

Guest editorial: The architecture of accountability: algorithmic autonomy and the discipline of stewardship

Introduction part I: The Blueprint of complicity

For decades, the competency paradigm has served as a dominant boundary condition in leadership theory. Predicated on the assumption that organizational outcomes are the determinative outcomes of personal agency and interpersonal dynamics, models such as transformational leadership (Bass, 1985) and leader–member exchange (LMX; Graen and Uhl-Bien, 1995) have tended to bracket a leader’s interpersonal traits from the performance realities of the architectures leaders authorize. As a result, these frameworks can be structurally underspecified for the present moment: the rise of the algorithmic agent as a primary organizational actor whose effects are mediated through scale, opacity and automated optimization.

Grounding our inquiry into sociotechnical systems (STS) theory (Trist, 1981), this volume treats leadership in the AI era as an emergent property of interaction between social actors and technical infrastructures. We argue that a traditional unit of analysis in organization development (OD), the leader’s mindset, no longer suffices on its own when algorithmic architectures function as consequential organizational actors. Following Drucker’s (1954) framing of management as a social organ responsible for institutional performance, we move beyond a primarily utilitarian inquiry into how leaders use AI to improve productivity and instead examine how organizational responsibility is operationalized, or surrendered, through the architectures that leaders authorize.

The urgency of this critique is no longer theoretical; it is being adjudicated in real time (Grabenstein, 2026a, b). As seen in the 2026 landmark trial in Los Angeles, Meta’s CEO was confronted with alleged systemic outcomes of Instagram’s architecture, sharpening what this volume terms the *performance gap* between stated values and observable results (Tan, 2026). Drucker’s “mirror test” offers a governance-relevant ethic for this gap: ethics requires that a leader’s actions remain compatible with the person they can live with seeing “in the mirror” each morning (Drucker, 1999/2005). When leaders authorize systems of *infinite reach* yet represent the resulting harms as operationally intractable (too complex, too diffuse or too difficult to govern), the managerial posture shifts from stewardship to authorization without effective oversight, leaving a social organ unmanaged at scale.

The Kaplan Paradox as a structural breach

We anchor this inquiry in the “Kaplan Paradox”: a condition in which a leader’s systemic liability expands in direct proportion to algorithmic reach, even as operational transparency, and thus perceived managerial control, diminishes. The paradox is sustained by shadow text: layers of algorithmic operation that are functionally inaccessible to the human agent. When leaders authorize opaque architectures through a digital “click-to-approve” autonomy, they define an automated reality while remaining legally and socially tethered to its outcomes.

As Griffin (2025) demonstrates, this “black box” can mask the deliberate ethical agency of developers – agency that management has a professional duty to govern. In agency-theoretic terms, leaders remain principals even when decision rights are delegated to technical agents;

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interests: The author declares no competing interests.



delegation changes monitoring and incentives, not responsibility for outcomes (Eisenhardt, 1989). By failing to embed the ecosystems of responsibility advocated by Stahl (2023), leaders enter a governance mode that resembles an ethical exception (Agamben, 2005), in which normal expectations for explanation, auditability and duty of care are suspended in practice. Reliance on delegation that is not readily auditable constitutes a structural breach of the duty of care, rendering leadership responsible for systemic malfunction, even when leaders experience diminished operational visibility.

The architecture of accountability: a diagnostic audit

The papers in this part provide the diagnostic and statutory components required to repair this professional breach.

- (1) *The audit (Seema; Khassawneh)*: Governance begins with the diagnostic identification of systemic “cracks” in which the machine becomes a site of human harm.
- (2) *The rule-book (Al Harrasi; Billsberry)*: Correction requires a fundamental remapping of organizational authority, reconceptualizing the leader’s role not as a manager of behavior but as a *manager of authorization logic*.
- (3) *The SOPs (Henderson; Sposato)*: Stewardship is transformed from a vague leadership trait into a measurable, auditable organizational reality through standard operating procedures (SOPs) for system maintenance.
- (4) *Material externalities (Araújo; Sadiku-Dushi)*: The audit must finally account for *ecological denial*. The CARE Governance Framework functions as a technical hard wall, operationalizing the evaluation of material costs and providing the structural resolution required for sustainable stewardship.

Introduction part II: The control room (the manual override)

If Part I provides the diagnostic audit of the “Kaplan Paradox”, Part II turns to the rigorous technical labor of the *active steward*, managing not only people but also the integrity of the systems through which power is exercised within the organizational control room (Drucker, 2010). Within this frame, we explicitly reject the managerial sedative – the assumption that empathy or culture alone can resolve the Kaplan Paradox. Empathy without systemic maintenance is not governance; it is performance.

