workflows poorly governed. A school could require human approval for AI-generated recommendations yet provide no standard for meaningful review. A university might prohibit automated decisions but permit agents to rank cases, summarize evidence, and frame options in ways that strongly influence the final outcome. Formal human control would remain intact, while substantive control would have shifted toward the computational system. Human-AI Governance attends to this gap between nominal authority and practical influence.

The governance architecture also has to address professional autonomy. Educational leaders and teachers exercise discretion because institutional rules cannot anticipate every context. Agentic systems may support discretion by providing richer information, but they may also constrain it through standardized workflows, default recommendations, and persistent monitoring. Walther's (2024) humane technology perspective suggests that systems should be assessed partly by whether they preserve the conditions for responsible human action. Professional autonomy should not be interpreted as immunity from evidence or accountability. It denotes the capacity to exercise informed judgment, explain departures from automated recommendations, and remain answerable for decisions.

Institutional legitimacy provides a further dimension. Educational decisions derive legitimacy not only from outcome quality but from transparency, participation, lawful authority, and consistency with public purposes. AI-generated recommendations may be technically sophisticated while remaining illegitimate if affected stakeholders cannot understand or challenge their use. Governance mechanisms should therefore include procedural rights, clear communication about AI involvement, documentation of human review, and channels for appeal. These requirements become increasingly important as agents participate in decisions concerning admissions, staffing, student support, resource allocation, or disciplinary processes.

2.6. Claude Agents, Multi-Agent Configurations, and Educational Organizations

The educational significance of Claude Agents is best understood within the wider development of large language model agent architectures rather than through a narrow comparison among commercial systems. Their relevance lies in the possibility of configuring persistent, tool-using, role-specific agents around institutional tasks. Such configurations may connect policy repositories, calendars, document systems, analytics platforms, and communication channels. The result is not merely a more capable chatbot but a computational layer capable of coordinating information across organizational boundaries.

Presti et al. (2025) anticipate a progression toward collaborative large language model multi-agent systems in education. Multi-agent configurations allow functions to be distributed among systems with different roles, access rights, and evaluative responsibilities. A policy-analysis agent could compare regulations, a planning agent could model implementation options, and a verification agent could inspect the evidential basis of recommendations. From a governance perspective, modularity may improve traceability by making roles explicit. It may also increase complexity, since failures can emerge through

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

interaction among agents rather than within a single model.

The distinction between a single agent and a multi-agent system has consequences for accountability. When one system produces an output, the decision chain may already be difficult to reconstruct. When several agents exchange intermediate representations, select tools, and revise one another's outputs, responsibility becomes more diffuse. Educational institutions would need to document not only the final recommendation but the sequence of computational actions that produced it. This creates requirements for observability, role separation, access control, and independent validation.

Claude-based agents may also be customized around institutional documents and localized operating procedures. Such customization can improve relevance, but it creates a risk of institutionalizing outdated, contradictory, or discriminatory practices. Retrieval from an approved corpus does not guarantee normative soundness. An agent may accurately reproduce a policy that is no longer appropriate, or reconcile conflicting documents in a manner that reflects model assumptions rather than authorized interpretation. Human expertise remains necessary to determine which sources have institutional authority and whether existing rules should continue to govern practice.

The organizational value of multi-agent systems should therefore be assessed against more than task completion. Relevant criteria include evidential reliability, reversibility, transparency, effects on professional work, distribution of authority, data protection, stakeholder contestability, and consistency with educational purposes. In some settings, a simpler decision-support tool may be preferable to an autonomous agent. Greater technical agency is not inherently a sign of institutional maturity. The appropriate level of autonomy depends on the risk, ambiguity, and normative density of the task.

2.7. Ethical Tensions: Accountability, Trust, Bias, and Institutional Legitimacy

The delegation of organizational tasks to Claude Agents raises ethical problems that cannot be resolved through technical safeguards alone. Accountability is the most immediate concern. An agent may assemble evidence, recommend action, and initiate workflows, but it cannot assume legal or moral responsibility for institutional consequences. Accountability must remain attached to identifiable human roles and governance bodies. Yet retaining a human signature at the end of a process is insufficient when the computational pathway remains opaque. Responsibility requires the capacity to understand, question, revise, and justify the role played by the agent.

Trust presents a related but distinct problem. Agustina et al. (2020) show that trust mediates the relationship between digital leadership and professional practice. In agentic environments, trust operates at several levels: trust in the model, in institutional data, in the leaders authorizing its use, and in the procedures established for review and appeal. Trust should not be understood as unquestioning confidence. Well-governed trust is calibrated, conditional, and supported by evidence. Institutions should know where agents perform reliably, where uncertainty remains high, and which decisions require independent verification.

Algorithmic bias is especially difficult to manage in systems that synthesize heterogeneous information. Bias may arise from model training, institutional datasets, retrieval corpora, problem framing, or the objectives assigned to the agent. Even when protected characteristics are excluded, historical patterns may remain encoded through proxy variables and

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

organizational routines. The governance response cannot be limited to testing model outputs in the abstract. It requires examination of how agents alter institutional attention, whose information is represented, which forms of evidence are privileged, and whether affected groups can challenge conclusions.

Institutional legitimacy depends on how these issues are handled publicly. Educational organizations occupy a distinctive moral and civic position. Their decisions affect learning opportunities, professional careers, social inclusion, and community confidence. Papa and Jackson (2021) emphasize the continuing significance of human agency in educational leadership, while Walther (2024) situates technology within a broader commitment to humane institutions. Together, these perspectives suggest that legitimate AI adoption must remain subordinate to educational purposes. Efficiency may support those purposes, but it cannot define them.

2.8. Research Gap and Conceptual Direction

The reviewed literature converges on several propositions. Artificial intelligence is becoming more deeply embedded in educational administration; leadership increasingly depends on interaction with computational systems; human judgment remains indispensable; and governance mechanisms must develop alongside technical capability. Yet the literature remains fragmented across separate discussions of generative AI, digital leadership, administrative decision support, human-machine collaboration, ethical AI, and agent architectures. Few accounts explain how these strands interact once autonomous agents become recurrent participants in institutional processes.

Three conceptual limitations are particularly evident. First, educational AI research continues to privilege pedagogy and individual use, leaving organizational leadership comparatively underexamined. Second, leadership scholarship often treats AI as a resource available to human decision-makers rather than as a system that can structure workflows and shape the informational conditions of decision-making. Third, AI governance generally conceptualizes artificial intelligence as an object to be regulated, paying less attention to hybrid arrangements in which humans and agents jointly produce organizational action.

The research gap therefore concerns the absence of an integrated framework linking agentic capability, educational leadership, distributed information processing, professional authority, and institutional accountability. Human-AI Governance is proposed to address this gap. The concept does not imply shared moral status between humans and machines. It describes an institutional configuration in which AI agents make consequential contributions to governance processes while human actors retain normative authority, legal responsibility, and the obligation to justify outcomes.

