Cambridge Forum on AI: Culture and Society (2026), **2**, e13, 1–15
doi:10.1017/cfc.2026.10023

Cambridge Forum

POSITION PAPER

Empirical AI Ethics: Grounded, relational, transformative

Paula Helm[1] and Selin Gerlek[2]

[1]Empirical Ethics of the Computational, Center for Critical Computational Studies, Goethe University Frankfurt, Frankfurt am Main, Germany and [2]Philosophy of Technology and Politics, Department of Philosophy, University of Amsterdam, Amsterdam, Netherlands
Corresponding author: Paula Helm; Email: helm@c3s.uni-frankfurt.de

(Received 25 June 2025; revised 5 March 2026; accepted 19 March 2026)

Abstract
Mainstream artificial intelligence (AI) ethics, with its reliance on top-down, principle-driven frameworks, often fails to account for the situated realities of diverse communities affected by AI. Critics have therefore argued that AI Ethics increasingly serves corporate interests through practices of "ethics washing," operating as a tool for public relations rather than a means of preventing harm or advancing the common good. In response, this paper adopts a Science and Technology Studies perspective to interrogate AI Ethics itself as a matter of concern. This perspective brings into view a core tension between vertical approaches (external, top-down, principle-based) and horizontal approaches (from-within, risk-mitigating, implementation-oriented). By tracing how these models have shaped the discourse, we show their relevance, as well as their limitations. To move beyond these limitations, we propose a third model that integrates empirical research on local lifeworlds with explicit value commitments. We outline this empirical approach through four shifts: first, a transition from top-down reasoning toward empirical grounding; second, a transition from abstract principles to situated orders of worth; third, a proposal to operate across scales by attending to assemblages of concern; and fourth, a redirection from risk mitigation toward the cultivation of technologies of hope.

Keywords: AI Ethics; care ethics; empirical philosophy; ethics washing; matters of concern; Science and Technology Studies

1. Introduction

Artificial intelligence (AI) Ethics has expanded rapidly as both an academic field and a governance infrastructure. It responds to the manifold challenges raised by the multidimensional phenomenon of AI. Building on the Science and Technology Studies (STS) tradition of tracing sociotechnical phenomena as heterogeneous assemblages (Law 1999), we conceptualize AI as a shifting configuration of interrelated modes of ordering. These modes include AI as a scientific discipline, a technological paradigm, a set of routines and practices, a sociotechnical imaginary and a political–economic regime.

AI Ethics has gained prominence because it articulates normative agendas for how to engage with various critical dimensions of the heterogeneous AI assemblage. Yet this prominence stands in tension with the persistence, and in some domains intensification, of AI-related harms. Despite extensive debates on fairness, cases of algorithmic discrimination continue to surface (Birhane et al. 2021). Privacy concerns have been described as sliding into privacy nihilism (Gertz 2024), while dystopian scenarios are rendered ordinary, with rising investments in autonomous weapon systems

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

(Widder et al. 2024). The troubling centralization of power enabled by data-intensive, deep-learning AI has accelerated with breakthroughs in transformer architectures (Luitse and Denkena 2021), and the energy demands of data centers are reaching unprecedented levels (Dodge et al. 2022).

In light of this list of accelerating problems, the discrepancy between ethical aspiration and empirical reality has led to critiques of "ethics washing" (Bietti 2019), where ethics functions as reputational risk management. This has been followed by "ethics bashing," where critical scholars and civil society actors see ethics as complicit with, rather than a challenge to, power (Phan et al. 2022; van Maanen 2022). This paper contributes to these debates by taking a step back and making AI Ethics itself the matter of concern. We ask: *How is AI Ethics enacting values in relation to AI? where do its core problems originate from? and how can we address them?*

To approach these questions, we draw from STS, which has long examined how disciplines coproduce the very phenomena they study: how population statistics construct categories of people (Hacking 1990), how biological taxonomies invent species classifications (Bowker and Star 1999) and how algorithmic predictions shape possible futures (Mackenzie 2005). Why is the same kind of attention not directed to ethics? Just like the scientific disciplines STS engages with, ethics is not a neutral arbiter of moral truth. Instead, it warrants scrutiny as a specific epistemic culture that coproduces the values, norms and moral guidelines it often presents as universal, objective and timeless (Knorr-Cetina 1999).

The de-essentialization perspective on ethics that we are proposing is inspired by Donna Haraway's concept of situated knowledges, which challenges the "god tricks" of seeing everything from nowhere and calls for "politics and epistemologies of location, positioning, and situating," in which partiality, rather than universality, becomes the basis for credible claims (Haraway 1988). An important but often overlooked implication of Haraway's account is that situated knowledge is not merely local. It depends on networks of relations that make translation possible across diverse communities. The resulting tension between the local and the planetary is equally central to the empirical approach to AI Ethics we propose as a complement to existing frameworks.

On the one hand, AI Ethics requires a vocabulary to confront the large-scale systemic and ecological harms that may follow from concentrated corporate influence, generalist claims of "Foundation Models" and the related infrastructural requirements. On the other hand, established concepts like fairness, transparency or accountability cannot be treated as self-evident or universally stable. If they are invoked as fixed principles, they risk reproducing the very "god trick" Haraway cautions against. For this reason, we argue that alongside its principle-driven and rule-based traditions, AI Ethics must also cultivate approaches that examine how values are articulated, contested and enacted in situated practices, and how AI systems interact with those practices in turn. This dual commitment to shared vocabularies and situated investigation is the core of the empirical complement we advocate.

