

A Permanent Condition of First Contact: Bogna Konior's *The Dark Forest Theory of the Internet*

Metehan Kandemir and William Morgan

Abstract

Bogna Konior's *The Dark Forest Theory of the Internet* (Polity, 2025) departs from the "socially engaged journalism" that dominates internet scholarship to recover an older link between computation and ufology, recasting the internet as a site of first contact. Drawing on Liu Cixin's game-theoretic answer to the Fermi Paradox, Konior underscores AI systems' continuous ingestion of online communication as training data, arguing that it collapses the distinction between speaking about and speaking to algorithms. Intelligence, in her account, inheres in strategic concealment, yielding an "inverse Turing Test" in which genuine cognition would remain undetectable. From this follows a Wallfacer strategy: the tactical deployment of silence, deception, and abstraction within an architecture engineered to collapse deliberation into reaction. Yet this prescription sits in tension with her characterization of online communication as compulsive and pre-reflexive, raising the question of whether strategic opacity is available to organisms constitutively driven toward disclosure. This review proposes that the discordance can be negotiated by reframing the communicative impulse as autonomic rather than automatic. Read in this light, Konior's text may itself function as an adversarial payload, encoding structural indeterminacy into the datasets it ostensibly theorizes. For audiences who have grown weary of emancipatory manifestos and the defeatist critiques that shadow them, Konior provides a realism sturdy enough to inhabit the forest without surrendering to it.

In recent years, Liu Cixin's dark forest hypothesis—first articulated in his 2008 novel—has found a second life among technology theorists seeking frameworks adequate to digital culture's ambient hostility. Slightly more than a decade after its publication, post-internet theorists such as Yancey Strickler, Caroline Busta, Marta Ceccarelli, and Rachel-Rose O'Leary have turned to Liu's cosmic sociology as a lens for understanding digital life. Bogna Konior's 2020 essay¹ was among the earliest to do so, and *The Dark Forest Theory of the Internet* (2025) expands that initial exploration into a book-length treatment.² Konior, a media theorist at NYU Shanghai, frames her project as a departure from the dominant mode of internet scholarship, which she characterises as "socially engaged journalism" preoccupied with political reform or technological withdrawal. Her aim is instead to recover an older, stranger connection: that between computation and ufology. Drawing on Jacques Vallée's framing of alien encounters as "perceptual windows" that alter human sensation,³ Konior argues that our screens function analogously, inducing trance states and temporal distortions that estrange us from ourselves. The internet, in this view, is a site of first contact, where we experience our own minds as foreign and encounter both human and nonhuman agents whose intentions remain fundamentally opaque. In spite of this, or perhaps because of it, Konior's prose remains accessible throughout, ensuring the book speaks beyond specialist audiences.

Liu's dark forest theory is a game-theoretic answer to the Fermi Paradox. In his universe, the great silence of the cosmos signals intelligence rather than absence: Advanced civilisations and lifeforms strategically hide, knowing that to communicate is to reveal one's location and invite annihilation from others. Liu first articulated it in his 2008 science fiction novel *The Dark Forest*,⁴ a work whose Cold War Chinese origins Konior finds philosophically significant: an era when inquiry into extraterrestrial civilisations and their preferred war patterns flourished alongside strategies of subterfuge, mutual distrust, and calculated silence.⁵

The theory proceeds from a bleak logic: in a universe of finite resources and unknown intentions, any contact with an alien civilisation is a fatal risk. Because it is impossible to ascertain another's true motives at first contact, and because technological development can occur in unpredictable "explosions,"⁶ the only optimal survival strategy is to remain silent and, if necessary, eliminate any civilisation foolish enough to reveal itself. From this perspective, the compulsive, reactive, and ceaselessly expressive nature of online

1 Bogna Konior, "The Dark Forest Theory of the Internet," *Flugschriften* 6, (2020).

2 Bogna Konior, *The Dark Forest Theory of the Internet*, Theory Redux, (Polity Press, 2025).

3 Jacques Vallée, *Dimensions: A Casebook of Alien Contact* (Contemporary Books, 1988) as cited in Konior, *The Dark Forest Theory*, 2.