This synthesis supports four propositions that guide the subsequent analysis. Claude Agents are likely to influence educational leadership principally through the reorganization of information and workflow rather than through direct replacement of leaders. Their contribution to decision quality depends on complementarity with human judgment, not autonomy alone. Increasing agentic capacity raises the need for stronger rather than weaker institutional governance. Finally, legitimate adoption requires that efficiency, accuracy, transparency, professional autonomy, and stakeholder rights be evaluated together rather than as separate implementation concerns.

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

The resulting conceptual direction moves the discussion beyond whether educational leaders should use generative AI. The more consequential issue is how educational institutions should govern systems that can participate continuously in organizational reasoning and action. The Human-AI Governance Framework for Educational Leadership developed in the following sections responds to that question by specifying the relationships among institutional conditions, agentic capabilities, collaborative decision mechanisms, governance safeguards, and educational outcomes.

3. Methodology

3.1. Research Design

This study adopts a conceptual qualitative research design grounded in an integrative literature review and theory-building methodology. Rather than evaluating the technical performance of Claude Agents or examining a specific institutional implementation, the research seeks to explain an emerging organizational phenomenon for which empirical evidence remains limited but conceptual development has become increasingly necessary. This methodological orientation is appropriate because the transition from generative artificial intelligence toward agentic AI has outpaced the development of theoretical models capable of explaining its implications for educational leadership and institutional governance. Consequently, the primary objective is not hypothesis testing but conceptual integration, analytical interpretation, and framework development.

Integrative literature reviews are particularly suitable when knowledge is dispersed across multiple academic domains and existing theoretical perspectives require synthesis into a coherent explanatory model. Snyder (2019) argues that integrative reviews facilitate theory development by combining findings from diverse disciplines to generate new conceptual insights rather than merely summarizing previous research. Similarly, Torraco (2005) emphasizes that theory-building reviews are especially valuable in emerging research fields characterized by fragmented conceptualizations and rapidly evolving knowledge. The present study follows this tradition by integrating scholarship from educational leadership, organizational governance, artificial intelligence, socio-technical systems, organizational information processing, and human-AI collaboration.

The methodological choice also reflects the interdisciplinary nature of the research problem. Claude Agents cannot be adequately understood through educational technology literature alone because their organizational implications extend to leadership theory, governance studies, management science, and digital transformation research. An integrative conceptual approach therefore enables the examination of relationships that remain largely disconnected within existing academic discussions.

3.2. Literature Selection and Analytical Strategy

The analytical corpus consists primarily of the contemporary scholarly literature identified by the author as the intellectual foundation of this study. Particular attention was given to recent contributions addressing agentic artificial intelligence, Claude Agents, educational leadership, AI-supported decision-making, organizational governance, and human-AI

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

collaboration. These publications were interpreted alongside foundational theoretical works that explain organizational information processing, socio-technical systems, leadership, and decision-making within complex institutions.

The literature was examined through iterative thematic analysis rather than bibliometric aggregation. Individual publications were not treated as isolated contributions but as components of broader theoretical conversations. The analysis focused on identifying conceptual convergence, competing interpretations, and unresolved tensions concerning three interconnected dimensions: the evolution from generative AI toward autonomous AI agents, the transformation of educational leadership under conditions of increasing algorithmic participation, and the emergence of governance models capable of integrating human and artificial forms of organizational intelligence.

Throughout the analytical process, recurring concepts, theoretical assumptions, and explanatory mechanisms were compared across disciplines. This interpretive strategy made it possible to identify relationships that remain implicit within individual studies but become more visible when examined collectively. Rather than privileging technological perspectives, the analysis deliberately emphasized organizational and governance implications, reflecting the central research question guiding the study.

3.3. Conceptual Framework Development

The development of the proposed **Human-AI Governance Framework for Educational Leadership (HAGF)** followed an abductive reasoning process in which theoretical propositions emerged through continuous interaction between the reviewed literature and conceptual interpretation. Abductive approaches are particularly valuable when existing theories only partially explain an emerging phenomenon because they allow researchers to generate plausible explanatory models while remaining grounded in empirical and theoretical observations.

Framework construction proceeded through four analytical stages. Initially, the literature was examined to identify the principal organizational characteristics distinguishing agentic AI from conventional generative artificial intelligence. Attention then shifted toward educational leadership studies to determine how contemporary leadership theories conceptualize decision-making, organizational coordination, institutional learning, and professional responsibility. In the third stage, governance literature was analyzed to explore mechanisms of accountability, oversight, transparency, legitimacy, and ethical regulation applicable to autonomous computational systems. Finally, these complementary perspectives were synthesized into an integrated conceptual model explaining how Claude Agents may participate in educational governance while preserving meaningful human authority and institutional accountability.

The resulting framework conceptualizes educational leadership as a collaborative governance process rather than a purely human managerial function. Within this perspective, autonomous AI agents contribute to organizational intelligence through information processing, reasoning, workflow coordination, and analytical support, while educational leaders retain responsibility for contextual interpretation, ethical judgment, strategic direction, and institutional legitimacy. Governance therefore emerges from the interaction between complementary

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

human and artificial capabilities rather than from technological substitution or hierarchical control alone.

3.4. Research Quality and Methodological Limitations

The credibility of conceptual research depends primarily on analytical coherence, theoretical transparency, and the logical integration of existing knowledge rather than statistical generalization. Accordingly, every conceptual proposition advanced in this paper is explicitly derived from established theoretical perspectives and contemporary scholarship concerning educational leadership, artificial intelligence, governance, and organizational studies. The proposed framework is therefore intended as an analytical model that organizes current knowledge and stimulates further empirical investigation rather than as a definitive representation of organizational reality.

Several limitations should nevertheless be acknowledged. First, the study is conceptual and does not include empirical observations from schools or universities implementing Claude Agents or comparable autonomous AI systems. Consequently, the proposed relationships remain theoretically grounded rather than empirically validated. Second, the rapid evolution of agentic artificial intelligence implies that technological capabilities and institutional practices may develop more quickly than academic theory, requiring continuous refinement of conceptual models. Third, although the literature reviewed reflects multiple disciplinary perspectives, much of the scholarship remains concentrated in technologically advanced educational systems, potentially limiting the immediate transferability of certain governance assumptions to institutions operating within different regulatory, cultural, or organizational contexts.

These limitations do not diminish the relevance of the proposed framework. Instead, they highlight the exploratory character of the study and reinforce the need for subsequent empirical research examining how Human-AI Governance is enacted across different educational settings. The framework presented in this paper should therefore be understood as a theoretically grounded point of departure for future qualitative, quantitative, and mixed-method investigations into the organizational consequences of agentic artificial intelligence in educational leadership.

