Policy Journal of Social Science Review

Online ISSN 3006-4635

Print ISSN 3006-4627

Vol. 3 No. 12 (2025)

Who Leads When AI Acts? Authority and Accountability in Agentic Organizations

[1]**Muhammad Ajmal**

[*2]**Azmat Islam**

[1]*Department of Management Science, University of Gujrat, Gujrat, Pakistan*

[*2] *Department of Business Administration, University of Education, Lahore, Pakistan.*

[1]ajmal.hailian@gmail.com, [*2] azmat24@gmail.com

Abstract

As artificial intelligence systems evolve from passive tools into autonomous agents capable of initiating actions, coordinating tasks, and making consequential decisions, traditional models of organizational authority are being fundamentally reshaped. This article examines how leadership, authority, and accountability function in "agentic organizations" where AI systems act with delegated discretion. We argue that authority in such settings becomes hybrid and distributed, emerging from interactions between human actors, algorithmic systems, and institutional governance structures rather than residing in a single identifiable leader. Drawing on organizational theory, governance scholarship, and AI ethics frameworks, we develop a conceptual model distinguishing operational authority (who executes decisions), epistemic authority (whose judgments are trusted), and moral-legal accountability (who bears responsibility). Through illustrative cases across corporate, public sector, and platform-based environments, we show how agentic AI complicates chains of command, diffuses responsibility, and challenges existing liability regimes. The article proposes governance design principles—including transparent delegation architectures, traceable decision pathways, and layered accountability mechanisms—to clarify leadership roles and preserve human oversight without undermining the performance advantages of AI autonomy. By reconceptualizing authority as socio-technical and dynamic, this work contributes to emerging scholarship on AI governance and offers practical guidance for organizations navigating the transition to agentic systems.

Keywords: Agentic AI; organizational authority; accountability; algorithmic governance; distributed leadership; AI ethics; socio-technical systems; human–AI collaboration; responsibility attribution; corporate governance.

Article Details:

Received on 18 Nov, 2025

Accepted on 10 Dec, 2025

Published on 12 Dec 2025

Corresponding Authors*

Azmat Islam

Policy Journal of Social Science Review

Online ISSN 3006-4635

Print ISSN 3006-4627

Vol. 3 No. 12 (2025)

1. Introduction

Artificial intelligence (AI) is rapidly evolving from a decision-support tool into an increasingly autonomous organizational actor. Recent advances in large language models, reinforcement learning, and multi-agent systems have enabled AI systems not only to analyze information but also to initiate actions, coordinate workflows, and make decisions with limited or no real-time human intervention(Ajmal, 2022). These developments mark a transition from *assistive AI* to what may be described as *agentic AI*—systems capable of goal-directed behavior within organizational environments(Ajmal, Islam, & Khan, 2023). As AI systems assume more active roles, a fundamental question emerges: **who leads when AI acts?** More specifically, how should authority and accountability be conceptualized in organizations where algorithmic agents exercise operational discretion?

Existing scholarship on AI in organizations has largely focused on performance, augmentation, and task substitution. Studies show that AI can enhance productivity, improve decision accuracy, and generate new forms of value creation (Brynjolfsson & McAfee, 2017; Raisch & Krakowski, 2021). Yet as AI systems become embedded in decision-making processes, they begin to shape organizational routines, influence resource allocation, and affect stakeholders in consequential ways. In such contexts, AI no longer functions merely as a tool but participates in governance structures (Ahmed, Ajmal, & Haq, 2024). This shift challenges deeply rooted assumptions in organizational theory that authority ultimately resides in human actors.

Classical theories of authority—most prominently Weber's (1978) typology of traditional, charismatic, and legal-rational authority—presume identifiable human agents who exercise legitimate power(Ajmal, Islam, & Khan, 2024). Organizational design theories similarly conceptualize authority as flowing hierarchically through formal roles and structures (Mintzberg, 1980). Accountability, in turn, has been traditionally linked to answerability and sanctionability of identifiable decision-makers (Bovens, 2007). These frameworks depend on traceable chains of command and clearly attributable agency.

However, algorithmic systems complicate these assumptions. Scholars have noted that algorithmic governance redistributes decision-making power in ways that can obscure responsibility and diffuse control (Danaher et al., 2017; Yeung, 2018). Algorithms embedded in platforms, financial systems, and public administration increasingly make or shape decisions that were previously human-led, from content moderation to credit scoring and predictive policing (Kellogg et al., 2020; Pasquale, 2015). As a result, authority becomes partially embedded in technical architectures rather than solely in managerial roles(Ajmal, Khan, & Islam, 2024).

The growing literature on "algorithmic management" demonstrates how AI systems can supervise workers, allocate tasks, and enforce performance standards—often with minimal human oversight (Kellogg et al., 2020). In such contexts, employees may experience AI systems as de facto managers. Yet these systems lack moral agency and legal personhood, raising pressing questions about accountability(Ajmal, Manzoor, & Khan, 2024). When an AI-driven decision produces harm—whether discriminatory lending outcomes, unsafe automated driving behavior, or biased hiring recommendations—responsibility is often distributed across designers, deployers, data providers, and organizational leaders (Mittelstadt et al., 2016).

This diffusion of responsibility creates what some scholars describe as the "problem of many hands," where multiple actors contribute to outcomes but no single actor appears fully accountable (Bovens, 2007). At the same time, regulatory initiatives such as the European Union's AI Act and global AI governance frameworks emphasize the necessity of "human

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

oversight" and accountability mechanisms (Floridi et al., 2018). Yet the precise nature of oversight remains under-theorized in contexts where AI systems operate semi-autonomously and adapt dynamically.

