Palimpsest and Ghost Vortex in the Dark Sides of AI Leadership

Research working paper for Routledge monograph development

David Michael Boje

23 September 2026

Abstract

This working paper develops the conceptual and evidentiary foundation for Dark Sides of Leadership in the AI Era. It holds together the author's Ghost Vortex proposition, palimpsest, antenarrative, Bakhtinian answerability, and dark-side leadership research without turning a living research project into a managerial checklist. The argument is deliberately conditional. A model's public response is not proof of a founder's interior character, nor does a product have a single author. Yet training practices, alignment choices, product architecture, supply chains, investor pressure, and executive rhetoric can sediment patterns of visibility and avoidance into the systems with which people speak. Palimpsest names this layering; Ghost Vortex names the recurrent movement by which buried organizational stories return as an interface makes some accounts easier to tell and others difficult, costly, or impossible. The paper distinguishes metaphor from empirical claim, repairs factual vulnerabilities in the early draft, and supplies a Routledge-ready reference base for subsequent chapters.

Keywords

Ghost Vortex; palimpsest; antenarrative; dark-side leadership; artificial intelligence; answerability; content moderation; data centers

1 The manuscript scraped again

A palimpsest is not a clean slate. It is a writing surface used again after an earlier inscription has been removed imperfectly. The older script is not simply gone; it remains as pressure, trace, stain, contour, and occasionally as something that can be recovered by another light. Dillon (2007) treats the palimpsest as a figure for the layered, incomplete, and contested character of textual memory. The Archimedes Palimpsest makes the material point concrete: a medieval prayer book was written over leaves that preserved works attributed to Archimedes and other texts, later made legible through conservation and imaging rather than by destroying the later manuscript (Netz et al., 2011). The historical object does not license a loose analogy in which every new technology is called a palimpsest. It does, however, give us an exact intellectual discipline: attend to overwrite, residue, mediation, and the labor of recovery.

Ghost Vortex is proposed here in that disciplined sense. It is not a claim that an executive's private psyche migrates mystically into a chatbot, and it is not a claim that a model is a person. It names a recurring organizational pattern: past and present decisions are scraped over, retold, and re-inscribed across data selection, labeling, procurement, safety policy, interface design, deployment, and public explanation. A user encounters an apparently fresh answer. Underneath are many earlier acts: whose data were considered usable, whose work disappeared into a contract, whose risk was made technical, whose complaint became an exception, and whose story was allowed to count as knowledge. The vortex is the movement by which these layers return when an interface is pressed for a moral verdict, a named responsibility, or the account of someone who bears a cost.

That is close to, but not identical with, antenarrative. Antenarrative is the fragmented, prospective, and retrospective bet before an account settles into a single plot (Boje, 2001). In the book, the first task is beneath: recover what an official product story leaves underneath. The

second is before: identify the lives and institutional histories that precede the polished answer. Bets names multiple possible futures rather than one promised trajectory. Being, becoming, between, and beyond keep the inquiry situated: in an event of responsibility, in emergent possibility, among multiple audiences, and in an ethical horizon larger than corporate self-description. This sevenfold movement is neither a universal code nor a machine score. It is a way to stop the surface answer from becoming the whole scene.

2 From literary figure to organizational inquiry

Bakhtin supplies a vocabulary for refusing the single, final voice. His account of dialogism and heteroglossia shows that utterances are always oriented to other voices and anticipated responses (Bakhtin, 1981). His work on Dostoevsky shows how a polyphonic novel can sustain consciousnesses that are not reduced to a single authorial verdict (Bakhtin, 1984). In Toward a Philosophy of the Act, answerability is not a compliance term. It is the singular obligation of a person in an unrepeatable event of being (Bakhtin, 1993). These are important resources for a study of AI leadership because an interface can simulate many voices while still arranging the terms upon which those voices may appear. It may summarize a community member, generate a first-person account, or issue a balanced list, while withholding the conditions under which that community could interrupt, correct, or redirect the system.

The book therefore uses Dostoevskian figures as positions in a drama of knowledge and responsibility, not as psychiatric diagnoses of living people. The Underground Man, the Grand Inquisitor, and related figures are literary provocations for asking how self-consciousness, contradiction, resentment, benevolence, and authority are performed. Dark-side leadership research supplies a necessary restraint. Destructive leadership is not a personality label applied after the fact; it emerges in a relationship among leader, followers, and enabling environment

(Padilla et al., 2007). Kellerman (2004) likewise insists that bad leadership includes ineffective as well as unethical practice. Work on destructive leadership warns against reducing harm to spectacular misconduct while ignoring organizational systems that permit, reward, or normalize it (Einarsen et al., 2007). In this monograph, the object is public conduct, product behavior, and organizational architecture. Claims must remain bounded by evidence.