To navigate between scales, we suggest attending to ethics in a dual register: as *Ethics* (capital *E*), the institutionalized academic field that has significantly shaped the contours of contemporary AI Ethics principles, guidelines, institutions, economies and policies; and as *ethical life*, the enactment of values in practice, a term we borrow from anthropologist Webb Keane (2015). It enables the description of everyday practices oriented toward achieving diverse visions of the moral good. Building on this differentiation, as a first step, this paper contributes to a de-essentialization of AI Ethics by situating it within the scholarly traditions from which it emerged. For this, the authors, trained in philosophy and anthropology, draw on their own experiences of working in Ethics across various institutions and countries over the years, partly identifying with the field, while never seamlessly fitting in. In the second part of the paper, we propose four strategies for addressing the apparent limitations and shortcomings of current AI Ethics: first, a shift from top-down reasoning toward empirical grounding; second, a shift from abstract principles to situated orders of worth; third, a proposal to operate across scales by attending to assemblages of concern; and fourth, a redirection from risk mitigation toward the cultivation of technologies of hope.

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

2. Situating AI Ethics

AI Ethics can be understood as a convergence of multiple scholarly traditions, shaped, in particular, by moral philosophy, bioethics, data and information ethics, and the philosophy of technology (Coeckelbergh 2020). Many of these efforts reflect what we describe as a vertical orientation: attempts to establish shared vocabularies and high-level normative reference points. At the same time, tensions around abstraction, implementation and power have prompted horizontal responses that seek to embed ethical considerations more directly within practices of design and deployment. As we will outline, both directions have made indispensable contributions, yet each encounters structural limits that have, in turn, fueled a growing critical discourse (Aradau and Blanke 2022; Mager et al. 2025; Powell et al. 2022).

These shifts in the prominence of AI Ethics are also visible in the publication record: the annual number of papers and books on the topic rose from fewer than 200 in 2005 to more than 1,500 in 2022, before falling again to below 800 in recent years, despite renewed cycles of AI hype and doom (see Appendix). It is precisely this dynamic that we seek to make sense of in this paper, and it motivates the complementary empirical ethics approach we propose. In our view, such a complement is necessary if we are to engage productively with the performative force of moral judgments, which is inseparable from the politics of technoscience.

2.1. AI Ethics as a historically vertical approach

AI Ethics has largely developed along what we call a *vertical ethics* orientation: efforts to articulate general normative frameworks for AI through guidelines, principles and vocabularies that can be adapted and used across domains and institutions (Jobin et al. 2019). Early work in AI Ethics sought to respond to the opacity and complexity problem of machine-learning systems by translating established ethical resources into the AI domain. Luciano Floridi and colleagues, for example, proposed an influential framework for a "good AI society" that foregrounds explainability, accountability and human-centric values (Floridi et al. 2018; Taddeo and Floridi 2018). This early focus has been foundational in shaping AI Ethics and can be seen as a contribution to a good-faith strand within vertical AI Ethics: attempts to establish shared reference points capable of informing public debate, regulatory efforts or institutional governance.

A wide range of vertical orientations has since developed around value- and rights-oriented conceptual toolkits, focusing on privacy, transparency, explainability, fairness, nondiscrimination and autonomy (Nyholm 2022). These orientations have been essential for rendering AI-related harms visible and addressable (Morley et al. 2023). At the same time, vertical orientations tend to operate at a high level of abstraction: values appear as stable reference points that are articulated independently of the concrete practices through which AI systems are designed, trained and deployed, while their meaning and practical consequences may remain underdetermined (Mittelstadt 2019). This does not mean to say that AI Ethics is dominated by *principlism* in the strict bioethical sense associated with the so-called "four-principles framework" (Beauchamp and Childress 2019). Rather, our concern is broader: vertical AI Ethics often relies on portable normative abstractions whose legitimacy stems less from consensus about their content than from their institutional legitimacy. Eventually, this orientation tends to position norms and values as *external* reference points, or in other words: ethical reflection is situated outside of sociotechnical practices.

This reliance on abstraction becomes problematic when vertical orientations are expected to guide action in contexts marked by conflicting normative pluralism or asymmetrical power relations (Bleher and Braun 2023; Clouser and Gert 1990). Privacy and fairness, for example, are neither self-evident nor univocally interpreted. Their significance and meaning depend on social position, institutional setting or domain (Bridges 2017). What counts as the "best" ethical interpretation may, in practice, be determined more by technical feasibility, economic viability or political expediency than deliberative agreement (Siffels and Sharon 2024). AI Ethics can thus provide a shared

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

language of contestation, but it cannot by itself ensure that this language is translated into accountable practice.

2.2. Corporate co-optation: a stress-test for vertical ethics

Corporate environments function as a stress test for these limitations. Major technology companies publicly endorse values such as fairness, sustainability and explainability (Hagendorff 2020) while simultaneously being criticized for practices that contradict these commitments, including unauthorized user data extraction (Zuboff 2019), biased systems (Buolamwini and Gebru 2018), industry serving benchmarks (Eriksson et al. 2025) and exploitative data labor (Miceli and Posada 2022). This pattern reflects what Sarah Ahmed calls "non-performativity": declarations of ethical commitment operate as substitutes for substantive change (Ahmed 2016). The problem is not that vertical ethics is *per se* reducible to "ethics washing," but that abstract value concepts are comparatively easy to appropriate while robust enforcement and accountability mechanisms remain missing.

Institutionalization efforts such as the GDPR, the EU's Trustworthy AI agenda and the UNESCO AI Ethics recommendations have contributed to embedding ethics into governance infrastructures, yet their practical effects remain uneven and highly dependent on implementation capacities (Goodman and Flaxman 2015). In parallel, a growing "economy of virtues" around AI has turned grassroots efforts and ethical expertise into a strategic resource, circulating between academia, industry and policy (Phan et al. 2022).

Vertical orientations have proven indispensable for establishing AI as a matter of ethical and critical concern. Yet they have also been limited when faced with the complex power structures embedded in AI deployment. We therefore argue that it is precisely the *vertical* orientation that has contributed to making AI Ethics not just powerful but, in many respects, counterproductive to basic aims like preventing harm or advancing the common good. This is because the vertical character leaves AI Ethics particularly susceptible to co-optation by industry players for adversarial purposes. By relying on a narrow set of external abstract principles, AI Ethics limits its ability to respond dynamically to the situational entanglements through which values are being enacted in various AI–human constellations. This is how the hype around AI Ethics, the tech industry's strategic influence and the absence of robust enforcement have brought to light systematic limitations of the vertical orientation, motivating the emergence of horizontal, practice-oriented approaches.

