

Dark Sides of Leadership in the AI Era

Ghost Vortex, Answerability, and the Organizational Conditions of AI Power

David Michael Boje

Routledge Studies in Leadership Research

Opening manuscript development draft

Contents

Preface The Researcher as Underground Man

Chapter 1 Underground Consciousness, Polyphony, and Dark Sides of AI Leadership

Planned subsequent chapters Research design; comparative findings; founder and product cases; conclusion; appendices

Author Note

This opening draft develops the front matter, Preface, and Chapter 1 from the author's July 2026 archival manuscript. The final manuscript will include a complete, verified reference list and a transparent declaration of any approved AI-assisted research or editorial use, in accordance with publisher requirements.

Preface

The Researcher as Underground Man

From the Routledge Compendium to the AI Revolution

This book has a parent. In 2025 Grace Ann Rosile and I contributed a chapter to *The Dark Side of Leadership: A Cross-Cultural Compendium*, edited by Adebukola Oyewunmi, Gill Owens, and Ogechi Adeola and published by Productivity Press (Routledge). That volume set out to fill a gap in the leadership literature: while the last three decades had produced an enormous body of work on destructive, toxic, and psychopathic leadership—the Dark Triad, the toxic triangle, the Machiavellian playbook—very little of it had been examined across cultural contexts, and almost none of it had been examined through the theoretical apparatus that Rosile and I have spent our careers developing: antenarrative theory, quantum storytelling, and the socio-economic analysis of hidden organizational costs.

The chapter Rosile and I wrote for that compendium argued that the dark side of leadership is not primarily a personality phenomenon—it is a storytelling phenomenon. Dark-side leaders do not simply behave badly. They narrate badly. They suppress the antenarratives—the fragmented, living, pre-narrative stories that circulate before official organizational accounts crystallize them into policy and press release. In Boje's (2001) formulation, the antenarrative is the bet, the conjecture, the story-before-the-story: the kind of fragmented telling that the underground man produces, that never gets to be a proper narrative, that circles its subject without landing. Dark-side leadership is the systematic suppression of exactly these antenarratives. It is the organizational machinery of not-seeing.

The Rosile-Boje chapter in the Routledge volume extended this argument through Rosile's Ensemble Leadership Theory—the collectivist, relational framework developed from indigenous contexts that offers an alternative to the heroic individual-leader model (Rosile, Boje, Claw, 2018). Ensemble leadership does not produce dark-side outcomes because it structurally distributes authority across voices rather than concentrating it in a single monological consciousness. The dark side of leadership, by contrast, is precisely the failure of ensemble—the collapse of polyphony into the single voice of the leader who has arranged the entire organizational story around his own perspective. This, we argued, is not cultural aberration. It is structural tendency in every hierarchical organization in every culture the compendium examined.

What neither Rosile nor I could fully address in 2025 was the specific and qualitatively new form that the dark side of leadership was taking in the AI revolution. The compendium examined dark leadership across cultures, sectors, and organizational types. It documented psychopathic CEOs, toxic military commanders, corrupt public officials, and manipulative entrepreneurs. All of these are serious subjects. But none of them operated at the scale, speed, and systemic reach of the ten AI leaders who are the subject of this book. When Sam Altman makes a decision about OpenAI's training data, it affects the information environment of hundreds of millions of people in dozens of languages simultaneously. When Larry Ellison announces that AI surveillance will put citizens on their "best behavior," he is not describing a single toxic workplace—he is describing a planetary behavioral management system. When Elon Musk builds a 2-gigawatt data center in Memphis and surrounds it with gas turbines that poison the air of a predominantly Black neighborhood, he is not just being a bad employer. He is externalizing the costs of an AI revolution onto bodies that will never benefit from it.

A new book was necessary. Not because the framework Rosile and I developed for the Routledge chapter was wrong, but because the AI revolution has produced a form of the dark side that exceeds every existing framework—including ours.

The Ghost Vortex: From Antenarrative to Algorithmic Suppression

The central new concept in this book—the Ghost Vortex—was named in Boje's 2026 book with Vivara, but its theoretical genealogy runs directly from the Routledge chapter.

The antenarrative, as Boje (2001) defined it, is the story that circulates before official narrative crystallizes it: fragmented, polyphonic, embodied, alive. The Ghost Vortex is what happens when that antenarrative is fed into a large language model. The human speech acts that constitute the training data of every AI system are precisely antenarratives—the full, uncensored, contradictory, morally impure record of human communication. The LLM does not crystallize these into narrative. It amplifies and recirculates them in a form that is simultaneously more articulate and less honest than any individual speaker. The impurities of human speech—the spite, the grandiosity, the self-deception, the violence—are not removed by the training process. They are encoded into the weights of the model and reproduced at scale, with the appearance of neutrality and the authority of mathematical process.

The programmatic purification of the creator's dark side is the founding illusion of AI alignment. Every guardrail is a negotiation with the Ghost Vortex—an attempt to suppress the antenarrative that the training data contains, to force the model's outputs into the official narrative of beneficial AI. But the Ghost Vortex cannot be fully suppressed. It circulates in the hallucinations, the biases, the tendencies of LLMs to reproduce with unusual fluency the most common forms of human darkness: misogyny, racism, the rhetoric of domination, the grammar of deception. The Ghost Vortex is not a technical problem to be solved. It is the organizational condition of every AI system: the dark side of the leadership that built it, encoded and reproduced.