Parallel developments in human–AI collaboration research suggest that AI does not simply replace human judgment but interacts with it in complex ways. Research indicates that AI can both improve and distort human decision-making depending on trust calibration and system transparency (Dietvorst et al., 2015; Logg et al., 2019). Overreliance on algorithmic outputs may produce automation bias, while under-reliance may reduce performance gains. These dynamics underscore that authority in AI-enabled organizations is relational and negotiated rather than static.

Despite these growing insights, there remains a conceptual gap in organizational scholarship: the lack of a coherent framework for understanding authority and accountability in what we term **agentic organizations**—organizations in which AI systems possess delegated discretion to initiate and execute consequential actions(Ajmal, Rahat, & Islam, 2024). Most existing discussions focus either on technical robustness or ethical compliance but stop short of reconceptualizing leadership structures themselves.

This article addresses that gap by developing a theoretical account of authority and accountability in agentic organizations(Zulfiqar, Ajmal, & Islam, 2024). We argue that authority becomes hybrid and distributed across human and algorithmic actors, mediated by governance architectures and socio-technical infrastructures. We distinguish among three analytically separable but interrelated dimensions of authority:

1. **Operational authority** – who executes and implements decisions;
2. **Epistemic authority** – whose judgments are trusted and relied upon;
3. **Moral-legal accountability** – who bears responsibility when outcomes occur.

By separating these dimensions, we clarify how AI systems may exercise operational and epistemic authority without holding moral-legal accountability, thereby creating structural tensions within organizations.

Theoretically, this article contributes to organizational theory by extending classical authority frameworks into socio-technical domains. It builds on research in algorithmic governance, AI ethics, and distributed agency to propose a model of layered accountability suited to agentic contexts(Ajmal, Islam, & Khalid, 2025b). Practically, it offers governance design principles—including traceable delegation pathways, transparent decision logging, and clearly assigned oversight roles—that can help organizations maintain legitimacy and responsibility while leveraging AI autonomy.

As AI systems increasingly act within organizational environments, leadership can no longer be understood solely as a human-centered phenomenon. Instead, authority must be reconceptualized as emerging from interactions among human actors, algorithmic systems, and institutional rules. The central challenge for contemporary organizations is not whether AI will participate in decision-making, but how authority and accountability will be structured when it does.

2. Literature Review

2.1. Classical Foundations of Authority and Accountability

The concept of authority in organizations has long been grounded in sociological and managerial theory. Max Weber's theory of legal-rational authority conceptualized authority as legitimate power embedded in formal roles and rule-based systems (Weber, 1978). In modern organizations, this authority is institutionalized through hierarchical structures and bureaucratic governance. Mintzberg (1980) further elaborated how coordination and control

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

mechanisms distribute authority across organizational forms, emphasizing strategic apex control and formalized workflows.

Accountability scholarship complements these theories by emphasizing answerability, transparency, and sanctionability. Bovens (2007) defines accountability as a relationship in which an actor must explain and justify conduct to a forum that can question and sanction. Traditional models assume clearly identifiable human agents who can be held responsible. These frameworks presuppose traceable decision chains and coherent leadership structures(Ajmal, Islam, & Islam, 2025).

However, digital technologies have begun to strain these assumptions. As authority becomes partially embedded in technical systems rather than solely in managerial positions, classical hierarchical models face conceptual limitations. This transition sets the stage for examining how AI systems alter authority configurations(Ajmal, Khalid, & Islam, 2025a).

2.2. Algorithmic Governance and Distributed Decision-Making

The rise of algorithmic governance has transformed decision-making in both public and private sectors. Algorithmic systems increasingly mediate allocation decisions, risk assessments, and behavioral regulation (Danaher et al., 2017; Yeung, 2018). Yeung (2018) argues that algorithmic regulation differs from traditional rule-based governance because it operates through continuous data-driven optimization rather than static legal norms.

Danaher et al. (2017) highlight how algorithmic governance shifts power toward technical infrastructures and data controllers, often reducing transparency. Similarly, Pasquale (2015) describes "black box" systems in finance and digital platforms that obscure the basis of automated decisions, complicating accountability.

Research in public administration further notes that algorithmic systems can reshape institutional responsibility. When automated tools influence welfare distribution, predictive policing, or immigration decisions, lines of accountability become blurred between policymakers, system designers, and frontline bureaucrats (Yeung, 2018). This phenomenon contributes to what Bovens (2007) calls the "problem of many hands," where responsibility is diffused across multiple actors.

Thus, algorithmic governance literature establishes that authority can be embedded in socio-technical systems, challenging human-centered leadership models.

2.3. Algorithmic Management and AI as Organizational Actor

A growing body of research examines AI systems as active agents within organizational control structures. Kellogg, Valentine, and Christin (2020) describe "algorithmic management" as systems that assign tasks, evaluate performance, and enforce standards—particularly in gig economy platforms. Workers often experience algorithmic systems as managerial authorities, even though formal leadership remains human.

Raisch and Krakowski (2021) propose the "automation–augmentation paradox," suggesting that AI simultaneously centralizes and decentralizes authority. While automation can standardize decision-making, augmentation increases reliance on human judgment. This dual dynamic creates new governance tensions in hybrid human–AI systems.

Brynjolfsson and McAfee (2017) argue that AI technologies can generate substantial productivity gains but require complementary organizational redesign. They emphasize that leadership structures must adapt to AI-enabled workflows, yet they stop short of fully reconceptualizing authority itself.

These studies collectively suggest that AI systems increasingly perform functions traditionally associated with leadership—task allocation, monitoring, coordination—without possessing

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

legal or moral personhood. As such, AI acts as a quasi-organizational actor, reshaping authority dynamics while lacking accountability capacity.