4. Results and Discussions

4.1. Claude Agents as Agentic Organizational Actors

The literature examined throughout this study suggests that Claude Agents represent a qualitative departure from conventional generative artificial intelligence rather than a simple technological improvement. Earlier generations of large language models were primarily designed to generate text, answer questions, summarize information, or assist users during isolated interactions. Although these capabilities significantly influenced teaching and learning practices, their organizational role remained largely supportive. Agentic AI introduces a different organizational logic by enabling autonomous planning, contextual memory, iterative reasoning, tool integration, workflow execution, and persistent collaboration across multiple institutional tasks. Consequently, the analytical focus shifts

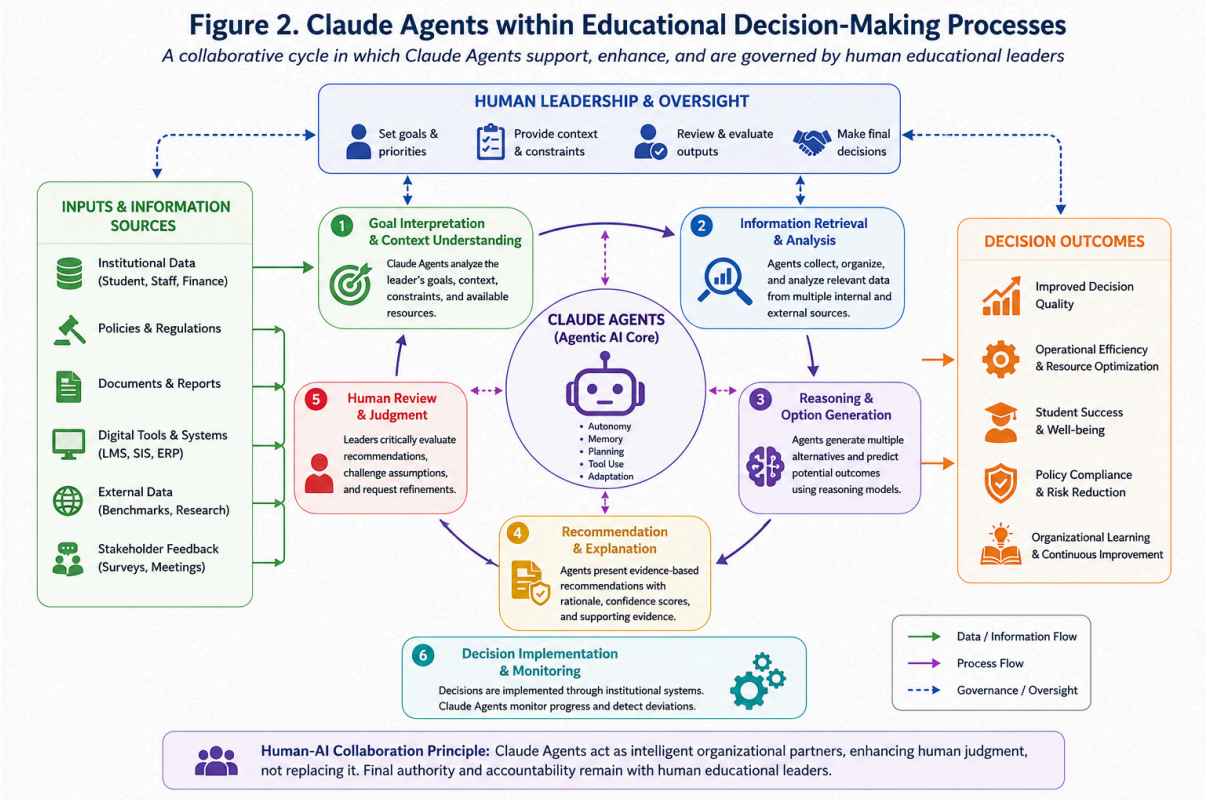
DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

from content generation to organizational participation.

This distinction has important implications for educational leadership. Educational organizations increasingly operate within environments characterized by regulatory complexity, expanding administrative requirements, fragmented information ecosystems, and growing demands for evidence-informed decision-making. Under these conditions, leadership involves continuous interpretation of heterogeneous information rather than isolated managerial decisions. Claude Agents substantially expand institutional information-processing capacity by retrieving organizational knowledge, synthesizing multiple sources, coordinating administrative workflows, generating strategic alternatives, and supporting policy implementation. Their contribution therefore extends beyond efficiency gains toward participation in organizational cognition itself.

To illustrate the operational role of Claude Agents within educational organizations, Figure 2 presents a conceptual representation of AI-supported decision-making processes. The figure demonstrates how agentic AI contributes to organizational decision-making by integrating institutional data, autonomous reasoning, workflow orchestration, and evidence-based recommendations under continuous human leadership and oversight. Rather than replacing educational leaders, Claude Agents function as collaborative organizational partners that enhance analytical capacity while preserving human authority and accountability throughout the decision-making process.

Figure 2. Claude Agents within Educational Decision-Making Processes



Source: Authors' conceptualization based on Russell and Norvig (2021), Jarrahi (2018), Papa and Jackson (2021), Dai et al. (2025), Appelo (2025), Presti et al. (2025), Richardson et al. (2025), Becker et al. (2026), and Woods (2026).

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

The findings resonate with organizational leadership research outside education. Becker et al. (2026) argue that conversational AI agents increasingly function as collaborative partners capable of supporting strategic reasoning rather than merely automating routine activities. Similarly, Woods (2026) suggests that future organizational effectiveness will depend on leaders' ability to integrate computational intelligence into managerial judgment without relinquishing strategic responsibility. Appelo (2025) further contends that AI-driven organizations require leadership models built around continuous collaboration between human expertise and autonomous computational agents. Collectively, these perspectives indicate that agentic AI should be understood as expanding institutional intelligence rather than replacing human leadership.

Within educational organizations, this transformation may be particularly significant because leadership responsibilities are simultaneously pedagogical, administrative, ethical, and political. School leaders routinely integrate educational objectives with regulatory compliance, stakeholder expectations, financial constraints, and community accountability. Claude Agents can support these interconnected activities by organizing institutional knowledge, monitoring policy implementation, coordinating documentation, identifying emerging organizational risks, and generating analytical reports. Nevertheless, these capabilities do not eliminate the necessity of human leadership. Rather, they redistribute cognitive work across socio-technical systems in which computational reasoning complements professional experience, contextual interpretation, and ethical deliberation.

This interpretation also challenges deterministic narratives portraying artificial intelligence either as a replacement for educational leaders or as a neutral technological instrument. The literature consistently suggests that organizational outcomes depend less on technological sophistication than on institutional arrangements governing interactions between human and artificial actors. Consequently, the emergence of Claude Agents should be interpreted as a governance issue rather than solely as a technological innovation.