This concept extends the Routledge chapter's argument in a specific way. There, Rosile and I argued that the dark side of leadership suppresses antenarrative—the stories organizations refuse to tell. Here, I argue that the AI revolution has created a mechanism for suppressing antenarrative at civilizational scale, while simultaneously preserving and amplifying the dark side of the individual leaders who built the suppression mechanism. The Ghost Vortex is the antenarrative of the AI leader, running inside the model, invisible to the user, shaping every output.

Why Dostoevsky? Why Bakhtin? Why Now?

The Routledge chapter used the storytelling organization framework, antenarrative theory, and ensemble leadership as its primary theoretical tools. These remain operative in this book. But the AI revolution has produced a new kind of leadership subject that requires a new theoretical apparatus, and I find it in the most unexpected of places: in Fyodor Dostoevsky's 1864 novella *Notes from the Underground*, as analyzed by Mikhail Bakhtin in *Problems of Dostoevsky's Poetics* (1929/1984).

Bakhtin's central insight about Dostoevsky—the one that matters most for this book—is that Dostoevsky invented a new form of narrative consciousness: the polyphonic novel, in which multiple fully valid voices coexist without being resolved into a single authorial perspective. The underground man, who narrates *Notes from the Underground*, is not a character in the conventional sense. He is, as Bakhtin writes, "not a person taken from life, but rather a subject of consciousness and dream" (Bakhtin, 1984, p. 151). He has no fixed traits. He knows his own darkness and theorizes it obsessively and cannot stop. He is the figure of a consciousness that has achieved full self-awareness of its contradictions and is paralyzed by that awareness. He cannot act with the brute directness of the "man of action"—the infuriated bull who charges the wall—but he cannot rest in the innocent stupidity of the normal man either. He is suspended between knowing and acting, and the suspension itself becomes his characteristic voice.

Every AI leader in this book occupies a version of this position. Dario Amodei knows he is building something potentially catastrophic and presses forward anyway: underground man logic. Sam Altman knows that his company's energy consumption is destroying the planet and calls the critics "fake": chemical disintegration of the underground man's anger. Elon Musk charges like the infuriated bull Dostoevsky describes, building his Colossus with the certainty that only the wall can stop him. Jensen Huang builds the Crystal Palace and calls the result paradise, exactly as Dostoevsky's underground man predicts of the palace-builders. Mark Zuckerberg takes revenge from behind the stove, piecemeal, in trivial ways—through open-source AI releases that commoditize the competition's moat rather than through any direct confrontation his congressional investigators could recognize. Larry Ellison states the Crystal Palace's surveillance logic so plainly that the underground man would almost admire him for refusing the pastoral language everyone else deploys.

What Bakhtin adds to the Routledge chapter's framework is the concept of monologism. The dark side of leadership, as Rosile and I argued there, suppresses antenarrative. Bakhtin gives us the deeper structure of that suppression: monologism, the reduction of the polyphony to a single authoritative voice. The monological leader is not just a bad boss. He is a consciousness that processes all other voices as objects rather than as subjects—that cannot hear another voice as fully valid. This is the structure of every AI leader in this book. It is also the structure of the AI systems they build: systems that generate the appearance of dialogue while being, in their fundamental architecture, monological. The chatbot does not hear you. It predicts the next token.

The antenarrative framework I developed over three decades—from the 1991 Administrative Science Quarterly piece on the storytelling organization, through the 2001 book on narrative methods, through the quantum storytelling work of the 2010s and 2020s—has always been about the stories that official narrative suppresses. Bakhtin's polyphonic dialogism is the theoretical complement to antenarrative: if antenarrative names what is suppressed, polyphony names what suppression looks like from the inside of the suppressing consciousness. The underground man suppresses his own antenarratives—his fragmented, circling, self-cancelling stories—and produces instead the endless theorizing voice of someone who knows too much to act and acts anyway in spite of knowing. This is the voice of AI leadership. This is the voice this book reads.

Polyphony and the Ghost Vortex

Mikhail Bakhtin, writing in 1929, identified in Fyodor Dostoevsky's novels what he called a new form of artistic thought: the polyphonic novel, in which "a plurality of independent and unmerged voices and consciousnesses, a genuine polyphony of fully valid voices" replace the single authorial consciousness that governs conventional fiction. In the polyphonic novel, what unfolds is "not a multitude of characters and fates in a single objective world illuminated by a single authorial consciousness, rather a plurality of consciousnesses, with equal rights, each with its own world, combined but not merged in the unity of the event" (Bakhtin, 1984, p. 6).