2.3. Horizontal orientations and their limitations

Horizontal approaches trace their origins to broader critiques of the political dimensions inherent in technological artifacts and seek to embed ethical reflection within the practices of design and deployment. Technologies are not ahistorical, instead, they are situated within culturally and politically contingent regimes of knowledge (Akrich 1992). This also applies to AI systems (Tacheva and Ramasubramanian 2023). From this perspective, AI does not merely mediate action but enacts specific epistemic assumptions, many of which are contested (Gerlek and Weydner-Volkmann 2025). Treating AI as shaped by and shaping culture (Seaver 2017), horizontal scholarship highlights how biases embedded in computer science become materialized in technical systems, while simultaneously amplifying existing epistemological regimes of dominance. One manifestation is the deeply rooted Anglocentrism that inheres in both the architectures of large language models and their training data (Bella et al. 2024).

Frameworks such as *Value Sensitive Design* have, early on, advanced the idea of embedding ethical commitments throughout the development process (Friedman 1997; Knobel and Bowker 2011; van de Poel 2013). These proactive approaches aim to align innovation with societal values during early and midstream development (Fisher et al. 2015), and they now thrive on the fertile ground of contemporary AI hype and doom discourse. Approaches such as Explainable AI, fairness measures, dataset

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

debiasing, value alignment and jailbreaking for AI safety represent attempts to bridge principle and practice. As such, they form part of a horizontal ethics that seeks to operate from within technological development rather than from above (in a vertical orientation). Still, this raises key questions: who is included in this "within"? who gets to decide? and whose voices remain excluded?

Despite being more implementation-oriented than vertical orientations, many of these frameworks continue to rely on the prior articulation of societal values that are subsequently translated into design requirements. In this sense, they do not abandon vertical norm articulation but relocate it into processes of technological development. As a result, much of this work remains shaped by a design discourse that defines ethical problems in technically solvable terms, despite acknowledging that technologies can reshape the moral situation in which they operate (Akrich 1992; Morozov 2014; Verbeek 2006). This is not to say that technological solutionism is bad *per se*, but it tends to mistake symptoms for causes (Helm and Hagendorff 2021; Siffels and Sharon 2024).

The so-called *Moral Machine* experiment exemplifies another approach that, while appearing horizontal, still engages in reductive framing. Designed to explore people's intuitive judgments on moral dilemmas involving autonomous vehicles (Awad et al. 2018), it used a global online survey to gather input on whom autonomous cars should prioritize in fatal scenarios. Participants chose between categories such as pedestrians versus passengers, young versus old or humans versus animals. Though the study revealed cultural variation and underscored the value of diverse input, its methodology remained top-down: expert-led, based on constrained moral logics, and disconnected from everyday experiences that are shaped by power asymmetries and practical concerns and constraints.

As AI Ethics becomes increasingly institutionalized, integrated into what Phan et al. describe as "economies of virtue" (Phan et al. 2022) and shaped by governments and corporations, approaches centered on abstract guidelines, risk mitigation, thought-experiments and design-level values prove insufficient. While horizontal scholarship moves ethical reflection closer to the practices than vertical approaches, systematic efforts to integrate research that investigates how ethical life unfolds in context remain scarce. What is needed, therefore, is not the abandonment of existing approaches but their empirical grounding: an orientation that examines how values are articulated, contested and enacted within the social, cultural, political and historical conditions from which technologies such as AI emerge and in which they operate and evolve.

3. Empirical AI Ethics

The following sections address the question of "how to close the loop" between situated observations and normative reasoning (Sharon and Koops 2021). We do this by outlining a four-pronged approach, from foundations and relations to scales and transformations. The combination of strategies we propose aims to empower AI Ethics as a transformative field capable of responding to community needs while actively shaping AI, rather than relegating it to a power-conserving, risk-mitigation role.

3.1 Foundations: from thought experiments to empirical grounding

Emerging in the late 1990s, philosophy's Empirical Turn describes a dynamic shift away from the dystopian critiques associated with thinkers such as Heidegger and Anders and toward scholars such as Don Ihde and Hans Achterhuis, who foregrounded technology's entanglement with everyday practices and institutions (Achterhuis 1992). This shift parallels a wider Practice Turn in the humanities and social sciences, which treats routines and material engagements as key sites where knowledge and norms are produced (Cetina et al. 2005) and where practice is treated as central for understanding how bodies, technologies, meanings and norms are co-constituted (Bourdieu 1977; Butler 1993). STS, then, extends these moves by converging in three core insights: first, phenomena are simultaneously epistemological and ontological, constituted through actor-networks (Callon 1984); second, agency is an effect of relational practice (Callon and Law 1997); and third, symbolic and technical mediations

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

coproduce phenomena in situated ways. Reality is inseparable from observation and intervention (Barad 2007).

These insights also relocate AI Ethics: from the application of principles to bounded systems toward the sociotechnical processes through which AI becomes a way of knowing and governing. Louise Amoore's emphasis on opacity and partial account-giving shifts ethical inquiry away from transparency demands and toward the practices through which algorithmic arrangements justify, distribute and deflect responsibility (2020). On this basis, empirical AI Ethics can be defined as the study of ethical life in AI assemblages. Lucy Suchman's critique of AI's "uncontroversial thingness" adds a temporal dimension: ethical debate often begins after categories have stabilized and responsibility has already been allocated across sociotechnical chains (2023). This relegates AI Ethics to a power-conserving risk-management enterprise.