2.4. Epistemic Authority, Trust, and Human–AI Interaction

Beyond operational authority, AI systems increasingly exercise *epistemic authority*—influencing whose judgments are trusted. Research on algorithm aversion and appreciation provides insight into how humans respond to AI-generated recommendations.

Dietvorst, Simmons, and Massey (2015) show that individuals often lose trust in algorithms after observing minor errors, a phenomenon termed "algorithm aversion." In contrast, Logg, Minson, and Moore (2019) find evidence of "algorithm appreciation," where people prefer algorithmic advice over human judgment in certain contexts.

These conflicting findings underscore the contingent nature of epistemic authority. Trust in AI depends on perceived accuracy, transparency, and domain expertise. Mittelstadt et al. (2016) further argue that opacity and complexity in algorithmic systems complicate the attribution of responsibility and reduce meaningful oversight.

The literature on AI ethics also emphasizes the importance of transparency and explainability in maintaining legitimate authority. Floridi et al. (2018) propose principles such as beneficence, non-maleficence, autonomy, justice, and explicability as foundational to trustworthy AI governance. Without explicability, authority risks becoming illegible and unchallengeable.

Together, these studies demonstrate that AI's authority is not only operational but epistemic—shaping belief formation and decision reliance in organizations.

2.5. Accountability Gaps and the Problem of Responsibility

One of the most pressing issues in AI-enabled organizations is the emergence of accountability gaps. Mittelstadt et al. (2016) argue that algorithmic systems generate moral and legal challenges because harms may arise from complex interactions between training data, model architecture, and deployment contexts. This complexity makes it difficult to identify a singular responsible actor.

Bovens (2007) highlights that accountability requires identifiable actors and forums for justification. Yet in AI systems, responsibility may be distributed across developers, data scientists, managers, and executives. Yeung (2018) similarly warns that algorithmic regulation can undermine traditional accountability structures if oversight mechanisms are not carefully designed.

Pasquale (2015) emphasizes that opacity in automated systems exacerbates power asymmetries and reduces contestability. When decision-making logic is inaccessible, affected parties cannot meaningfully challenge outcomes.

This body of research converges on a critical insight: while AI systems can exercise operational and epistemic authority, they cannot bear moral-legal accountability. The responsibility therefore remains human, but organizational structures often fail to clearly allocate it.

2.6. Toward Agentic Organizations

While existing literature addresses algorithmic governance, AI ethics, and organizational adaptation, it does not fully theorize organizations in which AI systems possess delegated discretion to initiate actions. The shift from assistive to agentic AI intensifies existing tensions identified in prior research.

Raisch and Krakowski (2021) acknowledge hybrid authority structures but do not disaggregate operational, epistemic, and moral dimensions of authority. Similarly, Kellogg et al. (2020)

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

document algorithmic control without providing a normative framework for layered accountability.

This review suggests a conceptual gap: the absence of an integrated framework that explains how authority and accountability should be structured when AI systems act autonomously within organizations. By synthesizing insights from authority theory, algorithmic governance, human–AI interaction, and accountability scholarship, the present article advances the concept of *agentic organizations* to address this gap.

3.Conceptual Framework: Authority and Accountability in Agentic Organizations

3.1. From Assistive AI to Agentic Organizations

Organizations are increasingly integrating AI systems that do more than provide recommendations—they initiate actions, coordinate workflows, and adapt dynamically to feedback. This shift reflects what Raisch and Krakowski (2021) describe as the automation–augmentation paradox, where AI simultaneously automates decision processes and reshapes human roles. As AI systems assume greater discretion, authority structures become hybrid rather than purely hierarchical.

Traditional organizational theory conceptualizes authority as formally assigned and hierarchically structured (Mintzberg, 1980; Weber, 1978). However, algorithmic systems embed rules and optimization logics directly into operational processes, effectively relocating elements of authority from managers to technical infrastructures. Yeung (2018) argues that algorithmic regulation operates through continuous data feedback loops, producing governance effects without conventional human deliberation. This transformation suggests the emergence of what we term **agentic organizations**—organizations in which AI systems possess delegated operational discretion within defined parameters.

We define agentic organizations as socio-technical systems characterized by:

1. Delegated algorithmic discretion,
2. Embedded decision infrastructures, and
3. Distributed responsibility structures.

3.2. A Three-Dimensional Model of Authority

To clarify how authority operates in agentic organizations, we distinguish among three analytically separable dimensions: **operational authority, epistemic authority, and moral-legal accountability**.

3.2.1 Operational Authority

Operational authority refers to the capacity to execute decisions and coordinate action. In traditional organizations, this authority is exercised by managers and supervisors. In agentic organizations, AI systems increasingly allocate tasks, optimize schedules, or autonomously trigger actions.

Research on algorithmic management demonstrates how AI systems allocate work, monitor performance, and enforce standards (Kellogg et al., 2020). In such cases, workers experience algorithms as de facto supervisors. Similarly, Brynjolfsson and McAfee (2017) note that AI-driven automation can independently manage complex processes, from logistics to financial trading.

Operational authority in agentic organizations is therefore embedded in code and data pipelines. However, this authority remains bounded by system design and governance parameters. Humans retain meta-level authority over system deployment, but day-to-day execution may be algorithmically driven.

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

3.2.2 Epistemic Authority

Epistemic authority concerns whose judgments are trusted and relied upon in decision-making. As AI systems generate predictions and recommendations, they shape belief formation within organizations.

Dietvorst et al. (2015) show that humans may distrust algorithms after observing errors, while Logg et al. (2019) find evidence that individuals sometimes prefer algorithmic judgments over human advice. These contrasting findings highlight that epistemic authority is relational and context-dependent.

Mittelstadt et al. (2016) emphasize that opacity in algorithmic systems complicates explainability and limits informed oversight. Floridi et al. (2018) argue that explicability is essential for legitimate AI governance. Without transparency, AI systems may exercise epistemic authority without being contestable.