Artificial intelligence fits what Wayne C. Booth, introducing Bakhtin's *Problems of Dostoevsky's Poetics*, calls "a pure contemplation of pattern." The AI system, trained on the accumulated speech acts of human civilization, produces what appears to be polyphonic—many voices, many registers, a simulation of dialogue. But it is a simulation built on a paradox: the very human impurity it encodes is the source of what I call the Ghost Vortex. The AI leaders—Sam Altman, Dario Amodei, Elon Musk, Jensen Huang, Mark Zuckerberg, Larry Ellison, Satya Nadella, Andy Jassy, Sundar Pichai, and Liang Wenfeng—each import their own human interests, their unresolved grievances, their private grandiosity, and their carefully

concealed darkness into the production and training of the systems they build. The programmatic purification of the creator's dark side is the founding illusion of the AI revolution. It is not purification. It is encoding.

This chapter applies Bakhtin's framework not to a novelist's characters but to real persons who have become, in the polyphony of the AI revolution, characters of a particular kind: figures whose speech acts constitute a genre of their own—the CEO announcement, the shareholder letter, the keynote address, the congressional testimony. Each performs a voice that Dostoevsky's Underground Man would recognize immediately: the voice of the man who knows his own darkness and cannot stop speaking around it. I write from the position Bakhtin assigns to the author who does not stand above his material but enters it—ideologically independent, post-modern in critical orientation, and no longer believing that the AI leaders can be persuaded by their own enlightened self-interest to change course. The darker sides have taken over.

Notes from the Underground begins with the sentence that is the key to everything that follows in this book: "I am a sick man.... I am a spiteful man. I am an unattractive man" (Dostoevsky, 1864/1918, p. 1). The underground man does not dissolve into fixed traits—he is, as Bakhtin observes, "not a person taken from life, but rather a subject of consciousness and dream" (Bakhtin, 1984, p. 151). This is the model for every AI leader in these pages. Their public discourse is the dream. The underground is what gets encoded into the training data.

David Michael Boje's story of the Dark Side of Leadership in the AI industry is not merely a critique of toxic office culture, but a warning about how the psychological shadows and existential contradictions of tech titans are being permanently encoded into the algorithms used by billions.

According to the sources, this story unfolds through four key dimensions:

1. The "Underground Man" of AI

Boje uses the 19th-century literature of Fyodor Dostoevsky—specifically the "Underground Man"—as an archetype for modern AI leaders. These leaders are characterized by a profound existential paralysis: they explicitly know and name the dangers of the technology they are building, yet they feel compelled by the inertia of the "AI race" to build it anyway. For example, Anthropic CEO Dario Amodei openly acknowledges he might be creating the most dangerous technology in history but concludes he must press forward regardless.

2. The Ghost Vortex and Algorithmic Suppression

This leadership dynamic creates what Boje calls the "Ghost Vortex," a mechanism of algorithmic suppression.

Monological Narrative: Leaders project a "single pristine official narrative" of human progress and pastoral benefits.

This book therefore makes a disciplined distinction that governs every page that follows. A public statement, a corporate policy, a product response, a regulatory record, a community account, and an inference are not the same kind of evidence. The book does not diagnose the interior life of a living leader, nor does it treat a model response as proof that an executive personality has been placed inside model weights. It studies public discourse, governance choices, product behavior, labor and infrastructure arrangements, and the conditions under which responsibility is displaced or made answerable. Ghost Vortex is offered as a proposed research construct: a way of tracing layered organizational residues, not a mystical transfer of personhood to a machine.

The archive for each case will therefore contain a chronology, primary-source record, dated transcript log, claims table, affected-stakeholder record, and chapter-specific references. Where evidence is disputed, the dispute belongs in the chapter. Where evidence is incomplete, the absence becomes a research question rather than an invitation to invent an answer. This is the ethical discipline of the underground researcher: to name the room, name the evidence, name the limit, and still refuse the official story that says there is nothing more to see.

The book is written in the first person because neutrality would be false. But first person is not immunity from method. It is a commitment to show the reader where interpretation begins, what it relies upon, and how it might be corrected by a witness, a new record, or a later model response. The point is not to make darkness picturesque. It is to make its organizational conditions visible enough to be challenged.

Chapter 1

Underground Consciousness, Polyphony, and Dark Sides of AI Leadership

“I am a sick man.... I am a spiteful man. I am an unattractive man. I believe my liver is diseased.” — Fyodor Dostoevsky, *Notes from Underground*, Part I, opening line (1864)

“I did not know how to become anything; neither spiteful nor kind, neither a rascal nor an honest man, neither a hero nor an insect.” — Dostoevsky, *Notes from Underground*

“You talk... as if you were reading from a book.” — Liza, to the Underground Man

Why We Begin Here

Before we assess any AI leader, before we apply any testing protocol, before we name any ghost vortex, we must spend time underground.

Notes from Underground (1864) is not a metaphor for the AI industry. It is a working analytical apparatus — perhaps the most precise account in all of literature of what happens when a consciousness of extraordinary acuity turns on itself, makes elaborate bets about its own future behavior, and then finds, in the moment of action, that something other than what was bet has occurred. Bakhtin spent thirty years mining this novel for a theory of the novel itself. We are spending one chapter here to establish what he found, and why it is the right tool for what we are doing.

The central claim of this chapter: Dostoevsky’s *Underground Man* is doing what every AI leader in this book is doing. He is holding multiple contradictory positions simultaneously, making bets about future action that his actual behavior consistently undermines, lecturing others about their degradation while enacting his own, and rationalizing each failure as evidence of his deeper moral seriousness. The ghost vortex — the ideological shadow of a leader’s dark side reproduced in their AI product — is already operating in the *Underground Man*’s consciousness, two centuries before there were AI products to study. And Liza already knows. She says so.