What is needed are approaches that move inquiry upstream. Ethnographic work on machine learning as a mode of experimentation on thought and conduct and critical analyses of how evaluation regimes function as infrastructures of power are exemplary here (Luitse et al. 2025, Mackenzie 2017). Investigations in the routines and realities of data work add another pointer to where an upstream empirical ethics must go, and what it must follow. It directs attention to the mundane but consequential sites where "AI" is assembled as a working system: the instruction sets, interfaces, taxonomies, quality checks and payment structures through which data are produced, cleaned, labeled and rendered compatible with model development (Miceli et al. 2021). These approaches make visible how ethical stakes are built into the field at the point of specification: in decisions about what counts as an admissible target, how "error" is rendered legible and acceptable, and how competitions, leaderboards and platform access distribute incentives, credibility and authority.

Such qualitative inquiry also clarifies why harms that are often discussed as "bias" or "misclassification" can be better understood as failures of sociotechnical recognition. Credit scoring provides a clear case. Systems trained on formal financial traces frequently misread informal practices in the Global South as indicators of risk or corruption, even when they reflect creditworthiness as it is locally produced through relations of solidarity and care (Heeks and Swain 2018). An empirical ethics approach does not simply register that an algorithm performs poorly on a given population. It empirically examines how model categories, data proxies and validation practices define what counts as legitimate economic behavior, and how these definitions voyage across contexts while presenting themselves as generalizable.

Close-up empirical studies stand in sharp contrast to sanitized thought experiments, such as the trolley-problem framing deployed in the Moral Machine experiment, which abstracts away the power-differentiated realities through which AI systems are specified, trained, validated, distributed and adopted. By making early and implicit judgments visible and by attending to the creative and unexpected ways in which technology is shaped, empirical ethics research does not merely add context to assessment. It supplies a foundational orientation as a practice-based, power-attentive inquiry into how ethical registers are materially assembled through situated practice.

3.2 Relations: from principles to orders of worth

The empirical approaches described above may raise questions about the status of normativity. Are principles to be understood as entirely relative to practice? The answer is no: principles do not become obsolete, but their status shifts. Judgment no longer derives from a standpoint of external reasoning; it is articulated and contested in relation to communities of practice.

This analytic shift positions us well to describe what, following Boltanski and Thevenot (2006), can be called "orders of worth": diverse justificatory logics that vary across communities of practice, domains and even individuals. Rather than following either a linear vertical logic or a linear horizontal logic, their framework can be used as a source of inspiration to study how actors articulate,

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

defend and enact values in practice and how such valuations become stabilized, challenged or overturned.

Attending to orders of worth helps make sense of recurrent conflicts between evaluative preferences: an industrial order may elevate efficiency and cost–benefit rationales, while a civic order may privilege equity and collective welfare. Taking a step back to identify the relevant and potentially conflicting orders of worth in a given field generates a context-sensitive understanding that reliance on principles (vertical), or on prestructured participatory designs (horizontal), often bypasses.

Close-up observation of ethical life may also bring into view the dominance of particular orders of worth imposed upon a community that does not fully share it. This can produce a profound desire for transformation, which becomes visible, for instance, when studying the frustrations of nurses working under increasingly technocratic regimes where efficiency is granted overriding force, yet in practice prevents them from enacting what they understand as "good care" (Pols 2015).

Furthermore, studying diverging orders of worth can bring into view the damaging consequences of vertical approaches that are too often power-preserving, as shown, for example, in the context of criminal justice. Canonical by now is the case of decision-support algorithms trained on historical crime data to estimate the likelihood of reoffending, thereby shaping parole decisions. In such high-stakes settings, it is vital that systems do not reproduce discriminatory patterns. Yet studies have shown that widely used fairness metrics often fail to capture intersectional inequalities (Barocas et al. 2023). Rather than mitigating injustice, they can inadvertently stabilize existing hierarchies. This becomes visible in a case where they have proven insufficient to prevent disproportionately harsh treatment of defendants with darker skin tones (Lum and Isaac 2016). Notably, qualitative reporting from interviews also points to forms of irritation and unrest not only among those who felt unfairly targeted but also among those ostensibly advantaged by the system (Angwin and Larson 2015).

An empirically grounded AI Ethics that begins with the systematic observation of practices, perspectives and consequences may challenge the dominance of certain orders of worth, confronting them with others that exist, albeit at the margins of power. For instance, systems previously evaluated through utilitarian or procedural notions of justice as fairness might be reassessed in light of historically situated conceptions such as epistemic justice (Helm et al. 2024). A reorientation can also involve extending domain-specific values into new territories, such as the application of care, traditionally confined to reproductive labor and health care, to fields like law enforcement (Asaro 2019). Care ethics, with its relational understanding of ethical life, resonates well with the project of empirical AI Ethics. As Joan Tronto famously stated:

> Care is a species of activity that includes everything that we do to maintain, continue, and repair our 'world' so that we can live in it as well as possible. (1998, p. 40)

Care ethics centers on experience and the moral urgency of attending to the needs of others, especially those rendered vulnerable or marginalized. By placing repair at its core, care ethics also foregrounds the historical embeddedness of ethical life, recognizing that before questions of procedural justice can be meaningfully posed, there is often a demand to redress the accumulated weight of past and ongoing harm.

What, then, would it mean to evaluate an AI system in criminal justice not through the logic of distributive fairness, commonly quantified via the proportional distribution of false positives and negatives, but through the logic of care? Such a shift would reorient our focus: from abstract metrics to historically accumulated vulnerabilities. It would prioritize the repair of social worlds, in which trust in the criminal justice system has, for many, been eroded by experiences of mistreatment, often based on markers of social difference. This erosion of trust is not incidental; it feeds into cycles of recidivism and violence, perpetuating the very forms of marginalization that AI systems purport to neutrally assess. Contrary to the orders of worth underlying distributive fairness, from a care ethics perspective, addressing algorithmic bias is not a matter of tuning models or refining datasets to level

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

the playing field. It becomes a demand to transform the social and historical conditions that produce inequality in the first place. Here, the primary order is to engage in the situated work of repair.