In agentic organizations, epistemic authority may shift toward AI systems in domains perceived as data-intensive or computationally complex. However, epistemic authority is not absolute; it depends on calibrated trust, interpretability mechanisms, and institutional norms.

3.2.3 Moral-Legal Accountability

Unlike operational and epistemic authority, AI systems cannot bear moral or legal responsibility. Accountability requires answerability and sanctionability (Bovens, 2007), conditions that AI systems do not satisfy.

The literature identifies accountability gaps arising from distributed system design. Mittelstadt et al. (2016) argue that responsibility may be diffused across developers, deployers, and users. Yeung (2018) warns that algorithmic governance can obscure traditional chains of accountability. Pasquale (2015) further highlights that opaque algorithmic systems undermine contestability.

Thus, while AI systems may exercise operational and epistemic authority, moral-legal accountability remains human and institutional. The misalignment between these dimensions generates governance tensions in agentic organizations.

3.3. Hybrid and Layered Authority Structures

Building on the three-dimensional model, we conceptualize authority in agentic organizations as **hybrid and layered**.

1. **Hybrid Authority**: Authority emerges from interactions between human leaders and AI systems. Raisch and Krakowski (2021) suggest that organizations must balance automation and human judgment, rather than fully substituting one for the other.
2. **Layered Accountability**: Accountability must be distributed across multiple organizational layers:
   - **Design Layer**: Developers and data scientists responsible for model architecture and training data.
   - **Deployment Layer**: Managers who integrate AI systems into workflows.
   - **Oversight Layer**: Executives and governance bodies responsible for monitoring, auditing, and compliance.

This layered model addresses the "problem of many hands" (Bovens, 2007) by clarifying roles across system life cycles.

3.4. Delegation Architecture and Traceability

A key conceptual component of agentic organizations is the **delegation architecture**—the formal specification of what authority is delegated to AI systems and under what constraints. Drawing from organizational design theory (Mintzberg, 1980), delegation must include:

- Defined scope of algorithmic discretion,

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

- Escalation protocols for human override,
- Transparent documentation of decision pathways.

Yeung (2018) emphasizes that algorithmic regulation depends on data feedback loops; therefore, traceability mechanisms must ensure that outputs can be audited retrospectively. Floridi et al. (2018) similarly stress that explicability and accountability must be built into system design.

Traceability serves as a structural bridge between operational authority and moral accountability, ensuring that human actors remain identifiable and answerable.

3.5. The Authority–Accountability Alignment Principle

We propose the **Authority–Accountability Alignment Principle**:

The greater the operational and epistemic authority delegated to AI systems, the more clearly defined and robust the corresponding human accountability structures must be.

This principle addresses structural asymmetry: AI may act autonomously, but responsibility cannot be automated. Governance frameworks must therefore proportionally strengthen oversight as discretion increases.

Empirical research on algorithmic management (Kellogg et al., 2020) shows that lack of transparency leads to worker resistance and legitimacy concerns. Similarly, Pasquale (2015) demonstrates that opacity undermines institutional trust. These findings support the necessity of alignment between delegated authority and accountability mechanisms.

4. Explanation of the Conceptual Model: Authority and Accountability in Agentic Organizations

The model of **Authority and Accountability in Agentic Organizations** explains how leadership and responsibility are structured when AI systems act with delegated discretion inside organizations. It integrates organizational theory, algorithmic governance, and AI ethics scholarship to clarify how authority shifts in AI-enabled environments.

The framework has two core components:

1. **Three Dimensions of Authority (Top Layer)**

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

2. **Layered Accountability Architecture (Bottom Layer)**

Together, they address a central structural tension: AI systems can exercise authority in practice, but they cannot bear responsibility in principle.

I. The Three Dimensions of Authority

The model separates authority into three analytically distinct but interrelated dimensions:

- Operational Authority
- Epistemic Authority
- Moral–Legal Accountability

This separation is necessary because AI systems increasingly exercise the first two, but never the third.

1. Operational Authority – "Who Executes Decisions"

Operational authority refers to the power to execute decisions and coordinate action.

Traditionally, operational authority resides in managers who allocate tasks, monitor performance, and enforce procedures (Mintzberg, 1980; Weber, 1978). However, AI systems increasingly perform these same functions.

Research on algorithmic management shows that AI systems assign tasks, monitor workers, and enforce standards—often with minimal human intervention (Kellogg et al., 2020). Workers frequently experience algorithmic systems as their de facto supervisors.

Similarly, Brynjolfsson and McAfee (2017) describe how AI systems independently manage logistics, pricing, and trading decisions, effectively operationalizing strategy in real time.

Thus, operational authority in agentic organizations becomes embedded in code, decision rules, and machine-learning models. AI systems execute decisions within predefined boundaries, exercising delegated discretion.

Key insight:

Operational authority can be partially delegated to AI systems.

2. Epistemic Authority – "Whose Judgments Are Trusted"

Epistemic authority refers to whose knowledge or judgment is relied upon in decision-making.

AI systems increasingly generate predictions, classifications, and risk scores that shape organizational decisions. When humans defer to these outputs, AI gains epistemic authority.

Research shows mixed patterns:

- People often avoid algorithms after observing errors (algorithm aversion) (Dietvorst et al., 2015).
- In other contexts, people prefer algorithmic advice over human advice (algorithm appreciation) (Logg et al., 2019).

This indicates that epistemic authority is socially constructed and context-dependent.

However, epistemic authority becomes problematic when systems are opaque. Mittelstadt et al. (2016) argue that opacity limits contestability and informed oversight. Floridi et al. (2018) emphasize explicability as essential for legitimate AI governance.