Although Drucker did not live to see the AI era, his core claim remains germane: managers are responsible for *designing* institutions to produce specific results and for *maintaining the legitimacy* of those results through *disciplined practice* (Drucker, 1954, 1973). Extending this legacy, we shift the unit of analysis from a conventional human-in-the-loop (HIL) model to a machine-in-the-human-loop (MIL) architecture, in which algorithmic agents are treated as subordinate within a shell of human stewardship (Shneiderman, 2022), maintaining that shell requires more than leadership style. It requires algorithmic discretion as a digital evolution of *street-level discretion*: practical judgment applied to mitigate the rigidity and externalities of rule-bound systems under real-world conditions (Lipsky, 1980).

In this control room view, the leader must serve as the *conscience of the architecture*. Refilling the managerial vacancy created by automated delegation requires three stewardship postures derived from Drucker’s functionalist ethics:

Mirror test. The mirror test operates as a hard constraint against the “I didn’t know” defense: if a leader authorizes an agent that acts in ways the leader would not personally endorse, the leader has failed in self-management (Drucker, 1999/2005).

Feedback analysis. Feedback analysis operationalizes accountability as a method: leaders must declare the system’s intent, compare expected outcomes with actual results and intervene, via manual override, when the two diverge (Drucker, 1999/2005).

Pioneering the future. For Drucker, management is an organ of society charged with making the future rather than inheriting it by default (Drucker, 1954, 1973). In this sense, the “don’t poke the bear” posture is not prudence but abdication: leaders cannot be passengers in machines they built but refuse to drive (Woffard, 2022).

When managers do not assume this architectural responsibility, a vacancy forms: a vacuum of integrity in which both steward and user are reduced to cogs within a system whose defaults become destiny (Trist, 1981). The papers in this part examine the mechanisms by which organizations can reclaim agency through the technical labor of algorithmic discretion, re-establishing the manual override as a professional duty rather than an optional intervention.

Reclaiming agency: the manual override

Management discipline in the AI era is defined by the manual override. We define algorithmic discretion as the digital evolution of *street-level discretion* required to mitigate the inherent rigidity of automated rule sets.

- (1) *AI crafting*: (Zu) identifies the leader’s mandatory technical right to adjust algorithmic outputs to fluid social realities and
- (2) *Logic-audit safety valve* (Sharma) operationalizes human intuition not as a *soft* feeling, but as the active steward’s final, high-stakes check against systemic collapse.

The human sensor layer

To govern the machine loop, the steward requires constant telemetry.

- (1) *Distributed agency*: (Biswas) provides a framework for orchestrating a hybrid network that serves as a sensor layer for cultural friction and
- (2) *Creative self-efficacy* (Aljuhmani) operationalizes human creative engagement as a systemic diagnostic; a drop in engagement serves as a telemetry signal that the architecture has become misaligned.

Together, they signal precisely when the architecture requires an active override.

The global bill: from literacy to liability

The greatest vulnerability of the system architect is operational opacity. As Zuboff (2019) warns, the ultimate cost of algorithmic governance is the erosion of the *right to the future tense*, the human capacity to act with intent.

- (1) *Digital literacy as ethical liability* (Jiang): establishes literacy not as a technical skill but as a macro-level benchmark for ethical liability. Literacy is the only mechanism capable of piercing the invisibility of the Kaplan Paradox; without it, stewardship is rendered impossible, and complicity shifts from accidental to negligent.
- (2) *National financial stability* (Hasnat): frames governance as a jurisdictional requirement. Without it, both the steward and the user forfeit their capacity for directed agency, surrendering the future to the default settings of an unmanaged system.

Conclusion

Together, this two-volume collection moves beyond the traditional boundaries of the competency paradigm to establish a formal architecture of accountability. By integrating the diagnostic rigor of an institutional audit with the functional requirements of the active steward, these papers provide a theoretical and practical roadmap for reclaiming agency within

machine-mediated organizations. This transition, from managing interpersonal behavior to governing authorization logic, is not merely a technical adjustment; it is a fundamental evolution of the leadership duty of care.

The Kaplan Paradox reminds us that as algorithmic reach expands, the vacuum of integrity created by automated delegation can no longer be filled by *managerial sedatives* or performative ethics. Instead, it requires the rigorous, technical labor of algorithmic discretion and the operational reality of the manual override. Ultimately, the control room is defined throughout this special issue as not only a technical mechanism of absolute authority but also a site of professional responsibility requiring the *discipline of practice*, where the machine is brought back under the stewardship of the human hand and the institutional conscience.

Christine Haskell

Carson School of Business, Washington State University, Pullman, Washington, USA

Ethics approval

This article is based on secondary analysis of published accounts and does not involve new research with human participants; therefore, ethics approval was not required.

Data availability

No new data was created or analyzed in preparation of this work. Data sharing is not applicable to this article.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the author used ChatGBT to assist with citation verification and formatting. After using this tool, the author reviewed and edited the content and takes full responsibility for the content of the published article.

Acknowledgments

This Special Issue is the result of a nine-month journey that began at the International Leadership Association (ILA) Virtual AI Summit 2025, following an editorial panel discussion I facilitated. I am deeply indebted to Stefanie Johnson for her early vision in appointing me as Guest Editor and for granting me the intellectual freedom to chart the architecture for this issue. I also thank Rickard Enstroem, who provided steady stewardship and vital continuity as he assumed the leadership of the *Leadership and Organization Development Journal*.