3.3 Scales: from micro–macro to assemblages of concern

Where empirical investigations into varied orders of worth reveal a demand for change, it may be fruitful to seek guidance from contexts that are most different. A plurality of conceptions of ethical life flourishes beyond the confines of dominant Western principles of fairness, accountability and transparency – the go-to principles of vertical ethics. Amerindian notions of Sumak Kawsay (buen vivir), for instance, foreground the entanglement of human and more-than-human actors by emphasizing conditions of living well in reciprocal relation with others and with the land (Acosta 2013). This orientation shares with the orders of care a rejection of anthropocentric and dualistic models of agency. Instead, both assume an infinite regress of relationality, where no actor, human or otherwise, exists in isolation. When applied to the domain of AI, such perspectives invite an ethical gaze that foregrounds the impact of technological systems on more-than-human well-being. Within this worldview, any AI innovation that causes environmental harm cannot plausibly be considered ethically desirable, rendering efficiency-based cost–benefit analyses as logically and morally untenable, even when short-term individual gains accrue (Taffel et al. 2022).

At the same time, these perspectives bring a tension into view: between situated inquiries into ethical life and the scale at which generalist AI models operate. Generalist models come with massive infrastructural apparatuses that can be socially and materially destructive. This matters for empirical ethics because such systems are typically justified through efficiency-oriented orders of worth (speed, throughput, optimization). The tension is that what counts as "efficient" or "successful" at scale is often achieved by shifting costs onto local communities of practice, but the sources of harms remain invisible and, hence, incontestable when inquiry focuses only on local context.

While it is essential to cultivate contextual sensitivity, we must also grapple with the undeniable need for coordinated responses to the global implications of AI. The deployment of energy-hungry models across sectors necessitates the establishment of regulatory processes aimed explicitly at safeguarding the flourishing of plural forms of life (Hecht 2018). How can we find modes to address the relationality of various scales, while not reducing sociotechnical complexity to quantifiable threats that translate into mere risk-mitigation measures or overly abstract frameworks that do not correspond to the situated reality of ethical life?

In a world of entangled infrastructures, planetary exhaustion and governing AI models, regulation demands more than policy; it requires onto-political awareness. The idea of publics as coming into being, theorized by Bruno Latour, and its methodological operationalization by Noortje Marres are instructive in this context (2024). Publics are not preexisting entities but emerge through controversy, breakdown and shared matters of concern (Latour and Weibel 2005), ultimately building assemblages of concern. Technologies such as AI, however, do not merely mediate communication or disseminate information; they participate in the assembly of publics by shaping what is rendered visible, debatable or ignored. If we assume that publics play a role in shaping regulation and development, and that AI is simultaneously co-constituting the very publics from which consensus is meant to emerge, we risk a recursive paradox: the goat becomes its own gardener. How to escape this dilemma?

What is needed are cross-scale interventions, which make us recognize how ethical life can find its expressions through assemblages: situated and emerging formations that link locality to wider networks of responsibility and meaning. A concrete methodological approach to studying such multiscalar assemblages lies in what Huub Dijstelbloem calls infrastructural investigations, which focus on the practices by which journalists, activists, artists and organizations contest diffuse forms of power (2021). Using techniques such as mapping, counter-databases, interactive visualizations and testimonials, these actors reconfigure relations of visibility and accountability. They counter forms of organized irresponsibility, evident in domains such as border politics, journalism and supply chains,

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

where harm is systemic yet dispersed, and difficult to attribute, a situation exacerbated by the opacity of AI black box technologies.

Whether instantiated through crude AI quality benchmarks (Raji et al. 2021), algorithmic news recommenders (Helberger 2021) or habit-forming interfaces (Ash et al. 2018), exposing sociotechnical vehicles as forms of power operating within broad ecosystems of influence challenges linear notions of causality and, by extension, vertical ethics and top-down governance. Making these dynamics legible and contestable as assemblages of concern can complement vertical ethics with a relational ethics of cross-scalar co-responsibility. A direction for the formation of such assemblages of concern as an ethics of co-responsibility comes from the notion of *pluriversality*, which is often referenced in tandem with the abovementioned *buen vivir*.

Inspired by the Mexican Zapatista movement, pluriversality invites us to envision "a world where many worlds fit." Rooted in decolonial and post-development thought, pluriversality creates space for diverse ways of knowing, being and doing (Escobar 2018). Embedded within it is not only a sharp critique of the universalizing tendencies of vertical ethics but also of the well-meaning horizontal orientations of co-creation, which can reproduce colonial patterns by exporting Euro-American values under the banner of participation (Helm et al. 2023). In contrast, pluriversality supports autonomous designs of sociotechnical assemblages: forms of thinking, knowing, making and doing that emerge from within communities of practice but respond to concerns that can only be understood by acknowledging interdependencies of ecosystems and infrastructures. With their planetary perspectives, yet concern for local autonomy, notions of pluriversality and buen vivir offer normative orientations for cross-scalar intervention. The idea is to reframe agency not as possession, but as circulation and to practice ethics not as distant assessment but as situated making and doing (Ratto 2011). In the following, we highlight strategies for the enactment of this idea.

3.4 Transformation: from risk mitigation to technologies of hope

To emancipate itself from its role as a whitewasher, rather than being driven by a relentless pace of innovation where damage prevention dictates the agenda, AI Ethics could contribute to manifesting alternatives. Just as the historical development of STS evolved from critique into an engaged program (Sismondo 2017), Ethics should not exhaust itself in abstract assessment. Such an orientation entails a conscious *ethical* choice not to dwell in despair but to engage in the generative power of concrete hope.

Amidst empirical contextualization, certain foundational commitments remain that most would recognize: if harm or destruction is observed, there is an imperative to respond. Yet before action can emerge, there must first be a sensation, a representation and an arousal that catalyzes the will to act, and from this, hope is born (Willow 2023). Not the dulled optimism that numbs into inaction (Berlant 2011) but a hope that energizes. The kind that Ernst Bloch (still haunted by the cruelties of the Second World War) termed *concrete hope* (1985). Conceived as something tied to the concrete, hope is not a projection onto distant or ill-defined futures. It emerges from engagement in feasible alternatives, which respond to local concerns but are open to scale. Concrete hope acknowledges conditions of constraint and powerlessness without equating improbability with impossibility (Han-Pile 2017).