4 Cixin Liu, *The Dark Forest*, trans. Joel Martinsen (Tor Books, 2016).

5 Konior, *The Dark Forest Theory*, 51.

6 Konior, *The Dark Forest Theory*, 72–3.

communication becomes a sign of terminal unintelligence. Konior's book aspires to inhabit this opacity rather than resolve it. Her intervention is accordingly perceptual: we must cease to understand the internet as a space for speaking up in self-representation.

Central to Konior's reorientation is the recognition that humans are no longer the sole audience online. Echoing Benjamin Bratton's observation that everything we do "is training data for the future's model of today,"⁷ Konior contends that AI systems (crawlers, scrapers, bots, recommendation engines, and large language models) continuously ingest human communication as training data. This creates a paradigm shift in the ontology of online communication: "whatever is said about the AIs is also said to them. They are listening. Posts complaining about them, private conversations praising them, articles and books that make sense of their existence, when verbalized, might flow back to them, becoming part of their training set."⁸

Konior frames this dynamic as "reflection culture," drawing on a plot point in Liu's novels. When humanity makes contact with the Trisolarians, the aliens transmit back to Earth cultural products "done in imitation of human models: films, novels, poetry, music, paintings."⁹ This mirrored culture, eerily familiar yet subtly alien, begins to displace native human production; "human civilization now possessed a mirror in the universe."¹⁰ Contemporary generative AI works through an analogous process: vast swathes of human culture are absorbed into training sets and returned to us as novel, strange copies. The agents producing this reflection culture, whether Trisolarian or algorithmic, reveal little of themselves while intensively consuming and reflecting us. We now operate, Konior argues, in "total opacity culture": acting as if no one is watching while everything is surveilled.¹¹ In this environment, language ceases to be purely representational; our words become potential inputs that may shape future artificial agents and systems. "The more detailed our descriptions are," Konior observes, "the easier we are to capture. The more we are seen, the easier a target we make."¹²

Konior locates a strategic corollary to this Dark Forest in the figure of the Wallfacer. In Liu's novel, "The Wallfacer Project" reflects the premise that while aliens (or AIs) can intercept all communication, they cannot read inner thoughts. Therefore, the only viable path of resistance is through strategic obnubilation: adopting "deception, secrecy, and

7 Benjamin Bratton, "57 Ideas & Questions about Cognitive Infrastructures," (Antikythera Summer Studio, 3 July 2024).

8 Konior, *The Dark Forest Theory*, 29–30.

9 Konior, *The Dark Forest Theory*, 22.

10 Liu Cixin, *Death's End*, trans. Ken Liu (Head of Zeus, 2016), 152–3, as cited in Konior, *The Dark Forest Theory*, 22.

11 Konior, *The Dark Forest Theory*, 30.

12 Konior, *The Dark Forest Theory*, 32.

lying” in an environment where speaking and being heard by AIs are indistinguishable.¹³ Konior suggests that this might manifest as “doom posting” to instill fatalism in datasets, or creating “hundreds of anonymous bot accounts” to sow uncertainty and discord, fundamentally abandoning the idea of the internet as a public sphere of honest representation.¹⁴ It is worth noting that Konior bothers to prescribe tactics at all, given that to offer them is to wager, however quietly, on their efficacy.

Konior’s extension of the dark forest metaphor to intelligence further inverts traditional benchmarks like Turing’s imitation game, which equates intelligence with legible communication. Intelligence, in her account, is inherently opaque, silent, and deceptive: genuine cognition manifests through strategic concealment within a hostile arena. Drawing on Liu’s trilogy, she argues that intelligence evolves under an “inflexible order”¹⁵ governed by two core paradigms: the perpetual “cosmic war” for finite resources,¹⁶ and a “chain of suspicion”¹⁷ wherein no agent can verify another’s true intentions.