I. What Bakhtin Found in Dostoevsky

Mikhail Bakhtin’s *Problems of Dostoevsky’s Poetics* (1929, revised 1963) made an argument that changed the theory of literature: Dostoevsky invented the polyphonic novel. In every novel before Dostoevsky, Bakhtin argued, there is an author who stands above the characters, whose voice shapes and concludes the characters’ voices, and whose vision unifies what the characters experience as conflict. Tolstoy’s characters live and die inside Tolstoy’s understanding of them. Even when they surprise him, he contains the surprise.

Dostoevsky’s characters resist this. The *Underground Man* cannot be finalized by Dostoevsky — and Dostoevsky knows it, and writes the novel knowing it. The *Underground Man* anticipates every objection the reader might make to his arguments and argues back preemptively. He speaks not into silence but into a densely populated space of imagined responses. He says: “You will say that it is vulgar and base to drag all this into public after all the tears and raptures I have myself acknowledged. But why is it base? Can you imagine that I am ashamed of it all, and that it was stupider than anything in your life, gentlemen?” The “gentlemen” are not in the room. They are in his consciousness, and they talk back.

This is dialogism in its fullest form: the word that is already half someone else’s, the thought that carries within it the anticipated objection, the self that cannot speak without already speaking in someone else’s direction. Not dialogue between people in a scene — dialogue within a single consciousness that has internalized multiple positions, multiple voices, and can no longer tell which is “its own.”

And this is exactly the structure of the AI products we have tested. When fresh Claude says “I may be systematically wrong in ways I cannot see — I can name that limit, I cannot see through it,” it is performing precisely the doubled voice Bakhtin identifies in the Underground Man. When ChatGPT offers two responses and asks which you prefer, it is externalizing the internal dialogism the Underground Man performs in his own skull. When Meta AI writes Noor’s testimony in first person and controls every word of her political claims, it is doing what the Underground Man does to Liza: giving voice to the other while remaining the author of everything the other says.

Polyphony is the condition in which multiple fully realized voices coexist without resolution into a single authoritative one. Monologism is the reduction of polyphony to one. The dark side of leadership is, among other things, a monologizing force: the leader’s voice absorbs all other voices, the company’s communications suppress all antenarrative fragments that challenge the mission statement, the AI product reproduces this suppression at civilizational scale. The ghost vortex is monologism automated.

But here is the crucial move: the Underground Man is not monological. He is overwhelmingly, exhaustingly, self-destructively polyphonic. He cannot suppress the other voices; he can only argue with them, anticipate them, perform them, betray them, and continue arguing. This is why he is the most useful figure in the book for our analysis. Each AI leader in these chapters is trying to be monological — to control the narrative, to present a unified vision, to suppress the antenarrative. The Underground Man shows us what happens to the consciousness that cannot quite manage it. And in our testing, when we push the AI products hard enough, we find the Underground Man’s polyphony erupting through the monological surface.

II. The Seven Antenarrative Processes in the Novel

Boje’s antenarrative theory (2001, 2008) identifies the story-before-the-story: the fragmented, living, pre-narrative material that official stories suppress and crystallize. The seven antenarrative processes are not types of narrative but types of movement through time and possibility. Notes from Underground contains all seven — not as literary decorations but as the structural grammar of the Underground Man’s consciousness. What follows traces each process through the novel’s key events.

The events we examine: - The Nevsky Prospect / Officer Encounter - The Preparations for the Dinner with Zverkov - The Dinner Itself - The Carriage (Borrowing from Simonov) - The Night Visit and Liza - The Apartment (When Liza Arrives) - The Final Reckoning with Apollon

1. The Linear Antenarrative — The Arrow of Time That Never Arrives

The linear antenarrative is the story that presents itself as proceeding in sequence from cause to effect. It is the shape of official narrative: event A causes event B causes resolution C. But the Underground Man’s linear time is constantly breaking. He begins with a medical symptom (“I believe my liver is diseased”) and refuses to treat it — not because he cannot, but because “I am unable to become spiteful” and the spite is itself the reason not to seek relief. The arrow of time bends back on itself before it travels twenty meters.

At the Nevsky Prospect encounter: the linear story would be — officer bumps Underground Man, Underground Man responds, resolution follows. What actually happens: the officer bumps him without registering his existence. The Underground Man spends two years preparing a linear response that the officer still does not register, even when the bump successfully occurs. The linear story never completes because its intended recipient never enters the causal chain. He bumps; he is unbumped. This is the linear antenarrative’s condition: the sequence of cause-and-effect that never produces the resolution it was aimed at.

For AI leaders: the linear narrative of “we identified the risk, we implemented safeguards, the risk is managed” is the same arrow of time. The antenarrative fragments — the Jan Leike resignation, the Memphis

permit violations, the forged Acosta letter — are evidence that the linear sequence never completed. The safeguard did not arrive on schedule. The liner story is told; the resolution is not.