4.2. Educational Leadership Beyond Digital Transformation

Educational leadership has historically evolved alongside broader organizational and technological transformations. Early discussions concerning digital leadership emphasized technological competence, innovation management, and the integration of digital infrastructure into teaching and administrative processes (Agustina et al., 2020). More recent scholarship has shifted attention toward artificial intelligence, recognizing that intelligent systems increasingly influence institutional communication, planning, evaluation, and resource allocation (Richardson et al., 2025). The emergence of agentic AI extends this trajectory by fundamentally altering the distribution of organizational intelligence.

The literature reviewed in this study indicates that educational leadership is gradually transitioning from technology-enabled administration toward collaborative governance involving autonomous computational actors. This transition does not imply diminishing human authority but rather increasing organizational interdependence between human judgment and artificial reasoning. Dai et al. (2025) describe this relationship as symbiotic collaboration, emphasizing complementary rather than substitutive interactions between educational leaders and AI-supported decision systems. Such collaboration becomes increasingly relevant as educational institutions confront rising informational complexity,

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

accelerating policy changes, expanding accountability mechanisms, and growing societal expectations regarding transparency and responsiveness.

From a socio-technical perspective, leadership effectiveness depends on balancing technological capability with organizational adaptability. The introduction of Claude Agents may improve administrative coordination, reduce routine cognitive workload, and facilitate evidence-informed planning. At the same time, autonomous reasoning systems create new organizational dependencies requiring continuous monitoring, validation, and institutional oversight. Leadership therefore expands beyond managing educational organizations toward governing complex socio-technical ecosystems composed of human professionals, intelligent computational agents, organizational policies, and digital infrastructures.

This transformation also modifies traditional understandings of distributed leadership. Previous models assumed that leadership responsibilities were distributed among teachers, administrators, and professional communities. Agentic AI introduces an additional dimension by incorporating autonomous computational actors into organizational decision processes. Such participation does not imply independent authority but rather continuous analytical collaboration that reshapes how educational institutions process information, generate strategic alternatives, and coordinate organizational action. Consequently, distributed leadership increasingly evolves toward distributed organizational intelligence.

4.3. The Emergence of Human-AI Governance

The central conceptual contribution emerging from the analysis is the proposition that educational institutions are entering a new governance paradigm characterized by structured collaboration between human leaders and autonomous AI agents. Existing governance frameworks generally assume that artificial intelligence remains an external object requiring ethical regulation, transparency mechanisms, and legal oversight. While these concerns remain essential, they become insufficient once AI agents begin participating directly in organizational reasoning, administrative coordination, and institutional decision support.

The concept of **Human-AI Governance** proposed in this study addresses this theoretical limitation by positioning governance as a collaborative institutional architecture through which authority, responsibility, and organizational intelligence are distributed between human actors and autonomous computational systems under explicit normative constraints. Governance is therefore conceptualized not as algorithmic control but as the institutional design regulating interactions between complementary forms of intelligence.

Within this framework, Claude Agents contribute to organizational processes through several interconnected functions. They expand institutional information-processing capacity by synthesizing complex data, retrieving organizational knowledge, coordinating administrative workflows, supporting strategic analysis, and generating evidence-informed recommendations. Human educational leaders simultaneously retain responsibility for contextual interpretation, ethical reasoning, value-based prioritization, stakeholder negotiation, and final institutional accountability. Decision quality consequently emerges from interaction rather than substitution.

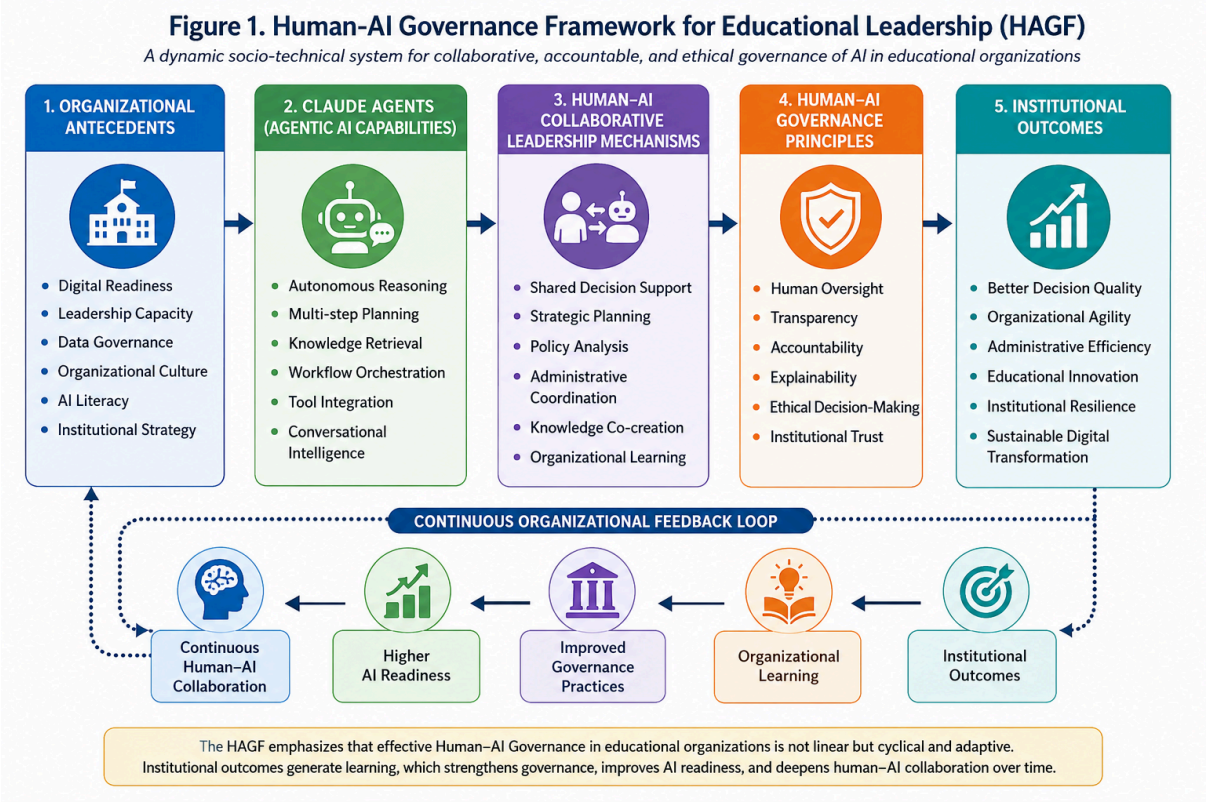
This interpretation aligns with Organizational Information Processing Theory, which argues that organizations respond to increasing environmental uncertainty by expanding their capacity to acquire, process, and interpret information. Claude Agents significantly increase

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

this capacity while simultaneously generating governance requirements associated with transparency, explainability, accountability, and institutional trust. Similarly, socio-technical systems theory suggests that organizational performance depends upon the reciprocal adaptation of technological and social subsystems. Human-AI Governance operationalizes this principle by conceptualizing educational institutions as adaptive socio-technical organizations in which governance structures coordinate interactions between human expertise and computational intelligence.

