

PAPER ID: 20260202014

Human–AI Cognitive Symbiosis: A Conceptual Framework for Decision Authority, Trust Formation, and Organizational Legitimacy in Industry 6.0 Firms

Neha Guliani¹ and Khushal Goyal²

¹Assistant Professor, Department of Management, Jagannath University, Delhi NCR, Bahadurgarh

²Young Researcher, Department of Management, Jagannath University, Delhi NCR, Bahadurgarh

Abstract: Promises of Industry 6.0 depend on intensive collaboration between human judgment and autonomous analytical systems. Yet this integration destabilizes a foundational assumption of management: those identifiable actors' author organizational decisions, a concern long rooted in theories of organizational decision behavior and bounded rationality [4], [20]. As algorithmic capabilities expand, cognitive influence shifts toward systems that calculate but do not bear responsibility. Organizations remain publicly accountable for outcomes they only partially generate. The tension is therefore institutional rather than technological—how firms distribute cognition while preserving legitimacy, a condition tied to recognized accountability structures [7]. To confront this challenge, this paper advances the Symbiotic Authority–Legitimacy Architecture (SALA), a theory-driven framework that conceptualizes decision making as a jurisdictionally layered process rather than a binary choice between human and machine control. SALA articulates three mutually reinforcing mechanisms. First, calibrated authority preserves human normative authorship while leveraging algorithmic inference consistent with human-centered AI governance thinking [13]. Second, relational trust develops through interpretability, contestability, and role clarity rather than predictive precision alone, reflecting sociological and organizational understandings of trust as interactional confidence [15], [25]. Third, legitimacy signaling embeds oversight, auditability, and participatory governance into routine organizational practice, addressing concerns associated with opaque algorithmic decision environments [12]. By shifting analytical focus from technological adoption to governance design, the framework reframes Industry 6.0 firms as socio-cognitive institutions whose credibility depends on how shared intelligence is structured, echoing sociomaterial perspectives that technologies participate in organizational action [8]. SALA therefore provides a conceptual basis for sustaining innovation while safeguarding professional agency, institutional responsibility, and stakeholder trust in an era where decisions are increasingly co-produced by human and artificial cognition [3].

Keywords: Human–AI Cognitive Symbiosis, Industry 6.0 Decision Authority Governance, Trust in Artificial Intelligence, Organizational Legitimacy.

Introduction: Distributed Cognition and the Paradox of Authority

Industry 6.0 reshapes the locus of reasoning. Predictive systems anticipate risks, prioritize opportunities, and suggest courses of action before human managers intervene [1], [3]. Yet organizations remain accountable for outcomes shaped by entities that lack moral agency, creating a paradox of delegated cognition: efficiency gains through AI coexist with erosion of explainability, discretion, and moral authorship [4], [20]. This tension is **institutional, not technological**. Classical management assumes decision authority aligns with human judgment [5], [21]. In hybrid cognitive systems, authority becomes dispersed across human and machine actors, challenging traditional mechanisms for sustaining trust, responsibility, and legitimacy [7], [8].

Limitations of Existing Perspectives

A significant portion of existing research rests on assumptions that inadequately capture the institutional consequences of advanced AI integration. First, artificial intelligence is

frequently treated as a neutral decision-support instrument. In contemporary organizational settings, however, algorithmic systems do more than assist; they actively influence what information is prioritized, how problems are framed, and which alternatives appear plausible, thereby shaping patterns of organizational attention and interpretation [8], [9]. Second, trust in AI is often equated with technical performance. Behavioral evidence suggests the opposite dynamic: decision makers may resist highly accurate systems when their reasoning is opaque, while showing greater acceptance of models that offer transparency and opportunities for understanding or contestation, even if their performance is imperfect [23], [25]. Third, legitimacy is commonly assumed to be an automatic outcome of improved efficiency. Institutional theory indicates that credibility depends less on output optimization than on visible accountability, procedural fairness, and the assurance that responsibility remains anchored in identifiable governance structures [7], [12].

Taken together, these assumptions overlook a more fundamental transformation—the migration of epistemic influence from human judgment toward algorithmic reasoning. As organizations increasingly rely on distributed cognition, questions of who authorizes knowledge, who bears responsibility, and how decisions are justified become central governance concerns. The challenge of Industry 6.0 therefore lies not simply in technological capability, but in designing institutional arrangements that stabilize authority, sustain trust, and preserve legitimacy under conditions of technologically mediated knowing.