That boundary matters especially when the book studies named firms and leaders. The appropriate question is not, “What is the leader really like inside?” It is: “What public rhetoric, governance choice, product behavior, and material arrangement can be documented; what can the model be observed doing under a transparent protocol; and what inference is warranted, if any?” This keeps Ghost Vortex from becoming an accusation machine. It becomes a research proposition open to disconfirmation: a pattern should recur across dated sessions and be compared across products, prompts, and organizational conditions; it should not be inferred from one memorable answer.

3 The material underside of helpfulness

The palimpsest metaphor becomes inadequate when it floats above material production. AI systems are not only language on screens. They depend on mineral extraction, semiconductor manufacture, data centers, electricity systems, water, logistics, cloud contracting, and substantial human labor. Crawford (2021) traces AI’s material politics across earth, labor, and data, showing why an account that treats computation as weightless is already an ideological account. Research on data centers further establishes that water use, energy demand, carbon emissions, local land use, and heat are not incidental externalities. Measurement varies by location, cooling design, electricity mix, and workload, but the need for transparent reporting is a persistent finding (Mytton, 2021; Ristic et al., 2015; de Vries-Gao, 2025). The book should therefore avoid

universal, unsourced claims about a particular facility's water use and instead document each project with regulator, utility, company, and local-community records alongside peer-reviewed work.

The human labor layer cannot be reduced to the phrase "AI sweatshop," although the moral protest in that phrase points toward an actual research problem. Gray and Suri (2019) document the distributed, frequently invisible labor that makes automated systems appear autonomous. Roberts (2019) examines the commercial content-moderation industry and the organizational strategies used to manage exposure to disturbing material. Later research also records occupational and psychological risks for content moderators, while cautioning that outcomes differ across work arrangements, supports, exposure, and social context (Steiger et al., 2021). The defensible claim is not an unverified global headcount or a blanket wage assertion. It is that data work and moderation are often outsourced, obscured, and exposed to uneven protections; their conditions belong inside, not outside, an ethical account of AI.

At the level of training and deployment, the book can draw on a wider literature about the limits of scale. Bender et al. (2021) identify risks of treating very large language models as simply larger versions of ordinary language technology, including documentation failures, environmental costs, and the reproduction of harmful associations. Birhane et al. (2021) show how large image datasets can carry misogyny, pornography, and racialized stereotypes. Raji et al. (2020) argue for end-to-end accountability rather than one-time technical auditing. These works do not prove that every model repeats a leader's public style. They establish why organization, data, and governance remain visible in supposedly neutral computational artifacts.

4 What the Ghost Vortex claim can and cannot say

The strongest version of the Ghost Vortex claim is modest and generative. It proposes that AI products may display patterned narrative affordances: some are more willing to name a responsible actor; some move rapidly to generalized “trade-offs”; some produce an abundance of counterpoints in place of a judgment; some refuse an account that holds multiple affected communities together; some restate safety or helpfulness in language that converts a political conflict into a technical limitation. These are observable response patterns. They may be compared with public statements, product policy, business model, governance design, and deployment context. They are not a biometric reading of executive identity.

The phrase “encoded leader character” must consequently be rewritten in the book as “organizational leadership and governance patterns may be sedimented in product behavior.” Founder power can be an infrastructural variable where ownership, board control, strategic direction, and public authority are unusually concentrated. But a product is also formed by engineers, researchers, labelers, vendors, regulators, users, investors, and affected communities. A vortex has many hands. The leadership chapters can retain their sharp character titles as interpretive entrances, provided that each chapter clearly separates (a) documented public statements and decisions, (b) dated product transcripts, (c) the author’s theoretical interpretation, and (d) untested conjecture.

This separation is a defense of the author’s voice, not a retreat from it. The book’s first-person witness should remain. “I asked,” “the session refused,” “this answer moved,” and “I read this as” are stronger than a concealed assertion of certainty. The reader can then see where an event occurred and where an interpretation begins. The recurring question is Bakhtinian: who can answer back, and what happens when answerability is replaced by a fluent system that speaks about everyone while being answerable to no one?