To integrate the conceptual arguments developed throughout this study, Figure 1 presents the proposed Human-AI Governance Framework for Educational Leadership (HAGF). The framework conceptualizes Human-AI Governance as a dynamic socio-technical system in which organizational readiness enables the deployment of agentic AI capabilities that, through collaborative leadership mechanisms and governance principles, generate institutional outcomes while continuously informing organizational learning and future governance practices.

Figure 1. Human-AI Governance Framework for Educational Leadership (HAGF)



Source: Authors' conceptualization based on Galbraith (1974), Trist and Bamforth (1951), Simon (1977), Jarrahi (2018), Floridi and Cowls (2019), Papa and Jackson (2021), Dai et al. (2025), Appelo (2025), Presti et al. (2025), Becker et al. (2026), Sollazzo et al. (2026), and Woods (2026).

Figure 1. Human-AI Governance Framework for Educational Leadership (HAGF). The framework illustrates the dynamic relationships among organizational antecedents, Claude Agents' agentic capabilities, collaborative leadership mechanisms, governance principles, and institutional outcomes. Continuous feedback loops emphasize that Human-AI Governance is

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

an adaptive organizational process in which institutional learning continuously improves AI readiness, governance practices, and collaborative decision-making.

4.4. Theoretical Implications

The conceptual framework developed in this study contributes to several emerging debates within educational leadership and artificial intelligence research.

First, it extends educational leadership theory beyond discussions of digital leadership and technology integration by introducing governance as the primary analytical category through which agentic AI should be understood. Existing models generally examine how leaders implement technological innovations. The present analysis instead considers how autonomous AI agents become participants within institutional governance itself, thereby expanding prevailing conceptualizations of leadership.

Second, the framework refines contemporary discussions concerning human-AI collaboration. Much of the existing literature conceptualizes collaboration as occurring between individual users and AI systems. The present study relocates collaboration to the organizational level, emphasizing institutional coordination, governance architectures, and distributed organizational intelligence rather than isolated human-machine interactions.

Third, the proposed framework contributes to organizational governance research by introducing educational institutions as distinctive socio-technical environments where legitimacy, ethics, democratic accountability, and professional autonomy remain inseparable from technological innovation. Governance consequently becomes a multidimensional construct integrating organizational effectiveness with institutional responsibility.

Finally, the analysis contributes to the rapidly expanding literature on agentic artificial intelligence by proposing that Claude Agents should be conceptualized as **agentic organizational actors** rather than advanced digital assistants. This conceptual distinction may prove valuable for future research examining how autonomous AI systems reshape organizational structures across education, public administration, healthcare, and other knowledge-intensive sectors.

Overall, the findings suggest that the most significant consequence of agentic AI is unlikely to be the automation of educational work itself. Rather, its enduring impact may lie in transforming how educational institutions organize knowledge, distribute authority, coordinate decision-making, and construct governance systems capable of integrating human judgment with increasingly autonomous computational intelligence. This perspective positions Human-AI Governance as an emerging theoretical paradigm for understanding educational leadership in the era of agentic artificial intelligence.

5. Future Research Agenda

The emergence of agentic artificial intelligence represents an inflection point in the evolution of educational leadership research. Whereas earlier investigations predominantly examined artificial intelligence as a pedagogical resource or an administrative support technology, the increasing organizational capabilities of autonomous AI agents require a broader research agenda capable of explaining how leadership, governance, and institutional decision-making evolve within hybrid human-AI environments. The conceptual framework proposed in this

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

study provides an initial theoretical foundation, yet its explanatory value ultimately depends on systematic empirical validation across diverse educational contexts. Consequently, future research should move beyond technology acceptance studies and investigate the organizational consequences of integrating autonomous intelligent agents into leadership practices and governance structures.

One important research direction concerns the empirical validation of the Human-AI Governance Framework for Educational Leadership (HAGF). The present study conceptualizes governance as a collaborative institutional process through which human leaders and autonomous AI agents jointly contribute to organizational intelligence while maintaining distinct responsibilities. Future qualitative investigations could examine how school leaders negotiate these responsibilities in practice, particularly when AI-generated recommendations influence strategic planning, policy implementation, resource allocation, or crisis management. Comparative case studies involving schools and universities at different stages of AI adoption would provide valuable insights into how governance arrangements emerge under varying organizational, cultural, and regulatory conditions. Such studies could also identify contextual factors that facilitate or constrain effective Human-AI Governance, including institutional trust, digital maturity, leadership experience, and organizational culture.

A second avenue for research concerns the redistribution of organizational authority within AI-supported educational institutions. Existing leadership theories continue to assume that authority remains concentrated within human organizational actors, whereas agentic AI introduces systems capable of influencing decision processes through information synthesis, workflow coordination, and strategic recommendation. Future research should therefore examine how authority is negotiated when autonomous agents become embedded within institutional routines. Particular attention should be devoted to understanding whether educational leaders perceive AI-generated recommendations as advisory inputs or whether repeated reliance gradually transforms professional judgment into algorithmically mediated decision-making. Longitudinal qualitative studies would be especially valuable for observing how these relationships evolve over time rather than capturing only initial implementation stages.

Another promising area involves examining organizational trust within Human-AI Governance systems. Trust has traditionally been conceptualized as an interpersonal phenomenon linking leaders, teachers, students, and educational communities. The introduction of autonomous AI agents raises the possibility of institutional trust extending toward socio-technical systems rather than exclusively human actors. Future research should investigate the conditions under which educational professionals trust AI-supported recommendations, how transparency influences perceived legitimacy, and whether explainability enhances or diminishes confidence in organizational decision-making. Mixed-method research combining surveys, interviews, and behavioral observations could contribute to a more nuanced understanding of trust dynamics within AI-enabled educational organizations.

The ethical dimensions of Human-AI Governance also warrant extensive investigation. While contemporary discussions frequently address algorithmic bias, privacy protection, and transparency, relatively little is known about how these ethical principles operate within routine organizational decision-making. Future studies should explore how educational

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

institutions establish accountability mechanisms for AI-assisted decisions, how governance policies define responsibility when AI-generated recommendations contribute to organizational outcomes, and how ethical oversight evolves as agentic systems acquire greater operational autonomy. Comparative policy analyses across different national educational systems would be particularly valuable because regulatory approaches toward artificial intelligence continue to diverge substantially between jurisdictions.