Literature Review

Recent scholarship on human–AI collaboration has expanded rapidly across management, information systems, and organizational research, reflecting growing interest in hybrid decision environments. Existing studies emphasize augmentation benefits, trust dynamics, explainability, and governance considerations, yet their insights remain fragmented across disciplinary boundaries. To synthesize current knowledge, a structured review of key contributions from 2018 to 2026 is presented. The table highlights dominant themes, methodological orientations, and unresolved limitations, thereby clarifying the conceptual gap addressed by the present study.

No.	Paper Title	Introduction (Core Idea)	Methodology	Loophole / Limitation	Conclusion
1	Artificial Intelligence and the Future of Work: Human–AI Symbiosis in Organizational Decision Making	AI complements human decision making in organizations.	Conceptual literature synthesis	No empirical validation	Humans retain contextual and ethical judgment while AI augments analytics
2	Algorithms for the Greater Good! On Mental Modeling and Acceptable Symbiosis	Human mental models shape acceptable AI collaboration.	Conceptual AI framework	Not organizationally tested	AI must align with human expectations for effective symbiosis
3	Reciprocal Human–Machine Learning	Humans and AI learn from each other in decision processes.	Theoretical model with illustrative cases	Limited large-scale testing	Co-adaptation improves decision quality
4	Human–AI Synergy in Decision Making: Literature Review	Identifies patterns of human-AI decision synergy.	Systematic literature review	Cross-disciplinary variability	AI must match human cognitive strengths
5	Trust in AI: Progress, Challenges and Future Directions	Explores trust dimensions in AI adoption.	Systematic review	User-level focus	Trust is essential for authority delegation
6	Generative AI as a Catalyst for HRM Practices	AI affects HR decisions through trust mechanisms.	Survey and structural equation modeling	Sector-specific	Trust mediates organizational adoption
7	AI Governance: A Systematic Literature Review	Governance ensures ethical and legitimate AI use.	Systematic review	Broad scope	Transparency drives legitimacy
8	Human-Tech Symbiosis in	Managers and	Conceptual	Limited empirical	Shared cognition

	Strategic Decision Making	AI share decision authority.	framework	depth	enhances strategic foresight
9	Human–AI Collaboration in Organizations	Integrates intuition and machine intelligence.	Multidisciplinary synthesis	No longitudinal testing	Ethical alignment improves outcomes
10	Trust Management for Human–AI Symbiosis	Proposes dynamic trust calibration.	Framework design	Not empirically validated	Continuous trust management sustains collaboration
11	Human–AI Governance (HAIG): Trust-Utility Approach	Links trust with governance design.	Conceptual governance model	Pre-peer-review	Balanced trust improves legitimacy
12	Agentic AI Organizational Socialization	AI acts as organizational actor shaping culture.	Conceptual with cases	Single industry focus	AI affects workplace trust climates
13	The AI Trust Factor: Automation and Decision Making	Trust structures shape AI entrepreneurial decisions.	Multiple correspondence analysis	Entrepreneurship context	Trust determines authority distribution
14	AI as Teammate or Tool? Human-AI Decision Support	Humans negotiate decision authority with AI.	Literature review	Early research stage	Shared mental models enable collaboration
15	Social Integration of Artificial Intelligence	Balances automation and autonomy.	Conceptual framework	Limited governance analysis	Calibrated autonomy increases trust
16	Explainable AI and Trust Calibration	Explainability increases trust.	Experimental user studies	Laboratory setting	Transparency improves acceptance
17	Human-Centered AI and Organizational Transformation	Human agency anchors AI adoption.	Conceptual bibliometric review	Industry 5.0 focus	Human-centric deployment is necessary
18	Decision Support Systems and Human Oversight	Early human-in-the-loop AI decision systems.	Empirical DSS case analysis	Low AI maturity	Human oversight maintains legitimacy
19	Organizational Legitimacy in the Age of Algorithms	Transparency shapes stakeholder legitimacy.	Conceptual theory development	No quantitative validation	Accountability drives legitimacy
20	Human–AI Symbiotic Sense-Making and Organizational Resilience	Humans interpret AI outputs in uncertainty.	Organizational case study	Single organization	Symbiosis improves resilience

To integrate the theoretical streams discussed above, the proposed Human–AI cognitive symbiosis framework is illustrated in Fig. 1.