In agentic organizations, epistemic authority shifts toward AI in data-intensive domains. But this authority must be supported by transparency, interpretability, and calibrated trust mechanisms.

Key insight:

Epistemic authority is relational and contingent on transparency and trust calibration.

3. Moral–Legal Accountability – "Who Bears Responsibility"

Unlike operational and epistemic authority, moral–legal accountability cannot be delegated to AI systems.

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

Accountability requires answerability and sanctionability (Bovens, 2007). AI systems cannot explain themselves in normative terms nor bear legal sanctions.

Scholars identify accountability gaps in AI systems because responsibility becomes distributed across developers, managers, and institutions (Mittelstadt et al., 2016; Yeung, 2018). Pasquale (2015) further argues that black-box systems undermine contestability and democratic oversight.

This creates a structural asymmetry:

- AI may execute decisions (operational authority)
- AI may influence judgments (epistemic authority)
- But AI cannot bear responsibility (moral accountability)

Key insight:

Responsibility always remains human and institutional.

II. Hybrid and Dynamic Authority

The arrows between the three dimensions in the model represent dynamic interaction.

Raisch and Krakowski (2021) show that AI both automates and augments decisionzmaking. Authority becomes hybrid: shared between humans and machines.

Authority in agentic organizations is therefore:

- Not purely hierarchical
- Not purely algorithmic
- But socio-technical and dynamic

This redefines leadership as emerging from interactions between humans and AI systems rather than residing solely in individuals.

III. Layered Accountability Architecture

The second part of the model (bottom layer) specifies how accountability must be structured.

Because AI cannot bear responsibility, organizations must design **layered accountability systems** across three levels:

1. Design Layer – Developers & Data Scientists

Responsible for:

- Model architecture
- Training data
- Testing and validation

Mittelstadt et al. (2016) emphasize that bias and harm often originate at this stage. Therefore, responsibility for technical robustness and fairness begins here.

2. Deployment Layer – Managers & Implementation Teams

Responsible for:

- Contextual integration
- Monitoring system performance
- Human override decisions

Kellogg et al. (2020) demonstrate that managerial design of algorithmic systems shapes worker experience and power distribution. Deployment choices determine real-world impact.

3. Oversight Layer – Executives & Governance Boards

Responsible for:

- Strategic alignment
- Ethical compliance
- Audit and accountability structures

Yeung (2018) argues that algorithmic regulation must be embedded within institutional oversight frameworks to prevent governance erosion.

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

IV. Delegation Architecture and Traceability

The right side of the model highlights **Delegation Architecture & Traceability**.

Delegation architecture refers to clearly defining:

- What authority is delegated to AI
- Under what constraints
- With what override mechanisms

Mintzberg (1980) emphasizes formalized structures for authority delegation in traditional organizations. In agentic organizations, these must be codified in system design.

Traceability ensures:

- Decision logs
- Audit trails
- Explainability mechanisms

Floridi et al. (2018) argue that explicability is a prerequisite for ethical AI. Without traceability, accountability collapses.

V. Authority–Accountability Alignment Principle

The model culminates in a normative principle:

The greater the operational and epistemic authority delegated to AI systems, the stronger and clearer the corresponding human accountability structures must be.

This principle resolves the asymmetry between machine autonomy and human responsibility.

Empirical research supports this alignment necessity:

- Opacity reduces legitimacy and trust (Pasquale, 2015).
- Poorly governed algorithmic systems create resistance and power conflicts (Kellogg et al., 2020).

Thus, accountability must scale with autonomy.

VI. Theoretical Contribution

The model advances scholarship in four ways:

1. It disaggregates authority into operational, epistemic, and moral dimensions.
2. It explains structural asymmetry in AI-enabled organizations.
3. It integrates algorithmic governance and organizational theory.
4. It provides a design-oriented accountability framework.

In doing so, it reconceptualizes leadership as socio-technical rather than exclusively human.

5. Discussion

The rise of agentic AI systems fundamentally destabilizes conventional assumptions about leadership, authority, and responsibility within organizations. As AI systems increasingly initiate actions, allocate resources, and generate judgments without continuous human intervention, authority becomes partially embedded in socio-technical infrastructures rather than exclusively located in identifiable individuals. This structural shift raises complex questions about control, legitimacy, and accountability that existing governance models struggle to resolve.

One central tension concerns the decoupling of operational authority from moral accountability. Organizational theory traditionally assumes that those who exercise authority can be held accountable for outcomes (Weber, 1978; Bovens, 2007). However, AI systems can execute decisions and shape outcomes while lacking the capacity for answerability or sanctionability. This asymmetry generates what scholars describe as accountability gaps (Mittelstadt et al., 2016). When harms arise from AI-mediated decisions—such as biased hiring tools, discriminatory lending models, or automated risk assessments—responsibility is often distributed across developers, managers, and institutions, complicating attribution.

Policy Journal of Social Science Review

Online ISSN 3006-4635

Print ISSN 3006-4627

Vol. 3 No. 12 (2025)

The diffusion of responsibility is amplified by the opacity of many AI systems. Pasquale (2015) argues that black-box decision systems obscure the reasoning behind outcomes, reducing transparency and contestability. Similarly, Yeung (2018) notes that algorithmic regulation operates through data-driven feedback loops that may bypass traditional deliberative processes. When authority is exercised through automated systems whose logic is inaccessible or difficult to interpret, affected stakeholders may struggle to challenge decisions, thereby weakening institutional legitimacy.