This orientation resonates with Appadurai's distinction between an "ethics of probability" and an "ethics of possibility" (Appadurai 2013, pp. 188, 295). The former limits us to the logic of what is most likely, a logic that binds us to focus on the visions of those who are in power. The latter, in contrast, opens spaces for imagination and expands our collective capacity to aspire. Empirical AI Ethics can move beyond the well-worn terrain of probability by opening itself to pluriversal figurations of technology.

An example comes from a project in which members of the Guarani-Kaiowá and Ikpeng-Xingu, Indigenous peoples of the Upper Amazon, were invited to imagine technologies they desired. Many participants responded with paintings of mythical beings: animals with human traits. At first, the

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

facilitators assumed the prompt had been misunderstood. But the images were intentional. Among them was the *Nyra Nhe'Éngatu*, a radiant hybrid of bird and human, drawn from Ikpeng cosmology. This creature bridges earth and sky – guardian of the unseen, responsible for the care of communication itself. It is said to visit those in need of sending urgent messages, especially in moments of injury or danger. One 12-year-old boy painted this being to represent what he believed a smartphone should do for his people (Bonami et al. 2026).

Unlike Western AI benchmarks, which are typically governed by anthropocentric ideals or operational efficiency, the Nyra Nhe'Éngatu functions as an onto-epistemic opening (Cadena 2015): an orienting force that resonates closely with ideas of care ethics while diverging fundamentally from anthropocentric frameworks of explainability or fairness. Through collaboration with local communities of practice, an empirical ethics approach has the potential to challenge dominant Silicon Valley narratives that valorize human-likeness as primary indicators of successful AI innovation (Brown et al. 2020). This dominant model presumes the inevitability of centralized, resource-intensive and infrastructure-heavy AI systems. The children of the Guarani-Kaiowá and Ikpeng-Xingu, in contrast, offer a radically different vision for AI. It is imagined not to replicate human capacities for convenience, nor driven by an obsession for explainability, but to assist in tasks that humans are inherently unable to perform, all while remaining within deeply humane registers of worth: supporting and caring for those in need.

By emphasizing possibilities rather than probabilities, we move away from mere reactionary responses to technological advances, risks and harms. An empirical ethics that seeks to be transformative could focus on elevating the hopes, desires and aspirations of otherwise marginalized and unheard people. This helps to move away from a mere critical perspective on AI as a "naughty child" needing discipline. And indeed, across the globe, numerous initiatives illustrate how machine learning can be harnessed to empower marginalized communities, for restorative justice, to challenge oppressive systems and foster domain-specific values. In the following, we provide two examples of such attempts to enact orders of worth through the development and adjustment of technical tools.

The *Live Language Initiative* serves as an example of a technology of hope that strives to elevate marginalized voices. This project functions as a database, catalogue and collaborative platform, surpassing existing language databases not in volume but in its diversity. Its name, Live Language, reflects the idea that languages are alive; they are born, evolve and die, along with the people who speak them (Koch et al. 2024). While most language databases focus on the 40 most resourced languages, thereby pivoting around English, Live Language strives to document the linguistic diversity of the remaining 6,960 languages and dialects spoken today (Skirgård et al. 2023). The vision is to collaborate with marginalized speakers around the world to create a pluriversal alternative to centralized models, fostering the development of tailored alternatives: a world where many languages fit.

Another noteworthy initiative focuses on efforts to enact domain-specific orders of worth, thereby countering the top-down tendencies associated with the algorithmization of the public sphere (Aradau and Blanke 2022). This approach resonates with the earlier discussion of the need to assess the enactment of values differently across distinct domains, where diverse communities of practice may prioritize different orders of worth. *Cultural citizenship*, for instance, has been established as a key order within the journalism sector. Ferraro et al. (2024) therefore prototypically enact four concrete steps for the redesign of recommender algorithms for music curation to reflect this order: (1) moving beyond a commercial orientation, (2) recognizing the cumulative influence of recommendations, (3) integrating a metric of commonality that measures how well recommendations familiarize a given listener population with specific content categories and (4) avoiding a top-down normative approach by complementing it with experiments that assess the effectiveness of commonality in achieving the target value. This reconfiguration of algorithmic recommendation in the music domain is a prime example of enacting an ethics of possibility to create technologies of hope.

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

4. Conclusion

Whether vertical or horizontal in orientation, AI Ethics' focus on abstract principles, technical "solutionism" and off-the-shelf participatory designs has left AI Ethics susceptible to co-optation by powerful industry actors, reducing it to a tool of risk management rather than a transformative force. We propose four strategies to overcome these limitations. First, we insist on the importance of empirically studying the enactment of values as varied orders of worth. These insights should not be seen as peripheral but as foundational for shaping regulatory frameworks that aim to be just and generative. To escape the narrow confines of a paradigm fixated on risk mitigation, we call for a reorientation toward possibility. In keeping with a practice-based ethics, we advocate for an ethics of hope grounded in the concrete: a commitment to making the possible liveable through situated, normatively oriented actions.

Ethical life does not begin with abstract principles but with practice and experience. To engage with this complexity, empirical ethicists may mobilize a rich repertoire of methods, from participant observation, interviews and creative storytelling to the design and curation of databases, archives and algorithms. Through this work, they gather what Sarojini Nadar has called "data with souls": forms of knowledge that carry with them the resonance of human aspiration, pain, fear and hope (Nadar 2014). These are not merely data points but fragments of life: gestures, stories, decisions, priorities, aspirations and spontaneous reactions. They are artifacts of ethical life that refuse to be flattened into metrics. Such data also remind us that many of the world's languages and cultures arise not from written scripts but from oral traditions. Much of what has endured across generations has done so without technological scaffolding, preserved not in code and energy-intensive data centers but in bodies, voices and memorized stories. This reminder offers an insight into our contemporary entanglements with AI. The belief that intelligent machines and digital archives are necessary to safeguard values is, at its core, an ideological assumption. It is part of a techno-solutionist worldview that positions AI as the inevitable vessel of human well-being.