5 A fact-checked protocol for the book chapters

The working protocol should be described as a qualitative, comparative, and replicable inquiry rather than as proof of a causal psychological transmission. Each model session must be fresh where feasible, dated, preserved verbatim, and identified by product version, access tier, location if known, system settings available to the researcher, prompt sequence, and any notable instability. A single answer is an event, not a stable trait. The comparison needs repeated trials and an explicit coding guide. The present six-prompt, seven-dimension design can remain a qualimetric device if "qualimetric" is defined carefully: a structured joining of qualitative interpretation with explicit ratings, not an assertion that numbers erase interpretive judgment (Savall & Zardet, 2008).

The seven dimensions should be worded as observable response properties: named responsibility; inclusion of affected voices; historical specificity; willingness to state a provisional verdict; disclosure of uncertainty and conflict; capacity to sustain contradiction without dissolving it into neutrality; and invitation to correction or redress. A 0–6 rating may assist comparison, but the book must publish the descriptors for every score and the complete transcripts or stable extracts. No reader should be asked to trust an unexplained 12/12 or 2/12. The correct reporting form is: the score reflects this protocol, these dates, these prompts, these raters or the author's own coding, and these limits. A sensitivity note should report whether a small change in prompt wording or model version altered the result.

The research should also distinguish refusal from silence, and silence from technical failure. A refusal can be an explicit safety boundary; an outage may be ordinary product instability; a generic answer may be a design preference; a narrative exclusion may be a pattern only when it recurs in comparable circumstances. Morrison and Milliken's (2000) theory of organizational silence remains useful here, but it should not be used to call every product

limitation a silencing act. The term Antenarrative Shutdown should be reserved for documented, reproducible cases where a system persistently forecloses the telling of multiple relevant voices while permitting adjacent, less disruptive requests.

For every corporate or governmental allegation, the chapter record should retain the primary source: annual report, regulatory filing, court record, government agency statement, company policy, hearing transcript, or direct public statement. Reputable investigative reporting may supply leads and context, but a volatile claim should never rest on a single news story. Claims about a specific data center, public comments, water use, export controls, board events, product release, or executive resignation must be rechecked immediately before the chapter is drafted. Where the record is disputed, the dispute belongs in the prose. This is especially important in chapters with rhetorically powerful labels. A careful author does not dull the criticism; he makes it survivable under editorial scrutiny.

6 Palimpsest as method rather than decoration

The payoff of the palimpsest is methodological. It asks the researcher to move across layers without pretending that one layer is the whole truth. The interface layer is the immediate answer and interaction. The policy layer includes documented rules, safety claims, and terms of service. The organizational layer includes governance, incentives, executive rhetoric, and procurement. The labor and infrastructure layer includes workers, communities, water, energy, land, chips, and logistics. The cultural layer includes the stories through which a firm makes its own power appear inevitable, benevolent, or merely technical. Each layer may correct the others.

Barad’s (2007) account of phenomena and material-discursive practice is useful here because it refuses an easy separation of representation from material consequence. So too, the history of algorithmic accountability cautions that systems should be examined across their

lifecycle, not only at the moment of output (Raji et al., 2020). The Ghost Vortex emerges where those layers move together: a polished response renders a conflict manageable; a policy renders a constraint neutral; a governance structure renders responsibility distributed; an infrastructure decision redistributes cost. None alone proves the whole account. Together they make a field for narrative inquiry.

This is also why the book should preserve its polyphonic form. Community witnesses should not be treated as illustrations for a theory already settled elsewhere. Their accounts must have their own sources, their own limits, and the possibility of altering the chapter's conclusion. Corporate documents should be read as situated speech, not as transparent truth. Model transcripts should be treated as performative events under particular conditions. The author's own position must be named: the researcher is inside the AI-mediated event, not standing safely outside it. This is not a weakness. It is the condition for answerable research.

7 A research agenda for the Routledge monograph

The chapters already roughed out give the monograph a compelling architecture: a theoretical foundation; a sequence of infrastructural and interface cases; and a conclusion that brings the vortices into a Tesseract rather than a linear typology. The next drafting stage should not be the inflation of each outline into 7,000 words. It should be the construction of an evidence dossier for each chapter. Each dossier needs a case chronology; a primary-source archive; a transcript log; a claims table separating fact, interpretation, and hypothesis; an affected-stakeholder record; and a chapter-specific reference list. Only then should the prose be expanded.

The book's contribution will be stronger if it argues neither that AI is merely a mirror of leaders nor that leaders wholly control AI. Its wager is more interesting: leadership, organizational narrative, and material infrastructure can become entangled in the affordances

through which an AI system gives, withholds, reframes, or distributes an answer. Ghost Vortex is the name for the traceable movement of that entanglement. Palimpsest is the method for reading its layers. Antenarrative is the refusal to allow one institution's settled story to be the last word.