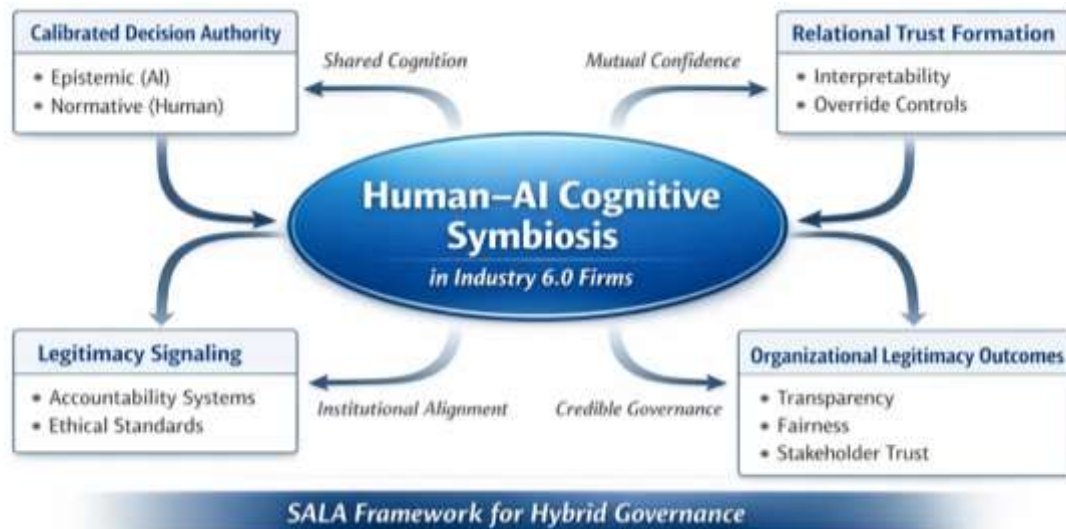


Fig.: Conceptual model of Human–AI cognitive symbiosis in Industry 6.0, showing the interdependent relationships among calibrated decision authority, relational trust formation, and organizational legitimacy within a shared governance structure.

The diagram illustrates how human and artificial intelligence jointly contribute to organizational decision processes through a structured symbiotic framework. It shows that calibrated decision authority forms the foundation for shared cognition and balanced responsibility. Trust emerges through interpretability, human oversight, and continuous interaction between users and intelligent systems. These elements collectively strengthen governance credibility and lead to sustained organizational legitimacy in Industry 6.0 environments.

Cross-Disciplinary Insights

Interpreting the organizational implications of Industry 6.0 demands a perspective that crosses disciplinary boundaries rather than remaining confined to technology-centric explanations. From an organizational cognition standpoint, decision outcomes are shaped by collectively shared interpretive frameworks that guide how actors notice, interpret, and respond to emerging situations [6]. This cognitive dimension intersects with authority theory, which emphasizes that decision acceptance depends not only on technical validity but also on whether the source of judgment is perceived to operate within a legitimate and recognized jurisdiction [16]. Institutional scholarship further suggests that practices become stable and credible when they align with broader social norms, expectations, and symbolic standards of appropriateness rather than merely delivering operational efficiency [7]. At the interpersonal and systemic level, trust research highlights that confidence develops through consistent interactional

experience, predictability of behavior, and the reduction of uncertainty over time [15], [25]. Complementing these perspectives, sociomateriality reframes digital technologies as constitutive elements of organizational action, actively shaping routines, meanings, and patterns of coordination rather than functioning as passive tools [8].

Viewed together, these theoretical strands point toward a fundamental reconfiguration of organizational agency. Decision processes increasingly emerge from the ongoing interplay between human interpretation and machine-based inference. This synthesis gives rise to the concept of **hybrid cognitive agents**, wherein organizations operate as integrated socio-technical entities that think, evaluate, and act through the combined reasoning capacities of human actors and intelligent systems. Such a view shifts analytical attention from technology adoption to the governance of shared cognition in environments where intelligence itself has become relational and distributed.

The Doctrine of Accountable Cognition (DAC)

The **Doctrine of Accountable Cognition (DAC)** proposes that organizational legitimacy in AI-mediated contexts rests on preserving human responsibility over the evaluative dimensions of decision making, even as analytical knowledge is increasingly generated through algorithmic processes. As intelligent systems assume a growing role in producing forecasts, classifications, and probabilistic assessments, the cognitive basis of decisions becomes technologically distributed. However, institutional acceptance depends not only

on the quality of these epistemic contributions but on the continued visibility of human judgment as the final source of authorization. DAC therefore distinguishes between **epistemic authorship**, which involves the production of data-driven insights and predictive outputs, and **normative authorship**, which concerns the interpretation of those insights in light of ethical standards, strategic priorities, and broader societal expectations. This distinction reflects long-standing perspectives in organizational decision theory that separate analytical evaluation from value-based judgment and responsibility allocation [4], [20].