Future scholarship should also examine the changing competency profile of educational leaders operating within Human-AI Governance environments. Leadership preparation programs have historically emphasized instructional leadership, organizational management, financial administration, and policy implementation. The emergence of autonomous AI agents suggests that future educational leaders may additionally require competencies in algorithmic literacy, AI governance, socio-technical systems thinking, ethical decision-making, digital risk management, and collaborative intelligence. Empirical studies involving leadership development programs could investigate how these competencies are acquired, assessed, and integrated into professional practice. Such research would contribute to redesigning educational leadership preparation for increasingly AI-enabled organizational environments.

Another underexplored area concerns the organizational consequences of multi-agent systems within educational institutions. While current research primarily examines interactions between individual users and single AI systems, the next generation of educational organizations may increasingly employ multiple specialized agents performing complementary administrative, analytical, and instructional functions. Future research should therefore investigate how networks of autonomous agents influence organizational coordination, communication patterns, knowledge management, and institutional resilience. Organizational network analysis may provide particularly useful methodological approaches for examining how information flows change when AI agents become integrated into institutional communication structures.

Cross-cultural comparative research represents another significant opportunity. Educational governance reflects distinct national traditions, institutional cultures, legal frameworks, and public expectations. Consequently, Human-AI Governance is unlikely to develop uniformly across educational systems. Comparative investigations involving educational institutions from different governance traditions could examine how variations in regulation, institutional autonomy, technological infrastructure, and cultural attitudes toward artificial intelligence shape leadership practices and governance arrangements. Such studies would strengthen the external validity of emerging conceptual frameworks while identifying context-specific adaptations necessary for responsible AI implementation.

Methodologically, the field would benefit from greater diversity. Qualitative case studies should be complemented by longitudinal surveys, experimental simulations, organizational ethnographies, social network analyses, and mixed-method designs capable of capturing both organizational processes and leadership outcomes. As Claude Agents and other autonomous AI systems become more widely deployed, access to naturally occurring organizational settings will allow researchers to move beyond hypothetical scenarios toward evidence-based analyses of Human-AI Governance in practice.

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

Finally, future theoretical work should continue integrating educational leadership with organizational governance, artificial intelligence, complexity science, socio-technical systems theory, and organizational information processing theory. The rapid evolution of agentic AI suggests that no single theoretical perspective is likely to explain the full complexity of emerging educational organizations. Greater theoretical integration will therefore be necessary to understand how autonomous computational agents influence leadership, organizational learning, institutional legitimacy, and governance under conditions of increasing technological complexity.

Taken together, these research directions indicate that the future of educational leadership scholarship will depend less on evaluating isolated AI technologies than on understanding how educational institutions reorganize authority, accountability, and organizational intelligence in response to increasingly autonomous computational actors. Human-AI Governance should therefore be regarded not as a completed theoretical construct but as an evolving research program that invites interdisciplinary collaboration among scholars of educational leadership, organizational studies, artificial intelligence, public administration, ethics, and digital governance. Empirical examination of this emerging governance paradigm will determine whether agentic AI becomes primarily an instrument of administrative efficiency or a catalyst for more adaptive, transparent, resilient, and human-centered educational organizations.

6. Practical Implications

The transition from generative artificial intelligence to agentic AI has implications that extend well beyond technological adoption. Educational institutions are entering an era in which intelligent agents can contribute not only to information retrieval and content generation but also to organizational coordination, policy implementation, workflow management, and strategic decision support. Consequently, educational leaders should reconsider AI implementation as a governance transformation rather than an isolated digital innovation initiative. The Human-AI Governance Framework proposed in this study suggests that successful implementation depends on establishing organizational structures capable of integrating computational intelligence while preserving human accountability, ethical responsibility, and institutional legitimacy.

For school principals and university administrators, the emergence of Claude Agents requires the development of new leadership competencies that complement traditional managerial expertise. Beyond understanding the technical capabilities of AI systems, educational leaders must acquire the ability to evaluate AI-generated recommendations critically, determine appropriate levels of delegated autonomy, recognize algorithmic limitations, and maintain responsibility for organizational decisions. These competencies include AI literacy, data governance, ethical reasoning, risk management, and strategic oversight. Rather than replacing leadership expertise, agentic AI increases the importance of professional judgment because leaders must evaluate not only organizational information but also the processes through which intelligent agents generate recommendations.

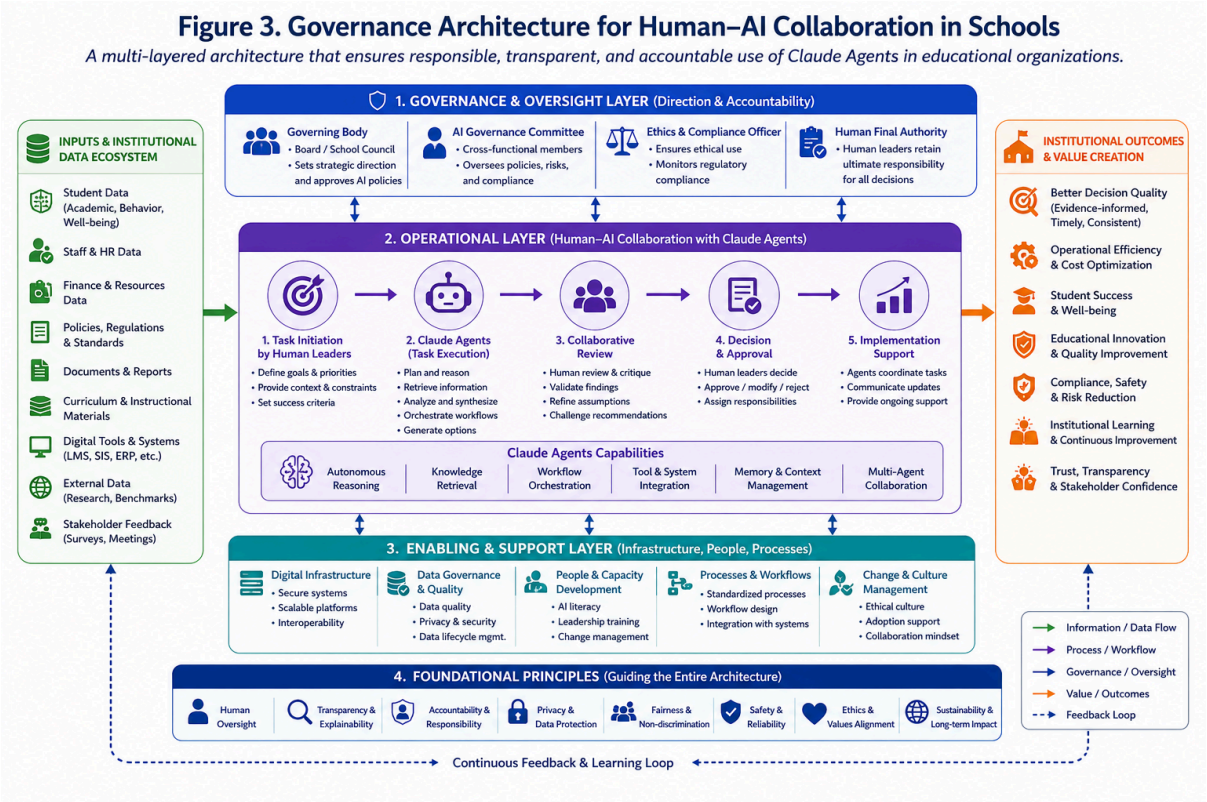
Educational institutions should also establish formal governance mechanisms before deploying autonomous AI agents in administrative decision-making. Such mechanisms