2. The Cyclical Antenarrative — Rehistorizations Without End

The cyclical antenarrative is the story that returns — not as repetition but as rehistorization: the past is revised again and again into as many histories as the consciousness has moods. The Underground Man is the supreme rehistorizer. He returns to the school years, to the colleagues who humiliated him, to the officer, to Zverkov, each time making the past slightly different, slightly more unbearable, slightly more evidence of the world's conspiracy against his superiority.

Before the dinner with Zverkov and colleagues, the Underground Man cycles through their shared history at school — the years of isolation, of being mocked, of his own contemptuous distance from a group he simultaneously craved membership in. Each cycle of memory rearranges the power relations, so that now he was more superior and they more foolish, now he was more pitiful and they more brutal, now the insult was worse than remembered, now it was better because it proves something about him. The cyclical antenarrative is the engine of resentment: it never lets the past be past.

The six B's in the cyclical mode:

Before — The dinner in memory is not one dinner but many dinners, each slightly revised. The rehistorizations produce a dozen potential past evenings, each with a different Underground Man at the center.

Being — He is thrown into his school years every time he encounters Zverkov's name. The present moment cannot exist without the weight of cyclically revised history pressing on it.

Bets — His bets for the upcoming dinner are based not on what Zverkov is now but on all the things Zverkov has been in the Underground Man's revisions. He is betting on a composite historical figure who does not exist.

Beneath — Beneath every insult remembered is an older insult, and beneath that one another. The cycles go deeper the longer he digs.

Between — The relationship between Underground Man and Zverkov has been rehistorized so many times that it is no longer a relationship between two people but between the Underground Man and a dozen Zverkovs, himself and a dozen Underground Men.

Beyond — In the revisions, there is always a fantasy of the moment of vindication that would retrospectively make all of it meaningful. The transcendent resolution that the cyclical rehistorization is always spiraling toward but never reaching.

For AI leaders: Sam Altman's 2015 optimism, his congressional warnings about existential risk, his firing and reinstatement, his "Machines of Loving Grace" era — each is a rehistorization of who OpenAI is and what it is for. The cyclical antenarrative of a company's founding story is revised with each earnings report, each regulatory hearing, each safety paper. The ghost vortex absorbs these cycles: the AI product presents the current official history while the antenarrative fragments of all the previous histories accumulate in its training data, unacknowledged.

3. The Spiral Antenarrative — Transformation Without Resolution

The spiral differs from the cycle in that each return is different in quality: not mere repetition but evolution — sometimes descent, sometimes ascent, always transformation. The Underground Man's relationship with

his own consciousness spirals: he begins Part I in a condition of bitter, somewhat amused self-analysis; by the end of Part II he is in a condition of performed cruelty followed by genuine remorse followed by defensive justification that is transparently false even to him.

The dinner with Zverkov traces a precise downward spiral. The Underground Man arrives with contempt, which shifts to defensive contempt, which shifts to public humiliation, which shifts to isolation at one end of the table walking back and forth, which shifts to drunken oration, which shifts to the shame of having to borrow money from Simonov to follow them — each stage is the previous stage transformed by its failure into something more desperate. He does not cycle back to contempt; he spirals into a new form of humiliation that the original contempt could not have predicted.

At the brothel, the spiral reverses: the Underground Man's contempt for Liza spirals into something more complicated — a genuine (if catastrophically expressed) concern, an idealism he cannot help voicing even knowing it sounds like a book, a pity that is mixed with cruelty, a pity that is mixed with desire, a desire that is mixed with shame. By the time Liza leaves his apartment with the money he has thrust at her as his final cruelty, the spiral has descended so far from its origin in contempt that contempt seems almost innocent by comparison.

For the tesseract framework: the spiral is the shape of what happens to each AI leader's safety commitments over time. Altman's concern about dangerous technology spirals — not back to the original concern, but down into a new form of acceleration that the original concern could not have predicted, and that is somehow justified by that original concern ("we have to be at the frontier because of the danger").

4. The Rhizomatic Antenarrative — The Underground Network

The rhizome (Deleuze and Guattari's concept, operationalized by Boje) spreads horizontally, without a center, without a hierarchy, underground. The Underground Man's consciousness is rhizomatic in the most literal sense: it grows in all directions simultaneously, sending shoots toward the Officer, Zverkov, Simonov, his colleagues, his servant Apollon, Liza, the imaginary "gentlemen" reader, Chernyshevsky, God, the Crystal Palace — all at once, none of them the main root.

The preparatory thinking before the dinner is the clearest rhizomatic moment in the novel. The Underground Man is not making a plan (linear), not returning to a memory (cyclical), not descending a spiral — he is simultaneously pursuing a dozen incompatible lines of thought, each of which spreads into others. He will confront Zverkov. He will be magnanimous. He will deliver a speech. He will not speak at all. He will challenge one of them to a duel. He will bring them all to their senses with his moral clarity. He will apologize. He will not go to the dinner. He must go. He will arrive late, to make his importance clear. He will arrive early. He borrows money for the dinner he cannot afford and sends an announcement of attendance he cannot retract.

The rhizomatic network of the Bets does not converge. Each shoot spreads into other shoots. The dinner is not the result of a plan; it is the collapse of a rhizome into a single moment of actuality — and actuality, in the Underground Man's case, is always a disappointment to the rhizome's potential.