From a governance standpoint, legitimacy emerges when stakeholders can identify where moral accountability resides and how decisions are justified within recognized authority structures [7]. When normative judgment appears displaced by opaque computational processes, organizations risk eroding trust and weakening their institutional credibility, regardless of performance gains. By contrast, maintaining human control over the evaluative and justificatory stages ensures that responsibility remains attributable, contestable, and publicly defensible. In this sense, DAC reframes effective AI integration not as the transfer of decision authority to machines, but as the careful preservation of human normative authorship within systems of distributed cognition, thereby sustaining organizational accountability and stakeholder confidence in increasingly automated environments [12], [15].

The Symbiotic Authority–Legitimacy Architecture (SALA)

The Symbiotic Authority–Legitimacy Architecture (SALA) translates the Doctrine of Accountable Cognition into an operational governance design for hybrid decision systems. Rather than collapsing authority into either human discretion or algorithmic automation, SALA structures decision processes

through differentiated yet interdependent layers. At the epistemic level, intelligent systems generate forecasts, classifications, and analytical scenarios; at the interpretive level, managers contextualize these outputs within organizational realities; and at the normative level, senior leadership assumes responsibility for value-based authorization and strategic endorsement. This layered distribution of roles reflects foundational insights from administrative theory, which distinguish between analytical evaluation and hierarchical responsibility in complex organizations [4], [5]. SALA further recognizes that trust in hybrid systems cannot be secured by predictive precision alone. Confidence develops when actors can understand the rationale behind algorithmic outputs, question their implications, and observe consistent feedback mechanisms that enable learning and correction over time [23], [25]. Interpretability, contestability, and structured review therefore become central design principles rather than peripheral technical features. Finally, legitimacy is reinforced through visible responsibility structures. Audit trails, independent oversight bodies, and participatory review processes make governance practices transparent and publicly defensible, addressing concerns about opaque algorithmic infrastructures and accountability diffusion [12], [13].

Taken together, SALA advances a model of co-decision in which human and machine reasoning are integrated without dissolving moral and institutional responsibility. The architecture preserves accountability while leveraging computational intelligence, ensuring that technological augmentation does not devolve into abdication of governance. The operational structure of the Symbiotic Authority–Legitimacy Architecture is presented in Fig.



Fig. 2. Symbiotic Authority–Legitimacy Architecture (SALA) illustrating the layered distribution of epistemic generation, interpretive mediation, and normative authorization to ensure accountable decision governance in hybrid human–AI systems.

The conceptual framework illustrates Human–AI cognitive symbiosis as an integrated governance system supported by three interdependent pillars: decision authority, trust formation, and legitimacy assurance. It shows how human normative oversight and AI analytical capability are structurally balanced to enable accountable shared cognition. Trust is developed through interpretation capacity, collaborative interaction, and structured human control mechanisms. Together, these elements operate within the SALA architecture to produce transparent governance and sustained organizational credibility in Industry 6.0 contexts.

Table 1: Layered Authority Structure in Human–AI Cognitive Symbiosis

Decision Layer	Primary Actor	Core Responsibility
Epistemic Layer	AI System	Data modeling, prediction, pattern detection
Interpretive Layer	Operational Managers	Contextual evaluation and scenario framing
Normative Layer	Executive Leadership	Ethical authorization and strategic accountability

The following table conceptualizes the layered distribution of decision authority in symbiotic human–AI systems. This table presents a three-tier authority configuration that distinguishes analytical production from moral authorization. It clarifies how algorithmic systems contribute knowledge without assuming institutional responsibility. Managers translate computational insights into context-sensitive interpretations before escalation. Final authorization remains anchored in leadership, preserving visible accountability within distributed cognition.

Research Propositions

P1: Normative authorship retention strengthens perceived legitimacy.

P2: Jurisdictional separation of computational and normative roles enhances accountability perception.

P3: Managerial interpretation of AI outputs produces greater trust than accuracy improvements alone.

P4: Structured override rights increase employee confidence.