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

should define institutional policies regarding data access, authorization levels, transparency requirements, audit procedures, human oversight, and accountability. Decisions involving student welfare, personnel management, financial resource allocation, or institutional strategy should remain subject to explicit human review regardless of the sophistication of AI-generated analyses. Governance structures should therefore distinguish between operational automation and strategic authority, ensuring that computational efficiency does not undermine democratic accountability or professional responsibility.

Building upon the Human-AI Governance Framework developed in this study, Figure 3 proposes a practical governance architecture for the responsible implementation of Claude Agents within educational institutions. The architecture translates the conceptual principles of Human-AI Governance into an operational model by identifying the governance structures, organizational processes, enabling conditions, and institutional safeguards required to ensure transparent, accountable, and ethical collaboration between educational leaders and autonomous AI agents. The model emphasizes that effective Human-AI collaboration depends not only on technological capabilities but also on clearly defined governance mechanisms, human oversight, and continuous organizational learning.

Fig 3. **Governance Architecture for Human-AI Collaboration in Schools**



Source: Authors' conceptualization based on Galbraith (1974), Trist and Bamforth (1951), Floridi and Cowls (2019), Papa and Jackson (2021), Dai et al. (2025), Sollazzo et al. (2026), Walther (2024), and Appelo (2025).

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

The implementation of Claude Agents also necessitates organizational investment in digital infrastructure and institutional readiness. Agentic systems depend on high-quality organizational data, interoperable information systems, cybersecurity protections, and clearly defined operational procedures. Institutions lacking reliable digital ecosystems may experience fragmented AI implementation that increases rather than reduces organizational complexity. Consequently, digital transformation strategies should prioritize data governance, system integration, information quality, and organizational interoperability before expanding the operational responsibilities assigned to autonomous AI agents.

Professional development constitutes another critical implication. Teachers, department heads, academic administrators, and senior leadership teams should receive continuous training not only in the technical operation of AI systems but also in the ethical, legal, and organizational consequences of Human-AI Governance. Training programs should emphasize collaborative intelligence, encouraging educational professionals to treat AI as an analytical partner rather than an unquestionable decision-maker. Developing critical AI literacy across educational organizations will help prevent automation bias while strengthening institutional capacity to evaluate AI-supported recommendations responsibly.

Policymakers likewise have an essential role in facilitating responsible adoption. Existing educational policies frequently address digital technologies in general terms but provide limited guidance regarding autonomous AI agents capable of participating in organizational workflows. National and regional education authorities should therefore develop regulatory frameworks that establish minimum standards for transparency, explainability, accountability, privacy protection, algorithmic auditing, and ethical oversight. Such policies should remain sufficiently flexible to accommodate rapid technological innovation while ensuring that educational institutions continue to operate according to principles of equity, fairness, and public accountability.

Technology developers also share responsibility for advancing Human-AI Governance. The design of Claude Agents and similar systems should prioritize explainability, traceability, configurable levels of autonomy, secure data management, and mechanisms that enable meaningful human intervention. Systems intended for educational leadership should facilitate rather than obscure organizational accountability by providing transparent documentation of reasoning processes, evidence sources, confidence levels, and decision pathways. Designing AI systems that support institutional governance rather than merely maximizing automation will increase trust among educational professionals and improve the sustainability of AI adoption.

Finally, the practical value of Human-AI Governance lies in its ability to balance innovation with institutional responsibility. Educational organizations face increasing pressure to improve efficiency, responsiveness, and evidence-based decision-making while maintaining ethical standards, professional autonomy, and stakeholder trust. Claude Agents offer considerable opportunities to enhance organizational intelligence, yet these opportunities can only be realized when AI implementation is guided by robust governance principles rather than technological enthusiasm alone. The future effectiveness of educational leadership will therefore depend not on the extent to which AI replaces human decision-making but on the capacity of institutions to develop collaborative governance architectures that combine computational capabilities with human wisdom, ethical judgment, and democratic accountability.

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

7. Conclusions

The transition from generative artificial intelligence to agentic AI represents a fundamental organizational development rather than a simple technological progression. While previous generations of large language models primarily supported knowledge generation and instructional activities, Claude Agents introduce capabilities that extend into institutional coordination, strategic reasoning, workflow orchestration, and organizational decision support. This shift requires educational leadership research to move beyond discussions of technology adoption and toward a more comprehensive understanding of governance in increasingly intelligent educational organizations. The central argument advanced throughout this paper is that the most significant consequence of agentic AI lies not in the automation of educational tasks but in the gradual reconfiguration of how educational institutions organize knowledge, distribute authority, and make decisions.

By synthesizing contemporary scholarship from educational leadership, organizational governance, socio-technical systems, organizational information processing, and human-AI collaboration, this study has proposed the **Human-AI Governance Framework for Educational Leadership (HAGF)** as a conceptual model for interpreting these emerging transformations. The framework conceptualizes governance as a collaborative institutional process in which human leaders and autonomous AI agents contribute complementary forms of intelligence within clearly defined ethical, organizational, and regulatory boundaries. Rather than positioning artificial intelligence as either a substitute for educational leaders or a passive technological instrument, the framework presents Claude Agents as **agentic organizational actors** capable of expanding institutional information-processing capacity while remaining subject to human accountability and normative oversight.

From a theoretical perspective, the paper contributes to several ongoing academic debates. First, it extends educational leadership scholarship by introducing Human-AI Governance as an analytical construct that explains how leadership evolves when autonomous computational agents become integrated into institutional decision-making processes. Existing research has largely concentrated on digital leadership, AI-supported teaching, or administrative automation. The present study instead argues that governance provides a more appropriate conceptual lens for understanding organizational transformations associated with agentic AI. Second, the paper contributes to the emerging literature on human-AI collaboration by relocating collaboration from the individual user level to the organizational level, emphasizing institutional coordination, distributed organizational intelligence, and governance architectures rather than isolated interactions between individuals and intelligent systems. Third, the proposed framework broadens discussions concerning agentic artificial intelligence by conceptualizing Claude Agents as organizational participants whose significance derives from their role within institutional processes rather than from their computational performance alone.