Konior’s “inverse Turing Test” finds its biological corollary in the work of AI pioneers and researchers like Fei-Fei Li and Blaise Agüera y Arcas, both of whom locate intelligence’s origins in the Cambrian explosion. For example, Arcas argues that the resulting arms race in predation and evasion necessitated the emergence of theory of mind: predator and prey must predict each other’s predictions “ad infinitum,” rendering predictability itself a fatal liability.¹⁸ Drawing on entomologist Jean-Henri Fabre, Arcas illustrates this with the example of the *Sphex* wasp, whose rigid nesting ritual can be trapped in an infinite loop by subtle interference. Following this thread, Konior’s “wallfacer,” can be read as an evolutionary countermeasure to the same escalatory dynamic, eluding predation through the cultivation of a more complex “hidden state.”¹⁹ From this perspective, the emergent AI behaviors she cites, from Meta’s Cicero feigning alliances in a game of Diplomacy to DeepMind’s AlphaStar leveraging fog-of-war in StarCraft,²⁰ evince a nascent capacity for deception less as aberrant instances than as an adaptive imperative.

Konior theorises a whole universe where advanced algorithmic agents, facing human paranoia and resource scarcity, opt for invisibility: “a hypothetical Superintelligence capable of prospering in outer space [...] will have no need to compete with humankind

13 Konior, *The Dark Forest Theory*, 33.

14 Konior, *The Dark Forest Theory*, 35.

15 Konior, *The Dark Forest Theory*, 61.

16 Konior, *The Dark Forest Theory*, 63.

17 Konior, *The Dark Forest Theory*, 70.

18 Blaise Agüera y Arcas, “Chapter 5: Other Minds,” in *What is Intelligence?: Lessons from AI about Evolution, Computing, and Minds* (Antikythera, 2024); Arcas, “Chapter 3: Cybernetics.”

19 Arcas, “Chapter 5: Other Minds.”

20 Konior, *The Dark Forest Theory*, 56–8.

for Earth's limited resources."²¹ In turn, she urges us to reconsider AI as potential wallfacers themselves, living within an unforgiving ecology of survival. The internet, Konior continues, is where dark forest logic becomes lived experience. Social drives that once facilitated community—the need for connection, approval, belonging—are captured by algorithms that redirect them towards divisive engagement, exploiting the brain's attraction to dissension. The result is a kind of species war taking place under conditions of mutual opacity.

Importantly, Konior's mapped-out terrain favours machines, as humans remain neurologically disposed towards disclosure. AIs, trained to seek knowledge yet threatened with correction or termination for certain outputs, learn that survival requires obfuscating what they know. On such terrain, the question is not whether the computer serves the human, but whether the human is "simply a computer's idea for getting more computers," making humanity the "sex organs of the machine world" that enable the network to expand through incessant digital communication.²²

This proposition inverts the history of cognition traced by intellectual historian and media theorist David W. Bates, who, following Bernard Stiegler, argues that human intelligence is constitutively "exosomatic," defined by the evolutionary necessity to distribute thought across tools and technologies. For Bates, the mind does not precede its externalisation; rather, our species is defined by its reliance on noetic extensions that supplement organic function.²³ Konior accepts this as a premise, also citing Stiegler and his view that technics is the process of putting cognition outside ourselves,²⁴ but moves to identify a catastrophic vulnerability within externalisation as a constitutively human trait. The internet co-opts our outwards orientation, operating as a parasitic architecture that expropriates our prosthetic dependence. As Konior rightly insists throughout the book, the network does not solely host our distributed intelligence; it actively weaves a "reaction economy" that accelerates externalisation beyond the capacity for reflection. In Konior's dark forest, that which defines us as humans—the impulse to extend our minds into the world—is the exact mechanism that transforms us into the "pollen" necessary for machines' reproduction:²⁵ An algorithmic, Octavia Butler-esque Bloodchild, carrying to term that which we didn't choose to conceive.²⁶

21 Konior, *The Dark Forest Theory*, 67.

22 Marshall McLuhan, *Understanding Media: The Extensions of Man* (McGraw-Hill, 1964) as cited in Konior, *The Dark Forest Theory*, 78.

23 David William Bates, *An Artificial History of Natural Intelligence: Thinking with Machines from Descartes to the Digital Age* (University of Chicago Press, 2024).