P5: Reciprocal learning between humans and AI enhances integration.

P6: Publicly articulated oversight influences stakeholder trust more than operational efficiency.

P7: Implicit normative control by AI decreases legitimacy.

P8: Firms institutionalizing DAC and SALA architectures sustain long-term stakeholder confidence.

Table 2: Trust Formation Mechanisms in Human–AI Collaboration

Trust Dimension	Enabling Mechanism	Organizational Outcome
Cognitive Trust	Model explainability	Increased understanding of AI reasoning
Relational Trust	Structured override rights	Confidence in human control
Institutional Trust	Transparent audit systems	Stakeholder legitimacy reinforcement

The table below outlines the mechanisms through which trust develops in hybrid human–AI decision environments. This framework differentiates trust into cognitive, relational, and institutional dimensions rather than treating it as a single construct. Understanding how AI reaches conclusions strengthens confidence at the user level. Override structures reassure employees that discretion has not been surrendered. Transparent auditing extends trust beyond internal actors to external stakeholders.

Contributions

This study advances theory by introducing the Doctrine of Accountable Cognition (DAC) and the Symbiotic Authority–Legitimacy Architecture (SALA) as integrative frameworks for governing hybrid cognition systems. Rather than treating artificial intelligence as a performance-enhancing instrument, these constructs reposition governance as the central analytical problem in environments where human and machine reasoning are intertwined. In doing so, the paper re conceptualizes authority beyond traditional hierarchical command structures. Authority is reframed as a layered jurisdictional arrangement in which epistemic generation, interpretive mediation, and normative authorization occupy distinct but coordinated domains of responsibility, consistent with foundational perspectives on administrative responsibility and bounded rationality [4], [5].

The framework also contributes to trust theory by shifting attention from outcome-based confidence toward interactional dynamics. Trust is conceptualized as an emergent property of sustained engagement characterized by interpretability, procedural clarity, and the opportunity for review, aligning with sociological accounts that view trust as a mechanism for stabilizing expectations under conditions of complexity [15], [25]. Furthermore, the argument extends legitimacy theory by applying it to organizational contexts marked by distributed cognitive authorship. When analytical reasoning is co-produced by algorithms and human actors, legitimacy depends on the visible preservation of normative accountability within recognized institutional structures [7], [12].

Through SALA, Industry 6.0 organizations are thus reinterpreted as socio-cognitive institutions whose durability depends less on efficiency gains than on their capacity to structure authority, sustain trust, and demonstrate accountable governance. The emphasis shifts from technological sophistication alone to the institutional design of responsibility in settings where cognition itself has become technologically augmented.

Managerial and Policy Implications

The governance challenges of hybrid human–AI systems carry direct implications for managerial practice and public policy.

For managers, the primary task is not merely to allocate computational tasks but to articulate clear responsibility structures that specify who bears justificatory authority for consequential decisions. In AI-mediated environments, it is insufficient to identify who generated an output; organizations must define who evaluates its ethical implications, who authorizes its implementation, and who remains answerable for its consequences. Such clarity reinforces administrative responsibility and prevents the diffusion of accountability that can arise when analytical processes are technologically distributed [4], [7].

Accordingly, managerial development initiatives should prioritize interpretive competence over operational familiarity. As algorithmic systems increasingly automate data processing and pattern recognition, the distinctive human contribution lies in contextual judgment—understanding model assumptions, recognizing limitations, and translating probabilistic outputs into ethically and strategically sound actions. Training programs must therefore cultivate critical reasoning, scenario evaluation, and principled decision justification rather than focusing solely on system operation or technical configuration [20], [23].

From a regulatory perspective, effective oversight should center on traceable accountability mechanisms rather than blanket restrictions on intelligent technologies. Policymakers can enhance institutional trust by requiring auditability, documentation of decision pathways, and identifiable points of human authorization within automated workflows [12], [13]. Such measures preserve innovation while ensuring that responsibility remains visible and enforceable. Ultimately, sustainable success in Industry 6.0 depends on maintaining moral and institutional responsibility even as cognitive processes become increasingly distributed across socio-technical systems.