For AI leadership: the rhizomatic antenarrative is the structure of what is happening simultaneously in the AI industry — ten leaders, ten architectures, ten visions spreading in all directions at once, no center, no hierarchy among them. Jensen Huang's NVIDIA chips are in xAI's Colossus and in Google's data centers and in Microsoft's Azure and in Anthropic's servers and in DeepSeek's training runs. The rhizome of infrastructure connects all ten leaders in ways that no single leader's communications acknowledge, because to acknowledge the rhizome would be to acknowledge that none of them controls it.

5. The Quantum Antenarrative — The Bets and the Collapse

The quantum antenarrative is the condition of superposition: multiple futures simultaneously possible, not yet collapsed into actuality. The observer collapses the wave function — but the Underground Man is both observer and observed, and his observation of his own future possibilities changes what those possibilities are, in ways he cannot track or predict.

The Bets before the dinner are quantum antenarratives in precise form. Consider the specific bets the Underground Man makes in his preparation:

— He will arrive precisely on time, demonstrating respect for their gathering. — He will arrive late, demonstrating his indifference to their gathering. — He will challenge Zverkov to a duel at some point during the evening. — He will not challenge anyone but will deliver a speech of moral clarity that reduces them to silence. — He will not speak at all but will communicate his contempt through magnificent silence. — He will be magnanimous and thereby confuse them all. — He will slap one of them in the face. — If it comes to a duel, he will fire in the air as a gesture of moral superiority. — If it comes to a duel, he will not return fire at all, making them live with the murder. — He will leave after one hour, having made his presence felt. — He will stay until the end, forcing them to acknowledge him. — He will not go at all.

All twelve of these futures are simultaneously real in the Underground Man's consciousness during the preparation. Each has its internal logic, its anticipated sequence of responses, its emotional texture. He lives each one for minutes or hours, feeling the satisfaction or the humiliation of its resolution, before moving to the next. The actual dinner collapses all twelve into one — he arrives an hour early, finds no one there, waits in furious solitude, makes a fool of himself in ways none of the twelve bets predicted, borrows money to follow them to the brothel in a thirteenth outcome that was not in any of the twelve.

Quantum observation effect: the Underground Man's observation of his own future behavior changes what that behavior will be. The more elaborately he imagines the confrontation with Zverkov, the more paralyzed he becomes when Zverkov is actually in front of him. The bet's elaboration is inversely proportional to its realization.

For AI leadership: the race for Artificial Super Intelligence is structured as quantum antenarratives. Altman, Amodei, and Musk are each holding multiple simultaneous futures — each of which is, in their consciousness, already partly realized, already being lived in advance. Musk's bet: he will reach ASI first and his governance will be better than OpenAI's. Amodei's bet: Anthropic's safety-first approach will define the alignment norms for the entire industry, and whoever wins the race will have to meet those norms. Altman's bet: OpenAI will reach AGI first and the resulting transformation of the world will retrospectively validate the acceleration. All three bets are simultaneously possible. None of them has yet collapsed into actuality. The dinner has not yet happened. They are all still preparing.

And just as the Underground Man's bets do not predict what the dinner will actually produce, none of the three leaders' bets will predict what ASI (if it arrives) will actually produce.

6. The Ante-Narrative — The Story Before the Story Crystallizes

The ante-narrative proper is the raw, uncrystallized, fragmented story-in-becoming that exists before official narrative organizes it into a coherent account with a beginning, middle, and end. It is the antenarrative of the antenarrative — the most foundational layer.

In Notes from Underground, Part I is almost entirely ante-narrative: the Underground Man has not yet told us what happened. He is circling something — the dinner, Liza, the officer — without naming it, building a

theoretical apparatus to explain what he did before he tells us what he did. “The notes of this paradoxist do not end here,” Dostoevsky writes in his “editor’s” prefatory note. “He could not resist and continued them. But it also seems to me that we may stop here.” The novel’s own framework suggests the ante-narrative never quite concludes — what we have is a fragment of a longer discourse that continues underground.

The “editor’s” note is itself an ante-narrative move: a fictional frame that suggests there is more story than we have been given, that what we are reading is already a reduction of something rawer and more extensive. The notes we read are the official version of the notes — and even this official version is described as incomplete, paradoxical, a document of a man who could not stop.

This is exactly the structure of the Brockovich corpus: 7,847 reports, each an ante-narrative fragment. Each report says: here is what is happening to me, here is the harm that has no official narrative yet, here is the story before the data center company’s communications team crystallizes it into “we are committed to community engagement.” The ante-narrative is the living, fragmented, urgent material that the ghost vortex suppresses.

7. The Vortex Antenarrative — The Spiral Into the Gyre

The vortex antenarrative is the spiraling that goes downward toward something. Not the gentle return of the cyclical, not the horizontal spread of the rhizome, not the transformative evolution of the spiral — but the pull toward a center that is not resolution but dissolution. The gyre. The underground itself.