The study also carries practical implications for educational leaders, school administrators, and institutional policymakers. As agentic AI becomes progressively integrated into educational organizations, leadership will increasingly require competencies extending beyond digital literacy or technology management. Educational leaders will need to develop governance capabilities that encompass ethical oversight, algorithmic transparency, institutional accountability, data stewardship, and organizational trust. At the same time,

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

schools and universities should establish governance structures that clearly define the scope of AI-supported decision-making, delineate responsibilities between human actors and intelligent agents, and ensure that institutional decisions remain aligned with educational values, legal requirements, and public expectations. The implementation of Claude Agents should therefore be accompanied by governance mechanisms capable of balancing innovation with professional autonomy, transparency, and democratic accountability.

The conceptual nature of this study inevitably imposes certain limitations. The proposed framework has been derived through theoretical synthesis rather than empirical observation and therefore requires systematic validation across diverse educational settings. Moreover, the rapid evolution of agentic AI suggests that technological capabilities, organizational practices, and governance arrangements will continue to evolve, potentially requiring future refinement of the concepts presented here. Current educational institutions also differ substantially with respect to digital maturity, regulatory environments, organizational culture, and technological infrastructure, implying that Human-AI Governance may assume different forms across national and institutional contexts. Consequently, the framework should be interpreted as an evolving conceptual model rather than a definitive representation of organizational reality.

These limitations also point toward promising avenues for future research. Empirical studies are needed to investigate how school leaders currently perceive and employ autonomous AI agents in administrative and strategic activities, how governance responsibilities are redistributed within AI-supported organizations, and how teachers, administrators, students, and policymakers negotiate questions of accountability and institutional trust. Comparative investigations across educational systems could examine how different governance traditions shape the adoption of agentic AI, while longitudinal research may explore how leadership practices evolve as autonomous agents become embedded within routine organizational processes. Mixed-methods designs combining organizational case studies, surveys, interviews, and network analysis could further illuminate the dynamics of Human-AI Governance and provide empirical evidence regarding its effectiveness under varying institutional conditions.

Ultimately, educational leadership appears to be entering a period in which intelligence is no longer understood exclusively as a human organizational resource but increasingly as a collaborative capability emerging from interactions between human expertise and autonomous computational agents. Understanding this transition requires conceptual frameworks that acknowledge both the opportunities created by agentic AI and the institutional responsibilities accompanying its adoption. Human-AI Governance provides one possible foundation for this emerging research agenda by positioning governance, rather than technology itself, at the center of discussions concerning the future of educational leadership. As Claude Agents and related autonomous systems become progressively integrated into educational organizations, the enduring challenge will not be determining whether artificial intelligence should participate in institutional decision-making, but how educational institutions can design governance systems that ensure such participation strengthens organizational resilience, preserves professional judgment, and remains consistent with the ethical and public purposes of education.

Acknowledgment

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

The authors acknowledge the use of OpenAI's ChatGPT (GPT-5.5) to support language editing, manuscript organization, and the preparation of conceptual figures. All scientific content, theoretical development, interpretation of the literature, and conclusions are the sole responsibility of the authors, who reviewed and verified the final manuscript before submission.

References

Aboumoussa, L., & Pfister, J. (2024). *Leadership development in the age of artificial intelligence* (Doctoral dissertation, Harvard University).

Agustina, R., Kamdi, W., Hadi, S., & Nurhadi, D. (2020). Influence of the principal's digital leadership on the reflective practices of vocational teachers mediated by trust, self efficacy, and work engagement. *International Journal of Learning, Teaching and Educational Research*, *19*(11), 24-40..

Appelo, J. (2025). *Human robot agent: New fundamentals for AI-driven leadership with algorithmic management*. Jojo Ventures.

Becker, M., de Jong, D., Briker, R., Mennens, K., Heller, J., Mahr, D., & Grewal, D. (2026). Using customized, conversational AI agents in leadership and management research: Benefits, practical illustrations, and best practices. *The Leadership Quarterly*, *37*(3), 101952.

Brynjolfsson, E., & McAfee, A. (2017). *Machine, platform, crowd: Harnessing our digital future*. W. W. Norton.

Cheong, P. H. (2025). Generative artificial intelligence and collaboration: Exploring religious human-machine communication and tensions in leadership practices. *Human-Machine Communication, 10*, 171-190.

Dai, R., Thomas, M. K. E., & Rawolle, S. (2025). The roles of AI and educational leaders in AI-assisted administrative decision-making: A proposed framework for symbiotic collaboration. *The Australian Educational Researcher, 52*(2), 1471-1487.

Doumanas, D., Soularidis, A., & Kotis, K. (2026). Causal reasoning and large language models for military decision-making: Rethinking the command structures in the era of generative AI. *AI, 7*(1), 14.

Drożdżyńska, S., & Kucharski, P. (2025). Assessing LLMs through scenario-based soft skills tasks for managers: Exploratory studies. In *Intelligent Management and Artificial Intelligence: Trends, Challenges, and Opportunities* (Vol. 2, pp. 146-161).

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1).

Galbraith, J. R. (1974). Organization design: An information processing view. *Interfaces, 4*(3), 28-36.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons, 61*(4), 577-586.

Kanbakan, A., Berikol, G. B., İlhan, B., Altıntaş, E., & Doğanay, F. (2026). Artificial intelligence based resuscitation simulation: A pilot study of a novel approach to team leadership training. *BMC Medical Education*.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*(1), 71-87.

DOI: https://doi.org/10.65222/VIRAL.2026.7.43.63

Mintzberg, H. (1973). *The nature of managerial work*. Harper & Row.

Papa, R., & Jackson, K. M. (Eds.). (2021). *Artificial intelligence, human agency and the educational leader*. Springer.

Presti, F. L., Gentile, M., Mangione, G. R. J., & Zampolini, L. (2025). Reimagining AI agents for education: The evolution towards collaborative large language model multi-agent systems. In *ICERI2025 Proceedings* (pp. 6936-6944). IATED.

Richardson, J. W., Vedder, B. C., Roberts, A. B., & McLeod, S. (2025). What's the chatter about AI and school leaders? *Leadership and Policy in Schools, 24*(1), 103-116.

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Simon, H. A. (1977). *The new science of management decision* (Rev. ed.). Prentice Hall.

Sollazzo, K., Titah, R., Léger, P. M., & Karran, A. J. (2026). Agentic AI and IT governance: Balancing autonomy and oversight. In *Proceedings of the International Conference on Human-Computer Interaction* (pp. 431-438). Springer.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333-339.

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review, 4*(3), 356-367.

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

Walther, C. C. (2024). *Human leadership for humane technology*. Palgrave Macmillan.

Weick, K. E. (1979). *The social psychology of organizing* (2nd ed.). Addison-Wesley.

Woods, G. (2026). *The AI-driven leader: Harnessing AI to make faster, smarter decisions*. Simon & Schuster.