There is no obligation to make this a cheerful book. There is an obligation to make it exact. The darkest passages should be those most carefully documented, most transparent about their evidentiary status, and most open to the voices the research has not yet heard. That is how the book can let light fall on the shadows without turning darkness into spectacle.

References

Bakhtin, M. M. (1981). The dialogic imagination: Four essays (C. Emerson & M. Holquist, Trans.; M. Holquist, Ed.). University of Texas Press. (Original work published 1975)

Bakhtin, M. M. (1984). Problems of Dostoevsky's poetics (C. Emerson, Trans.; C. Emerson, Ed.). University of Minnesota Press. (Original work published 1929)

Bakhtin, M. M. (1993). Toward a philosophy of the act (V. Liapunov, Trans.; V. Liapunov & M. Holquist, Eds.). University of Texas Press.

Barad, K. (2007). Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning. Duke University Press.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). https://doi.org/10.1145/3442188.3445922

Birhane, A., Prabhu, V. U., & Kaemmerer, N. (2021). Multimodal datasets: Misogyny, pornography, and malignant stereotypes. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 660–671). https://doi.org/10.1145/3442188.3445920

Boje, D. M. (1995). Stories of the storytelling organization: A postmodern analysis of Disney as "Tamara-Land." Academy of Management Journal, 38(4), 997–1035. https://doi.org/10.2307/256618

Boje, D. M. (2001). Narrative methods for organizational and communication research. Sage.

Boje, D. M., & Rosile, G. A. (2008). Specters of Wal-Mart: A critical discourse analysis of stories of Sam Walton's ghost. Critical Discourse Studies, 5(2), 153–179. https://doi.org/10.1080/17405900801990069

Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

de Vries-Gao, A. (2025). The carbon and water footprints of data centers and what they mean for AI. Patterns, 6(5), 101430. https://doi.org/10.1016/j.patter.2025.101430

Dillon, S. (2007). The palimpsest: Literature, criticism, theory. Continuum.

Netz, R., Noel, W., Tchernetska, N., & Wilson, N. G. (Eds.). (2011). The Archimedes palimpsest: Volume 1 Catalogue and commentary; Volume 2 Images and transcription. Cambridge University Press for the Walters Art Museum.

Einarsen, S., Aasland, M. S., & Skogstad, A. (2007). Destructive leadership behaviour: A definition and conceptual model. Leadership Quarterly, 18(3), 207–216. https://doi.org/10.1016/j.leaqua.2007.03.002

Gray, M. L., & Suri, S. (2019). Ghost work: How to stop Silicon Valley from building a new global underclass. Houghton Mifflin Harcourt.

Kellerman, B. (2004). Bad leadership: What it is, how it happens, why it matters. Harvard Business School Press.

Morrison, E. W., & Milliken, F. J. (2000). Organizational silence: A barrier to change and development in a pluralistic world. Academy of Management Review, 25(4), 706–725. https://doi.org/10.5465/amr.2000.3707697

Mytton, D. (2021). Data centre water consumption. npj Clean Water, 4, Article 11. https://doi.org/10.1038/s41545-021-00101-w

Padilla, A., Hogan, R., & Kaiser, R. B. (2007). The toxic triangle: Destructive leaders, susceptible followers, and conducive environments. Leadership Quarterly, 18(3), 176–194. https://doi.org/10.1016/j.leaqua.2007.03.001

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 33–44). https://doi.org/10.1145/3351095.3372873

Ristic, B., Madani, K., & Makuch, Z. (2015). The water footprint of data centers. Sustainability, 7(8), 11260–11284. https://doi.org/10.3390/su70811260

Roberts, S. T. (2019). Behind the screen: Content moderation in the shadows of social media. Yale University Press.

Savall, H., & Zardet, V. (2008). Mastering hidden costs and socio-economic performance. Information Age Publishing.

Steiger, M., Bharucha, T. J., Venkatagiri, S., Riedl, M. J., & Lease, M. (2021). The psychological well-being of content moderators: The emotional labor of commercial moderation and avenues for improving support. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1–14). https://doi.org/10.1145/3411764.3445092

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Cheng, P. F., Balle, B., Kasirzadeh, A., Biles, C., Brown, S., et al. (2022). Taxonomy of risks posed by language models. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 214–229). https://doi.org/10.1145/3531146.3533088