But the endurance of oral traditions teaches us otherwise: it is not technology that ensures the survival of meaning and worth; it is care. No algorithm can substitute this fidelity. This reorientation challenges the problem-solution logic that underpins much of AI Ethics (Birhane 2021), even in its more horizontal forms. It invites us to privilege meaning over metrics and relationality over optimization. In this light, empirical ethics is not merely a tool for resolving ethical dilemmas; it is the study and practice of concrete hope.

Funding statement. This work was generously supported by the Amsterdam Institute for Advanced Studies from 2023 to 2025.

Open access funding provided by University of Amsterdam.

Competing interests. The authors have no competing interests to declare.

References

Achterhuis, H. 1992. "De Maat van de Techniek: Zes Filosofen over Techniek. Günther Anders, Jacques Ellul, Arnold Gehlen, Martin Heidegger, Hans Jonas En Lewis Mumford." Baarn: Ambo.

Acosta, A. 2013. *El Buen Vivir: Sumak Kawsay, una oportunidad para imaginar otro mundo*. 1st ed. Barcelona: Icaria editorial.

Ahmed, S. 2016. "How Not to Do Things with Words." SSRN Scholarly Paper No. 3093652. Social Science Research Network, July 1. https://papers.ssrn.com/abstract=3093652.

Akrich, M. 1992. "The De-scription of Technical Objects." Shaping Technology/Building Society. Studies in Sociotechnical Change, 205. https://shs.hal.science/halshs-00081744.

Amoore, L. 2020. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham London: Duke University Press.

Angwin, J., and J. Larson. 2015. "Technical Response to Northpointe." ProPublica. Accessed January 10, 2022. https://www.propublica.org/article/technical-response-to-northpointe?token=JEL-Unlbmks60fsRpClZ_emcktORLywG.

Appadurai, A. 2013. *The Future as Cultural Fact: Essays on the Global Condition*. 1st ed. London New York: Verso Books.

Aradau, C., and T. Blanke. 2022. *Algorithmic Reason: The New Government of Self and Other*. Oxford: Oxford University Press.

Asaro, P. M. 2019. "AI Ethics in Predictive Policing: From Models of Threat to an Ethics of Care." *IEEE Technology and Society Magazine* 38 (2): 40–53. https://doi.org/10.1109/MTS.2019.2915154.

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

Ash, J., B. Anderson, R. Gordon, and P. Langley. 2018. "Digital Interface Design and Power: Friction, Threshold, Transition." *Environment and Planning D: Society and Space* 36 (6): 1136–1153. https://doi.org/10.1177/0263775818767426.

Awad, E., S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J.-F. Bonnefon, and I. Rahwan. 2018. "The Moral Machine Experiment." *Nature* 563 (7729): 59–64. https://doi.org/10.1038/s41586-018-0637-6.

Barad, K. 2007. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham London: Duke University Press.

Barocas, S., M. Hardt, and A. Narayanan. 2023. *Fairness and Machine Learning: Limitations and Opportunities*. Cambridge, Massachusetts: The MIT Press.

Beauchamp, T. L., and J. F. Childress. 2019. *Principles of Biomedical Ethics*. 8th ed. New York, NY Oxford: Oxford University Press.

Bella, G., P. Helm, G. Koch, and F. Giunchiglia. 2024. "Tackling Language Modelling Bias in Support of Linguistic Diversity." The 2024 ACM Conference on Fairness, Accountability, and Transparency, June 3, 562–572. https://doi.org/10.1145/3630106.3658925.

Berlant, L. 2011. *Cruel Optimism*. Durham: Duke University Press. https://doi.org/10.1515/9780822394716.

Bietti, E. 2019. "From Ethics Washing to Ethics Bashing: A View on Tech Ethics from within Moral Philosophy." SSRN Scholarly Paper No. 3513182. Social Science Research Network, December 1. https://papers.ssrn.com/abstract=3513182.

Birhane, A. 2021. "Algorithmic Injustice: A Relational Ethics Approach." *Patterns* 2 (2): 100205. https://doi.org/10.1016/j.patter.2021.100205.

Birhane, A., V. U. Prabhu, and E. Kahembwe. 2021. "Multimodal Datasets: Misogyny, Pornography, and Malignant Stereotypes." arXiv:2110.01963. Preprint, arXiv, October 5. https://doi.org/10.48550/arXiv.2110.01963.

Bleher, H., and M. Braun. 2023. "Reflections on Putting AI Ethics into Practice: How Three AI Ethics Approaches Conceptualize Theory and Practice." *Science and Engineering Ethics* 29 (3): 21. https://doi.org/10.1007/s11948-023-00443-3.

Bloch, E. 1985. *Das Prinzip Hoffnung - 3 Bände: Band 5: Das Prinzip Hoffnung*. 1st ed. Frankfurt am Main: Suhrkamp Verlag.

Boltanski, L., and L. Thevenot. 2006. *On Justification: Economies of Worth*, Erste Ausgabe Edition. Translated by C. Porter. Princeton: Princeton University Press.

Bonami, B., P. Helm, R. van Voorst, and A. da Silva. 2026. *Reimagining Knowledge in the Amazon Rainforest. Crafting Alternative Futures*. Leuven: Leuven University Press.

Bourdieu, P. 1977. "Outline of a Theory of Practice." In *Cambridge Studies in Social and Cultural Anthropology*, Translated by R. Nice, Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511812507.

Bowker, G. C., and S. L. Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, Massachusetts: The MIT Press. https://doi.org/10.7551/mitpress/6352.001.0001.

Bridges, K. M. 2017. *The Poverty of Privacy Rights*. 1st ed. Stanford, California: Stanford University Press.