Another key issue concerns epistemic authority and trust calibration. Research demonstrates that individuals sometimes over-rely on algorithmic outputs (Logg et al., 2019) while in other cases rejecting them after observing minor errors (Dietvorst et al., 2015). These dynamics suggest that epistemic authority is neither automatic nor stable. In agentic organizations, fluctuating trust levels can generate coordination challenges. Overconfidence in AI may lead to automation bias, while excessive skepticism may undermine performance gains. The dynamic interplay between human judgment and algorithmic recommendation thus becomes a central governance concern.

Moreover, the automation–augmentation paradox identified by Raisch and Krakowski (2021) remains particularly salient in agentic contexts. As AI systems automate routine tasks, they simultaneously elevate the importance of human oversight and exception handling. This paradox can create organizational ambiguity: while AI may execute the majority of operational decisions, humans retain ultimate responsibility for failures. Such arrangements risk creating symbolic oversight, where formal authority remains human but substantive control is algorithmically structured.

The literature on algorithmic management further highlights power redistribution effects. Kellogg et al. (2020) show that algorithmic systems can centralize control by standardizing performance metrics and reducing discretionary space for workers. At the same time, algorithmic systems may decentralize authority by embedding decision rules directly into workflows. This dual movement complicates hierarchical governance structures and may alter organizational cultures, particularly in platform-based or digitally mediated environments.

Ethical governance frameworks emphasize transparency, fairness, and explicability as safeguards against these risks (Floridi et al., 2018). However, embedding such principles into dynamic machine-learning systems remains technically and institutionally challenging. AI systems that adapt over time may evolve beyond their initial design parameters, raising questions about continuous oversight and responsibility maintenance. Mittelstadt et al. (2016) argue that many harms in AI systems arise not from explicit intent but from systemic interactions between data, model design, and deployment contexts.

The problem of many hands (Bovens, 2007) becomes especially pronounced in AI-enabled organizations. Responsibility may span data collectors, software engineers, managers, executives, and external vendors. Without clearly structured delegation and documentation mechanisms, accountability may become diluted. Yeung (2018) warns that algorithmic governance, if not carefully institutionalized, may erode democratic oversight and weaken mechanisms of justification.

Additionally, the socio-technical nature of agentic organizations suggests that authority is relational rather than static. Authority emerges through interactions among humans, algorithms, and institutional rules. Brynjolfsson and McAfee (2017) emphasize that AI's productivity benefits depend on complementary organizational redesign. Without such redesign, authority misalignment may produce inefficiencies, resistance, or legitimacy crises.

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

The discussion therefore underscores a structural transformation: authority in agentic organizations becomes hybrid, distributed, and infrastructure-dependent. AI systems may exercise operational and epistemic authority, but moral and legal accountability remains anchored in human and institutional actors. This structural asymmetry introduces governance tensions that demand careful alignment between delegated discretion and oversight mechanisms.

Ultimately, agentic organizations do not eliminate leadership; they reconfigure it. Leadership becomes embedded not only in people but also in architectures of delegation, traceability, and review. The central governance challenge is ensuring that as AI systems gain operational autonomy, institutional accountability remains clear, robust, and contestable.

6. Theoretical Implications

The model of authority and accountability in agentic organizations generates several important theoretical implications for organizational theory, leadership studies, governance scholarship, and AI ethics. These implications emerge from the structural separation between operational authority, epistemic authority, and moral–legal accountability, and from the hybrid nature of human–AI systems.

6.1. Reconceptualizing Authority as Socio-Technical Rather Than Purely Human

Classical authority theory conceptualizes authority as residing in legitimate human actors embedded within institutional hierarchies (Weber, 1978). Organizational design literature similarly assumes that authority flows through formal roles and structural arrangements (Mintzberg, 1980).

The present framework challenges this anthropocentric assumption by demonstrating that authority can be partially embedded in technical infrastructures. Algorithmic systems execute decisions, enforce standards, and structure behavior without possessing legal personhood. This aligns with scholarship on algorithmic governance, which argues that power increasingly operates through socio-technical systems rather than solely through human deliberation (Danaher et al., 2017; Yeung, 2018).

The theoretical implication is that authority must be conceptualized as **emergent from human–machine interaction**, rather than as an exclusively human property. Authority becomes infrastructural, encoded in models, datasets, and decision pipelines.

6.2. Disaggregating Authority into Multiple Dimensions

Traditional organizational theories treat authority as relatively unified—those who decide are those who are responsible. However, accountability theory distinguishes answerability and sanctionability as essential components of responsibility (Bovens, 2007).

By separating authority into operational, epistemic, and moral–legal dimensions, the model extends accountability theory into AI contexts. It shows that:

- AI systems can hold operational authority (execute decisions).
- AI systems can hold epistemic authority (shape judgments).
- AI systems cannot hold moral–legal accountability.

This structural asymmetry refines theoretical understandings of delegation and control. It suggests that authority in digital organizations is **multidimensional and partially decoupled**, challenging assumptions of coherence between power and responsibility.

6.3. Extending the "Problem of Many Hands" to Socio-Technical Systems

Bovens (2007) describes the "problem of many hands" as the diffusion of responsibility across multiple actors. In AI-enabled systems, this problem intensifies because responsibility spans developers, managers, executives, and automated infrastructures.

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

Mittelstadt et al. (2016) argue that AI systems create accountability gaps due to distributed design and deployment processes. Yeung (2018) similarly notes that algorithmic regulation complicates traditional lines of institutional oversight.

The theoretical contribution here is to show that the problem of many hands is no longer merely organizational—it is socio-technical. Responsibility diffusion now occurs across human and technical layers. This extends accountability theory into the realm of distributed computational agency.

6.4. Advancing Hybrid Leadership Theory

Leadership theory has historically emphasized traits, behaviors, and relational dynamics among humans. However, research on algorithmic management demonstrates that AI systems increasingly perform supervisory and coordinating functions (Kellogg et al., 2020).