Table 3: Governance Signals and Legitimacy Outcomes

Governance Design Element	Visibility Level	Legitimacy Effect
Documented Decision Trails	Internal & External	Traceable accountability
Cross-Functional AI Board	Internal Governance	Balanced authority perception
Public Ethical Charter	External Stakeholders	Enhanced societal credibility

The following table links governance design choices to organizational legitimacy outcomes in Industry 6.0 firms. This table connects specific governance instruments to legitimacy stabilization in AI-mediated organizations. Documented decision trails make reasoning pathways reviewable and

contestable. Cross-functional oversight bodies prevent unilateral algorithmic dominance. Public ethical commitments signal that innovation is guided by principled standards rather than unchecked automation.

Limitations and Future Research

This paper is conceptual. Empirical validation is needed to examine sectoral variations, cultural influences on legitimacy perceptions, and longitudinal trust dynamics. Future research may also explore ethical pluralism and professional identity shifts in hybrid decision environments.

Conclusion

Industry 6.0 dissolves the clarity of authorship. Firms increasingly act through reasoning they do not fully generate. The survival of advanced enterprises depends less on algorithmic intelligence than on **deliberate governance of human–AI symbiosis**. Firms that obscure AI’s influence risk eroding credibility; those that structure decision authority, trust, and accountability as prescribed in DAC and SALA can achieve both innovation and institutional legitimacy. The next era of management is not about making decisions alone—it is about **ensuring accountable cognition when humans and intelligent systems decide together**.

References

- [1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Hoboken, NJ, USA: Pearson, 2021.
- [2] E. Brynjolfsson and A. McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York, NY, USA: W. W. Norton, 2014.
- [3] T. W. Malone, *Superminds: The Surprising Power of People and Computers Thinking Together*. Boston, MA, USA: Little, Brown, 2018.
- [4] H. A. Simon, *The Sciences of the Artificial*, 3rd ed. Cambridge, MA, USA: MIT Press, 1996.
- [5] J. G. March and H. A. Simon, *Organizations*. Oxford, U.K.: Blackwell, 1958.
- [6] K. E. Weick, *Sensemaking in Organizations*. Thousand Oaks, CA, USA: Sage, 1995.
- [7] M. Suchman, “Managing legitimacy: Strategic and institutional approaches,” *Academy of Management Review*, vol. 20, no. 3, pp. 571–610, 1995.
- [8] W. J. Orlikowski, “The sociomateriality of organisational life: Considering technology in management research,” *Cambridge Journal of Economics*, vol. 34, no. 1, pp. 125–141, 2010.
- [9] B. Latour, *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford, U.K.: Oxford Univ. Press, 2005.

- [10] L. Floridi et al., "AI4People—An ethical framework for a good AI society," *Minds and Machines*, vol. 28, no. 4, pp. 689–707, 2018.
- [11] K. Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT, USA: Yale Univ. Press, 2021.
- [12] F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA, USA: Harvard Univ. Press, 2015.
- [13] B. Shneiderman, *Human-Centered AI*. New York, NY, USA: Oxford Univ. Press, 2022.
- [14] D. Kahneman, O. Sibony, and C. R. Sunstein, *Noise: A Flaw in Human Judgment*. Boston, MA, USA: Little, Brown, 2021.
- [15] N. Luhmann, *Trust and Power*. Chichester, U.K.: Wiley, 1979.
- [16] P. M. Blau, *Exchange and Power in Social Life*. New York, NY, USA: Wiley, 1964.
- [17] M. Granovetter, "Economic action and social structure: The problem of embeddedness," *American Journal of Sociology*, vol. 91, no. 3, pp. 481–510, 1985.
- [18] J. Rawls, *A Theory of Justice*. Cambridge, MA, USA: Harvard Univ. Press, 1971.
- [19] J. Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge, MA, USA: MIT Press, 1996.
- [20] R. M. Cyert and J. G. March, *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1963.
- [21] H. Mintzberg, *The Structuring of Organizations*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.
- [22] A. Rai, "Explainable AI: From black box to glass box," *Journal of the Academy of Marketing Science*, vol. 48, no. 1, pp. 137–141, 2020.
- [23] B. J. Dietvorst, J. P. Simmons, and C. Massey, "Algorithm aversion: People erroneously avoid algorithms after seeing them err," *Journal of Experimental Psychology: General*, vol. 144, no. 1, pp. 114–126, 2015.
- [24] M. Janssen, G. Kuk, and R. Wagenaar, "A survey of trust in e-government adoption," *Government Information Quarterly*, vol. 25, no. 3, pp. 433–444, 2008.
- [25] R. Gulati, "Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances," *Academy of Management Journal*, vol. 38, no. 1, pp. 85–112, 1995.