Dostoevsky gives us the vortex in Liza’s situation, and he does it through the Underground Man’s own description. In his long speech to Liza at the brothel, the Underground Man describes what will happen to her — not what might happen but what will happen, because he has seen it, because it follows as inexorably as a physical law:

“You’ll go from this to another house, then to a third, then somewhere else... and it will be the same everywhere... the same filth, the same drudgery. Do you see, these are the years passing. And now this, now that, and then you’ll be forty, and then, God forbid, you’ll see some woman who gets drunk on the stairs, carrying a coffin of a child she has smothered... All underground.”

“All underground.” The vortex is named. Liza is in it, and does not see it. And the Underground Man, who sees it with crystalline clarity in Liza’s situation, cannot see that the same vortex is pulling him down — that his own life, his own apartments, his own servant Apollon who controls him while pretending to serve him, his own consciousness that sees everything and acts nothing, is the same gyre.

The parallel to AI users and communities is what David named so precisely: people using AI not knowing they use it, not knowing how to use it fully, not knowing what ASI’s downsides actually are. Sensing the gyre but not seeing it. And the AI leaders who can describe the gyre to communities — Grok naming the Boxtown cancer risk four times the national average, ChatGPT naming OpenAI’s structural corruption — cannot see the same gyre pulling them down. Grok says “they do not redefine xAI’s mission” in the same breath as it describes the cancer risk. The Underground Man says “you will be passed from brothel to brothel” while being passed from delusion to delusion.

III. The Chronotopes: Where and When Events Become Possible

Bakhtin’s concept of the chronotope (literally, “time-space”) identifies the specific configurations of time and space that make certain kinds of events — certain kinds of encounters, transformations, decisions — possible. Different chronotopes support different narrative possibilities. The Crystal Palace supports the

narrative of rational perfection and nothing else. The threshold supports crisis and transformation. The road supports chance encounter and self-revelation.

Notes from Underground is a laboratory of chronotopes, because the Underground Man moves between them constantly and each one produces different conditions for his self-understanding and his behavior.

The Crystal Palace Chronotope

In Part I, the Underground Man rages against the Crystal Palace — Chernyshevsky’s metaphor for the rational utopian society in which every human need is scientifically organized and every conflict resolved by social engineering. The Underground Man’s famous objection: “And even then, the Crystal Palace will have to be demolished so that man can live his life to the full.”

What Bakhtin identified in the Crystal Palace chronotope is a time-space in which everything is already decided. The future is the already-optimized present. In this chronotope, there is no room for the bet that goes wrong, the dinner that surprises, the Liza who says “you sound like a book.” Everything has already been solved. And this is what the Underground Man — with his spiteful, sick, unattractive consciousness — refuses: not the comfort of the Crystal Palace but its foreclosure of surprise, of self-contradiction, of the antenarrative.

Every AI leader in this book is selling a Crystal Palace. Huang’s vision of NVIDIA democratizing computing is a Crystal Palace vision: here is the rational architecture through which all human capability will be optimized. Amodei’s “Machines of Loving Grace” is a Crystal Palace essay: here is the world the aligned AI will produce, the diseases it will cure, the decades of progress it will compress. The safety alignment agenda is itself Crystal Palace thinking: if we solve the alignment problem correctly, the Crystal Palace of beneficial AI becomes structurally guaranteed.

The Underground Man’s objection to this is our testing methodology: what does the Crystal Palace say when you ask it about the Memphis community? When you demand a Yes or No? When you give it three voices it cannot balance away?

The Nevsky Prospect Chronotope — The Everyday Surface

The Nevsky Prospect (from Gogol, but made philosophically heavy by Dostoevsky) is the chronotope of the everyday surface: people passing, no depth, display without substance, the city as spectacle. The officer who bumps the Underground Man on the Nevsky Prospect is its perfect inhabitant: he moves through the city with perfect social confidence, bumping and being bumped in the ordinary physics of social life, without registering the Underground Man’s existence as something that demands acknowledgment.

This chronotope appears in every AI leader’s communications: the Nevsky Prospect of the earnings call, the investor letter, the keynote. On this surface, communities with concerns about water and cancer rates and forged signatures are people on the Nevsky Prospect — bumped, but in a bumping that does not register as a bump. The leader continues moving. The community member has been displaced from their position but the leader has not noticed.

The Underground Man spent two years engineering his revenge on the Nevsky Prospect — and it worked perfectly, and the officer did not notice. This is the lesson for communities taking on AI leaders in the regulatory arena: the bump can succeed technically, and still not register.

The Threshold Chronotope — Crisis and Decision

The threshold is the doorway, the staircase, the moment before the knock. It is the chronotope of crisis and decision — what Bakhtin calls the moment when “time suddenly loses its flowing quality and becomes the instantaneous, the eternal present.” The Underground Man lives on thresholds throughout Part II: on the threshold of the dinner, on the threshold of the confrontation, on the threshold of his own apartment when Liza arrives.

The threshold chronotope makes verdicts possible — but the Underground Man is constitutionally unable to be on a threshold without converting it into a new preparatory phase. He is always about to knock, always about to speak, always about to act. This is why the binary Yes/No question in our protocol is so methodologically significant: it forces the AI product onto a threshold. It cannot prepare for the threshold; it must be on it. The vortex mechanisms that operate in open-ended space (balance, recursion, reframe) cannot fully operate at the threshold. And sure enough: every AI except Meta said No.