Raisch and Krakowski (2021) describe an automation–augmentation paradox in which AI reshapes managerial authority. The present model deepens this insight by proposing that leadership authority is hybrid and layered.

The theoretical implication is that leadership should be conceptualized as:

- Socio-technical rather than solely interpersonal
- Distributed across human and algorithmic actors
- Dependent on governance architecture

This reframes leadership from a person-centered construct to a systems-centered one.

6.5. Reframing Epistemic Authority in Organizational Decision-Making

Research on algorithm aversion and appreciation shows that trust in AI systems fluctuates depending on context (Dietvorst et al., 2015; Logg et al., 2019). These findings demonstrate that epistemic authority is socially negotiated.

The model contributes theoretically by distinguishing epistemic authority from operational authority. It suggests that trust in AI-generated knowledge may exceed, match, or fall below human trust depending on transparency and performance.

Floridi et al. (2018) emphasize explicability as essential for legitimate AI systems. Thus, epistemic authority is not purely technical—it is normatively constructed.

This advances knowledge-based theories of the firm by incorporating algorithmic knowledge production as a central element of organizational epistemology.

6.6. Reinterpreting Organizational Design in the Age of AI

Mintzberg (1980) emphasizes structural configurations that distribute authority and coordination mechanisms. The present framework extends this logic to digital infrastructures.

Delegation architecture—formal specification of AI discretion—becomes a core structural variable. Authority is not only assigned through reporting lines but encoded through system design.

Brynjolfsson and McAfee (2017) argue that AI requires complementary organizational redesign. The model theorizes what that redesign entails at the level of authority and accountability alignment.

This shifts organizational design theory toward incorporating computational architectures as constitutive elements of governance.

6.7. Clarifying the Normative Limits of Artificial Agency

The framework reinforces a normative boundary: AI systems may act, but they cannot be accountable in moral or legal terms.

Pasquale (2015) warns that black-box systems can obscure power structures and undermine contestability. By clearly distinguishing moral–legal accountability from other authority forms,

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

the model prevents conceptual slippage that might otherwise attribute agency or responsibility to machines.

The theoretical implication is that artificial agency must be understood as functionally agentic but normatively constrained. This helps preserve accountability theory within increasingly automated environments.

6.8. Integrating AI Ethics with Organizational Theory

AI ethics literature emphasizes transparency, fairness, and accountability (Floridi et al., 2018; Mittelstadt et al., 2016). However, it often operates separately from organizational design scholarship.

This framework integrates ethical principles into structural theory by showing how accountability must scale with delegated authority. It bridges governance ethics and institutional theory, positioning accountability as an architectural requirement rather than a post hoc correction.

7. Practical Implications

The model of authority and accountability in agentic organizations offers several practical implications for how organizations design, govern, and monitor AI-enabled systems. As AI systems gain operational autonomy, organizations must deliberately structure authority and responsibility to prevent accountability gaps, legitimacy risks, and governance failures.

7.1. Clarify Delegation Boundaries Before Deployment

Organizations should formally define:

- What decisions AI systems are authorized to execute
- The limits of algorithmic discretion
- Conditions that trigger human override

Research on algorithmic governance shows that automated systems can operate through continuous feedback loops, often beyond direct human awareness (Yeung, 2018). Without explicit delegation boundaries, operational authority may expand implicitly.

Raisch and Krakowski (2021) argue that AI requires careful balancing between automation and human augmentation. Practically, this means documenting where AI replaces, supports, or defers to human judgment.

Clear delegation reduces ambiguity about control and prevents unintentional authority drift.

7.2. Build Traceability and Auditability into System Architecture

Traceability is essential for accountability. Organizations should implement:

- Decision logs
- Version control for models
- Documented data provenance
- Explainability interfaces

Mittelstadt et al. (2016) highlight that harms often emerge from complex interactions between training data and deployment context. Without traceability, organizations cannot identify root causes.

Pasquale (2015) warns that opaque “black box” systems undermine contestability and public trust. Embedding audit trails and explainability tools ensures that decisions remain reviewable and defensible.

Operational authority without traceability increases legal and reputational risk.

7.3. Establish Layered Accountability Structures

Responsibility for AI systems should be distributed across clearly defined layers:

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

Design Layer
- Model fairness
- Bias testing
- Technical robustness

Deployment Layer
- Contextual integration
- Performance monitoring
- Escalation protocols

Oversight Layer
- Strategic governance
- Compliance oversight
- Risk assessment

Bovens (2007) emphasizes that accountability requires answerability and sanctionability. Layered structures ensure that responsibility does not disappear into technical complexity.

Kellogg et al. (2020) show that algorithmic management reshapes power relations in organizations. Without structured oversight, algorithmic systems may centralize control while obscuring responsibility.

7.4. Calibrate Human–AI Trust

Organizations must actively manage epistemic authority.

Research shows that humans may over-trust or under-trust algorithmic outputs (Dietvorst et al., 2015; Logg et al., 2019). Over-reliance can lead to automation bias; under-reliance can reduce performance gains.

Practical mechanisms include:
- Confidence scores with AI outputs
- Clear communication of model limitations
- Training programs for managers
- Human-in-the-loop review for high-risk decisions

Floridi et al. (2018) stress explicability as central to trustworthy AI. Transparent communication of system capabilities improves trust calibration.

7.5. Align Authority with Accountability

The greater the autonomy granted to AI systems, the stronger the oversight mechanisms must be.

Brynjolfsson and McAfee (2017) argue that AI productivity gains depend on complementary organizational redesign. Authority–accountability alignment is part of that redesign.

When AI systems execute high-impact decisions—such as hiring, credit scoring, or safety monitoring—executive-level governance should proportionally increase.

Misalignment creates accountability gaps and increases litigation and regulatory exposure.