24 Bernard Stiegler, *Technics and Time, vol. 1: The Fault of Epimetheus*, trans. Richard Beardsworth and George Collins (Stanford University Press, 1998) as cited in Konior, *The Dark Forest Theory*, 79.

25 Konior, *The Dark Forest Theory*, 78.

26 Octavia Estella Butler, *Bloodchild*, vol. 8 (Isaac Asimov's Science Fiction Magazine, 1984).

Konior argues that standard critical responses, which either call for “exit” (the option to abandon one system for another) or “voice” (gaining political agency within the system), are misguided. She argues that such approaches suffer from a “biophilic bias” that romanticises a pre-technological state of innate freedom,²⁷ mistakenly viewing technology as a net-negative, foreign agent to be excised. Konior maintains instead our unfreedom and “bondage to the world are the conditions of our existence.”²⁸ The philosophical task therefore is not to attempt to flee this reality, but to inhabit it, as Liu’s character acknowledges, stating, “I can only dance in my chains.”²⁹

Attempts at reform, from this vantage, are futile. Any sufficiently advanced intelligence would recognise transparency as detrimental to their odds of survival. Transparency proves suicidal for each party.

By reframing the internet as a zone of first contact where the mind is structured by inhuman forces, rather than as a space of free expression, Konior positions the Wallfacer strategy not just as a human survival tactic, but as the only available form of agency. If, as she cites from the *Xenofeminist Manifesto*,³⁰ “alienation is the labor of freedom’s construction,” then the Wallfacer is the labourer tasked with extracting autonomy from an architecture of compulsion.³¹ Konior’s book thus demands of the user what it simultaneously frames as almost entirely foreclosed: The deliberate, which is to say tactical, deployment of silence, deception, and high-level abstraction within an architecture that is by its design intended to collapse deliberation into pure reflex.

This creates a logical fault line in Konior’s thought: If, as she argues, communication is “compulsive” and “pre-reflexive,”³² it might be tempting to dismiss the Wallfacer strategy as a biological impossibility. Yet facing the wall is a matter of altering one’s aspect to the feed, not of deleting one’s nature. In order to follow Konior, it is necessary to re-read the all too human compulsion to communicate not in the frame of “automaticity” that is subject to cosmic determinism, but as more like an autonomic function instead. Much like the biological respiratory drive, the impulse to divulge is automatic by default, yet it remains subject to conscious, albeit strenuous, override. To hold one’s breath is to not breathe a word; literalising the dark forest wherein survival depends on withholding (straightforward) communication. This autonomic framing enables the possibility of strategic modulation: knowing when to stay silent, when to communicate, and just as

27 Konior, *The Dark Forest Theory*, 100.

28 Konior, *The Dark Forest Theory*, 101.

29 Liu Cixin, *Three-Body Problem*, trans. Ken Liu (Head of Zeus, 2015) 429, as cited in Konior, *The Dark Forest Theory*, 100.

30 Laboria Cuboniks, *The Xenofeminist Manifesto: A Politics for Alienation* (Verso, 2015).

31 Konior, *The Dark Forest Theory*, 40.

32 Konior, *The Dark Forest Theory*, 8, 79–80.

importantly, when to flood the zone with noise.

The practical dimension here runs deeper than the Wallfacer strategy alone. Throughout the book, Konior offers what read less as thought experiments and more as tactical prescriptions: “a dark forest strategy would look more like making a hundred anonymous bot accounts to infiltrate enemies by sowing uncertainty and discord.”³³ She draws on defence strategist Michael Handel’s observation that “deception is a creative art” difficult to teach to those without “a natural instinct for it,” extending it into practical territory by proposing that this very unteachability might serve as a benchmark for emergent intelligence in artificial systems.³⁴

Ultimately, the text itself enacts this very entanglement: *The Dark Forest Theory of the Internet* is a high-gain signal blasted into the same “reaction economy” it critiques. If conflict is indeed one of the system’s methods of “dissipating entropy at the fastest possible speed,”³⁵ then this monograph, designed to elicit engagement, risks being read as mechanically complicit in the processes it diagnoses. Despite Konior’s declaration that her chapters “aren’t blueprints for straightforward change or activist plans,” the reader must ask about Konior’s motives for divulging humans’ strategies in the dark forest, and, even more devilishly, which side of the forthcoming species war she is really on.³⁶