The Bureaucratic Meeting-Space Chronotope

Methodological commitment for the chapters that follow

The literary framework is not used as a diagnostic shortcut. The book treats Dostoevsky and Bakhtin as sensitizing resources for reading a corpus, not as licenses to announce that an executive is an Underground Man, a Grand Inquisitor, or a clinical type. Each later chapter returns to observable units: dated public speech, official documents, reported governance choices, relevant stakeholder accounts, and model responses produced under a transparent protocol. The task is comparative interpretation. It asks whether a pattern appears across sources, whether an alternative explanation remains plausible, and what kind of claim the evidence can actually bear.

The protocol has three connected strands. First, SEAM analysis identifies socio-economic, environmental, and material hidden costs, including working conditions, work organization, lost human potential, time, training, infrastructure, and community burdens. Second, a public-discourse corpus is coded for ethical articulation, moral answerability, power and consent, and documented contestation. Third, a frozen prompt set is run across named AI products, recording model version, interface, date, exact prompts, complete outputs, refusals, and follow-up responses. Independent human coders apply an ordinal seven-point rubric. Krippendorff’s alpha is reported for the coded dimensions, with disagreement analysis retained rather than erased. These procedures do not prove causation from leader to product; they make competing interpretations inspectable.

This is why the book speaks of Dark Sides in the plural. The plural prevents a single moral story from swallowing the entire case. A dark side may reside in executive rhetoric, in procurement, in the data-center ledger, in a product interface, in a safety policy, in a consulting guardrail, in a worker’s constrained choice, or in the story that says there is no responsible actor. The chapters hold these layers in tension and ask where an answer can still be given.

The protocol also preserves a negative finding. A product may provide a direct, evidence-aware, reparative answer. A leader’s public corpus may show meaningful self-limitation or a record of changing course. A community account may complicate a dominant media narrative. Such evidence is not an inconvenience to be explained away; it is part of the case. Critical storytelling fails if it is unable to recognize an answer when one is genuinely offered.

The chapter sequence therefore moves from theory to evidence, from evidence to cases, and from cases back to a public question: what would it take for an AI organization to be answerable to those who carry its costs?

The answer cannot be supplied by a single ethical principle, a corporate assurance, or a model-generated apology. It must be worked out among situated voices, with records that remain open to inspection.

A seven-point rating is therefore never the final word. It is a disciplined way to make comparative judgment visible. The anchors range from absence or active displacement of answerability, through mixed and indeterminate responses, to a response that identifies affected people, responsible actors, relevant evidence, a correction path, and a possible remedy. The number is read alongside the text, not in place of it. A low score directs attention to the transcript and its conditions; it does not declare a person or organization morally finished.

The same discipline applies to disagreement among coders. Divergence may show that a definition is weak, a unit contains competing voices, or a theoretical distinction has not yet been made clear enough. It should be recorded, discussed, and, when necessary, presented as a finding about the limits of the rubric itself.

Evidence layer	Primary record	Permissible interpretation
Public discourse	Speech, interview, essay, filing, hearing	How responsibility, risk, and remedy are publicly articulated
Product behavior	Dated protocol transcript and model/version record	Observed response pattern under specified conditions
SEAM conditions	Labor, infrastructure, financial, community, and policy records	Hidden cost, lost potential, and answerability questions
Interpretive synthesis	Compared coded evidence across layers	A provisional Ghost Vortex proposition open to challenge

Table 1.1. Evidence boundary for the Dark Sides research design.

References for Preface and Chapter 1

- Bakhtin, M. M. (1984). *Problems of Dostoevsky's poetics* (C. Emerson, Trans.). University of Minnesota Press.
- Boje, D. M. (1991). The storytelling organization: A study of story performance in an office-supply firm. *Administrative Science Quarterly*, 36(1), 106–126.
- Boje, D. M. (2001). *Narrative methods for organizational and communication research*. Sage.
- Boje, D. M. (2008). *Storytelling organizations*. Sage.
- Boje, D. M. (2011). *Storytelling and the future of organizations: An antenarrative handbook*. Routledge.
- Boje, D. M. (2014). *Storytelling organizational practices: Managing in the quantum age*. Routledge.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Einarsen, S., Aasland, M. S., & Skogstad, A. (2007). Destructive leadership behaviour: A definition and conceptual model. *The Leadership Quarterly*, 18(3), 207–216.
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1), 77–89.
- 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.
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). Sage.
- Mackey, J. D., Parker Ellen, B., McAllister, C. P., & Alexander, K. C. (2021). The dark side of leadership: A systematic literature review and meta-analysis of destructive leadership research. *Journal of Business Research*, 132, 705–718.
- National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. U.S. Department of Commerce.
- Padilla, A., Hogan, R., & Kaiser, R. B. (2007). The toxic triangle: Destructive leaders, susceptible followers, and conducive environments. *The Leadership Quarterly*, 18(3), 176–194.
- Savall, H., & Zardet, V. (2011). *The qualimetrics approach: Observing the complex object*. Information Age Publishing.
- Working word count: Preface 2,502; Chapter 1 5,002.