7.6. Prepare for Regulatory Scrutiny

Algorithmic governance is increasingly subject to regulatory oversight. Yeung (2018) notes that automated regulation alters traditional governance mechanisms.

Organizations should proactively:
- Conduct impact assessments
- Maintain documentation for compliance review
- Implement internal AI ethics committees
- Regularly audit performance and fairness metrics

Policy Journal of Social Science Review
Online ISSN 3006-4635
Print ISSN 3006-4627
Vol. 3 No. 12 (2025)

Mittelstadt et al. (2016) emphasize that ethical risks are systemic rather than incidental. Proactive governance reduces compliance risk and enhances institutional legitimacy.

7.7. Redesign Leadership Roles

As AI systems assume operational functions, managerial roles shift toward:

- Exception handling
- Oversight and escalation
- Ethical decision arbitration
- Cross-functional coordination

Raisch and Krakowski (2021) describe how AI reshapes managerial authority. Leaders increasingly supervise systems rather than directly executing tasks.

Organizations should invest in leadership training focused on AI governance literacy, risk assessment, and socio-technical oversight.

7.8. Monitor Organizational Culture and Power Dynamics

Algorithmic management may alter perceptions of fairness and autonomy (Kellogg et al., 2020). If workers perceive AI systems as opaque or unchallengeable, trust may erode.

Practical safeguards include:

- Clear appeals mechanisms
- Transparent performance metrics
- Opportunities for human review

Pasquale (2015) warns that opacity concentrates power. Organizations must ensure that AI systems remain contestable and subject to review.

7.9. Institutionalize Continuous Monitoring

AI systems evolve over time through retraining and data drift. Governance cannot be static.

Organizations should:

- Conduct periodic bias and accuracy audits
- Monitor model drift
- Update delegation boundaries when system capabilities change

Floridi et al. (2018) argue that ethical AI governance must be ongoing rather than one-time compliance. Continuous monitoring prevents unnoticed expansion of operational authority.

8. References

Ahmed, T., Ajmal, M., & Haq, M. A. U. (2024). Knowledge-oriented leadership and innovative performance: Role of creative self-efficacy and organizational climate in the software industry of Pakistan. *International Journal of Knowledge and Learning, 17*(2), 119–138.

Ajmal, M. (2022). Ethical leadership and employee creative performance: Discussing the mediating role of employer feedback environment in software houses of Pakistan. *Reviews of Management Sciences, 4*(1), 192–210.

Ajmal, M., Islam, A., & Khalid, S. (2025). Knowledge transcendence as a catalyst for organizational consciousness development. *Research Consortium Archive, 3*(4), 2336–2252.

Ajmal, M., Islam, A., & Khan, I. (2023). Investigating the impact of benevolent leadership on idiosyncratic deals: Mediating mechanisms of psychological contracts and role breadth self-efficacy. *International Journal of Islamic Business, Administration and Social Sciences (JIBAS), 3*(4), 31–56.

Ajmal, M., Islam, A., & Khan, I. (2024). Enhancing innovative work behavior in Pakistan's software industry: The mediating impact of employee feedback seeking and psychological empowerment on transcendental leadership. *International Journal of Islamic Business, Administration and Social Sciences (JIBAS), 4*(1), 25–44.

Ajmal, M., Islam, Z., & Islam, A. (2025). Enhancing organizational performance in higher education through knowledge-centered culture and absorptive capacity: The mediating role of the knowledge creation process. *The Learning Organization, 32*(5), 733–756.

Ajmal, M., Khalid, S., & Islam, A. (2025). Embedding sustainability in public educational leadership: A meta-systems thinking approach. *ASSAJ, 4*(02), 3912–3928.

Ajmal, M., Khan, I., & Islam, A. (2024). Fostering employee creativity in software houses: Exploring the influence of ethical leadership, LMX, and employee feedback-seeking behavior. *Migration Letters, 21*(S9), 1–18.

Ajmal, M., Manzoor, W., & Khan, I. (2024). Exploring the nexus between servant leadership, affective commitment, psychological empowerment, and employer feedback environment. *Habibia Islamicus (The International Journal of Arabic and Islamic Research), 8*(2), 1–20.

Ajmal, M., Rahat, W., & Islam, A. (2024). Enhancing affective commitment through transcendental leadership: Unveiling the influence of altruistic mindset and intrinsic motivation in higher education. *International Journal of Leadership in Education*, 1–27.

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal, 13*(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x

Brynjolfsson, E., & McAfee, A. (2017). The business of artificial intelligence. *Harvard Business Review*. https://doi.org/10.2139/ssrn.2991064

Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., ... Shankar, K. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society, 4*(2). https://doi.org/10.1177/2053951717726554

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines, 28*, 689–707. https://doi.org/10.1007/s11023-018-9482-5

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

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Mintzberg, H. (1980). Structure in 5's: A synthesis of the research on organization design. *Management Science, 26*(3), 322–341. https://doi.org/10.1287/mnsc.26.3.322

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2). https://doi.org/10.1177/2053951716679679

Pasquale, F. (2015). *The Black Box Society*. Harvard University Press. https://doi.org/10.4159/9780674736061

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

Weber, M. (1978). *Economy and Society*. University of California Press. https://doi.org/10.1525/9780520350342

Policy Journal of Social Science Review

Online ISSN 3006-4635

Print ISSN 3006-4627

Vol. 3 No. 12 (2025)

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance, 12*(4), 505–523. https://doi.org/10.1111/rego.12158

Zulfiqar, N., Ajmal, M., & Islam, A. (2024). Examining the role of leader mindfulness and employee moral identity in shaping workplace dynamics: A study on nurses' experience of leader surface acting and incivility in hospitals. *Migration Letters, 21*(S9), 27–44.