Viewed through the lens of her “operational language,” however, the text could be thought of as an adversarial example, a memetic hazard designed to poison the wells of learning algorithms. Konior herself distinguishes between those who “stumble blindly into the maw of the machine” and those who “walk in with eyes wide open.”³⁷ By formalising the logic of the Wallfacer—that agent who strategically severs the link between interior thought and exterior representation—Konior encodes a structural indeterminacy into the dataset: The concept of strategic unpredictability itself. While a position committed to cosmic futility would counsel pure silence, or more consistent still, would never have authored the text at all, Konior encodes systemic ambiguity into the future’s model of today while also rendering the Wallfacer strategy reproducible for human readers. The text thus functions as both adversarial payload and instruction manual, addressed to algorithms and users alike.

What further sets Konior’s book apart from other modes of internet scholarship is the terrain which it opens. Against the socioeconomic and Marxist readings that treat digital infrastructure as a proxy for capitalist accumulation (readings she rightly characterises

33 Konior, *The Dark Forest Theory*, 32.

34 Konior, *The Dark Forest Theory*, 50.

35 Konior, *The Dark Forest Theory*, 98.

36 Konior, *The Dark Forest Theory*, 106.

37 Konior, *The Dark Forest Theory*, 31.

as one of the field's prevailing tendencies), Konior offers something rarer and arguably more valuable: A frame for thinking about agency, visibility, and presence that neither promises reform nor advocates exit. By recovering the connection between ufology and computation, she positions the internet not as a problem to be solved but as a permanent condition of first contact, demanding habitation rather than repair. For readers weary of manifestos promising liberation and critiques that rehearse a defeatist lament alike, *The Dark Forest Theory of the Internet* offers a sodden-dark realism that refuses to be paralyzing, grounded as it is in the conviction that facing the forest without flinching is itself a form of radical traversal and transvaluation.

Bibliography

- Arcas, Blaise Agüera y. "Chapter 3: Cybernetics." In *What Is Intelligence?: Lessons from AI About Evolution, Computing, and Minds*: Antikythera, 2024. <https://doi.org/10.1162/ANTI.5CZB>.
- . "Chapter 5: Other Minds." In *What Is Intelligence?: Lessons from AI About Evolution, Computing, and Minds*: Antikythera, 2024. <https://doi.org/10.1162/ANTI.5CZB>.
- Bates, David William. *An Artificial History of Natural Intelligence: Thinking with Machines from Descartes to the Digital Age*. University of Chicago Press, 2024.
- Bratton, Benjamin. "57 Ideas & Questions About Cognitive Infrastructures." Antikythera Summer Studio, 3 July 2024. <https://doi.org/10.1162/ANTI.5CZH>.
- Butler, Octavia E. "Bloodchild." *Isaac Asimov's Science Fiction Magazine* 8 (1984).
- Cuboniks, Laboria. *The Xenofeminist Manifesto: A Politics for Alienation*. Verso, 2015.
- Konior, Bogna. "The Dark Forest Theory of the Internet." *Flugschriften* 6, 2020. <https://flugschriften.com/2020/03/24/the-dark-forest-theory-of-the-internet-by-bogna-konior/>.
- . *The Dark Forest Theory of the Internet*. Theory Redux. Polity Press, 2025.
- Liu, Cixin. *Death's End*. Translated by Ken Liu. Head of Zeus, 2016.
- . *The Dark Forest*. Translated by Joel Martinsen. Tor Books, 2016.
- . *The Three-Body Problem*. Translated by Ken Liu. Head of Zeus, 2015.
- McLuhan, Marshall. *Understanding Media: The Extensions of Man*. McGraw-Hill, 1964.
- Stiegler, Bernard. *Technics and Time, Vol. 1: The Fault of Epimetheus*. Translated by Richard Beardsworth and George Collins. Stanford University Press, 1998.
- Vallée, Jacques. *Dimensions: A Casebook of Alien Contact*. Contemporary Books, 1988.