A project of this complexity relies on an invisible infrastructure of professional labor. I wish to extend a special thanks to my friend, colleague, and fellow editor, Suzanne Joy Clark, whose assistance in co-auditing and managing the extensive volume of reviews was indispensable to the quality of the final collection. I also wish to thank Lauren Malone and Emma Ferguson for their ongoing and remarkably patient assistance in navigating the administrative and technical hurdles of a two-volume production.

The theoretical anchor of this issue, the Kaplan Paradox, was sparked by Christopher Woffard's 2022 *Wired* article, which provided the necessary vocabulary for the operational invisibility I have witnessed throughout my thirty years in the technology industry. However, a concept becomes a discipline only under empirical pressure. I am indebted to the authors of these papers for their intellectual courage in reframing their research to align with this "Hard Governance" architecture.

Finally, I must acknowledge the peer reviewers. Although they remain anonymous, their contributions are evident on every page. In an era when *managerial sedatives* often pass for scholarship, these reviewers served as the primary auditors of the logic in this issue. Their rigorous, often necessarily uncomfortable critiques ensured that our Control Room was built on a diagnostic foundation, examining the institutional duty of care in an era of automated scale. Their labor is the silent, generative heartbeat of academic discipline.

References

- Agamben, G. (2005), *State of Exception*, (K. Attell, Trans.), University of Chicago Press.
- Bass, B.M. (1985), *Leadership and Performance beyond Expectations*, Free Press.
- Drucker, P.F. (1954), *The Practice of Management*, Harper & Row, New York, NY.
- Drucker, P.F. (1973), *Management: Tasks, Responsibilities, Practices*, Harper & Row.
- Drucker, P.F. (1999), "Managing oneself", *Harvard Business Review*, Vol. 83 No. 1, pp. 100-109, Reprinted from Harvard Business Review, January 1999.
- Drucker, P.F. (2010), *On Managing Yourself*, Harvard Business Review Press, Boston, MA.
- Eisenhardt, K.M. (1989), "Agency theory: an assessment and review", *Academy of Management Review*, Vol. 14 No. 1, pp. 57-74, doi: [10.2307/258191](https://doi.org/10.2307/258191).
- Grabenstein, H. (2026a), "Landmark trial accusing tech giants of harming children with addictive social media begins", *PBS NewsHour*, available at: <https://www.pbs.org/newshour/nation/landmark-trial-accusing-tech-giants-of-harming-children-with-addictive-social-media-begins>
- Grabenstein, H. (2026b), "What legal experts say about a major 'bellwether trial' over child social media addiction", *PBS NewsHour*, available at: <https://www.pbs.org/newshour/nation/what-to-know-about-a-trial-that-will-test-tech-giants-liability-for-child-social-media-addiction>
- Graen, G.B. and Uhl-Bien, M. (1995), "Relationship-based approach to leadership: development of leader-member exchange (LMX) theory of leadership over 25 years: applying a multi-level multi-domain perspective", *The Leadership Quarterly*, Vol. 6 No. 2, pp. 219-247, doi: [10.1016/1048-9843\(95\)90036-5](https://doi.org/10.1016/1048-9843(95)90036-5).
- Griffin, T.A. (2025), "A showing of hands: making visible the ethical agency of AI developers", Doctoral dissertation, Maastricht University.
- Lipsky, M. (1980), *Street Level Bureaucracy: Dilemmas of the Individual in Public Services*, Russell Sage Foundation, available at: <http://www.jstor.org/stable/10.7758/9781610447713>
- Shneiderman, B. (2022), *Human-Centered AI*, Oxford, 2022, Online edn, Oxford Academic, doi: [10.1093/oso/9780192845290.001.0001](https://doi.org/10.1093/oso/9780192845290.001.0001).
- Stahl, B.C. (2023), "Embedding responsibility in intelligent systems: from AI ethics to responsible AI", *Scientific Reports*, Vol. 13, 7973, doi: [10.1038/s41598-023-34622](https://doi.org/10.1038/s41598-023-34622).
- Tan, E. (2026), "Mark Zuckerberg takes the stand in social media addiction trial", *The New York Times*, available at: <https://www.nytimes.com/2026/02/18/technology/mark-zuckerberg-tech-addiction-trial.html>
- Trist, E.L. (1981), *The Evolution of Socio-Technical Systems: A Conceptual Framework and an Action Research Program*, Ontario Ministry of Labour, Ontario Quality of Working Life Centre.
- Woffard, B. (2022), "The infinite reach of Joel Kaplan, Facebook's man in Washington", *Wired*, available at: <https://www.wired.com/story/facebook-joel-kaplan-washington-political-influence/>
- Zuboff, S. (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, PublicAffairs.

Further reading

- Ross, S.A. (1973), "The economic theory of agency: the principal's problem", *The American Economic Review*, Vol. 63, pp. 134-139.