Brown, T., B. Mann, N. Ryder et al. 2020. "Language Models Are Few-Shot Learners." In *Advances in Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, 1877–1901, Vol. 33. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.

Buolamwini, J., and T. Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of the 1st Conference on Fairness, Accountability and Transparency, January 21, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html.

Butler, J. P. 1993. *Bodies that Matter: On the Discursive Limits of Sex*. London New York: Routledge.

Cadena, M. D. L. 2015. *Earth Beings: Ecologies of Practice across Andean Worlds*. Durham, NC: The Lewis Henry Morgan Lectures. Duke University Press.

Callon, M. 1984. "Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay." *The Sociological Review* 32 (1_suppl): 196–233. https://doi.org/10.1111/j.1467-954X.1984.tb00113.x.

Callon, M., and J. Law. 1997. "Agency and the Hybrid Collectif." In *Mathematics, Science, and Postclassical Theory*. Durham, NC: Duke University Press. https://www.degruyterbrill.com/document/doi/10.1515/9780822382720-006/pdf?licenseType=restricted.

Cetina, K. K., T. R. Schatzki, and E. von Savigny, eds.. 2005. *The Practice Turn in Contemporary Theory*. London: Routledge. https://doi.org/10.4324/9780203977453.

Clouser, K. D., and B. Gert. 1990. "A Critique of Principlism." *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 15 (2): 219–236. https://doi.org/10.1093/jmp/15.2.219.

Coeckelbergh, M. 2020. *AI Ethics. The MIT Press Essential Knowledge Series*. Cambridge, Massachusetts: MIT Press.

Dijstelbloem, H. 2021. *Borders as Infrastructure: The Technopolitics of Border Control*. Cambridge, Massachusetts: The MIT Press. https://doi.org/10.7551/mitpress/11926.001.0001.

Dodge, J., T. Prewitt, R. T. Des Combes et al. 2022. "Measuring the Carbon Intensity of AI in Cloud Instances." Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (New York, NY, USA), FAccT'22, June 20, 1877–1894. https://doi.org/10.1145/3531146.3533234.

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

Eriksson, M., E. Purificato, A. Noroozian et al. 2025. "Can We Trust AI Benchmarks? an Interdisciplinary Review of Current Issues in AI Evaluation." arXiv:2502.06559. Preprint, arXiv, May 25. https://doi.org/10.48550/arXiv.2502.06559.

Escobar, A. 2018. *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds*. Durham, NC: Duke University Press.

Ferraro, A., G. Ferreira, F. Diaz, and G. Born. 2024. "Measuring Commonality in Recommendation of Cultural Content to Strengthen Cultural Citizenship." *ACM Transactions on Recommender Systems* 2 (1): 1–32. https://doi.org/10.1145/3643138.

Fisher, E., M. O'Rourke, R. Evans, E. B. Kennedy, M. E. Gorman, and T. P. Seager. 2015. "Mapping the Integrative Field: Taking Stock of Socio-Technical Collaborations." *Journal of Responsible Innovation* 2 (1): 39–61. https://doi.org/10.1080/23299460.2014.1001671.

Floridi, L., J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, and E. Vayena. 2018. "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds and Machines* 28 (4): 689–707. https://doi.org/10.1007/s11023-018-9482-5.

Friedman, B. 1997. *Human Values and the Design of Computer Technology*. Cambridge: Cambridge University Press.

Gerlek, S., and S. Weydner-Volkmann. 2025. "Materiality and Machinic Embodiment: A Postphenomenological Inquiry into ChatGPT's Active User Interface." *Journal of Human-Technology Relations* 3 (March): 1–15. https://doi.org/10.59490/jhtr.2025.3.7387.

Gertz, N. 2024. *Nihilism and Technology*. Lanham (Md): Rowman & Littlefield.

Goodman, B., and S. Flaxman. 2015. "EU Regulations on Algorithmic Decision-Making and a "Right to Explanation"." Paper presented at WHI 2016. 2016 ICML Workshop on Human Interpretability in Machine Learning.

Hacking, I. 1990. "Making Up People." London/New York: Routledge.

Hagendorff, T. 2020. "The Ethics of AI Ethics: An Evaluation of Guidelines" *Minds and Machines.* 30 (1): 99–120. https://doi.org/10.1007/s11023-020-09517-8.

Han-Pile, B. 2017. "Hope, Powerlessness, and Agency." *Midwest Studies In Philosophy* 41 (1): 175–201. https://doi.org/10.1111/misp.12069.

Haraway, D. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14 (3): 575. https://doi.org/10.2307/3178066.

Hecht, G. 2018. "Interscalar Vehicles for an African Anthropocene: On Waste, Temporality, and Violence." *Cultural Anthropology* 33 (1): 109–141. https://doi.org/10.14506/ca33.1.05.

Heeks, R., and S. Swain. 2018. "An Applied Data Justice Framework: Analysing Datafication and Marginalised Communities in Cities of the Global South." *SSRN Electronic Journal*, ahead of print. https://doi.org/10.2139/ssrn.3425885.

Helberger, N. 2021. "On the Democratic Role of News Recommenders." In *Algorithms, Automation, and News*. New Yrok/London: Routledge.

Helm, P., G. Bella, G. Koch, and F. Giunchiglia. 2024. "Diversity and Language Technology: How Language Modeling Bias Causes Epistemic Injustice." *Ethics and Information Technology* 26 (1): 8. https://doi.org/10.1007/s10676-023-09742-6.

Helm, P., A. de Götzen, L. Cernuzzi, A. Hume, S. Diwakar, S. Ruiz Correa, and D. Gatica-Perez. 2023. "Diversity and Neocolonialism in Big Data Research: Avoiding Extractivism while Struggling with Paternalism." *Big Data and Society* 10 (2): 20539517231206802. https://doi.org/10.1177/20539517231206802.

Helm, P., and T. Hagendorff. 2021. "Beyond the Prediction Paradigm: Challenges for AI in the Struggle against Organized Crime." *Law and Contemporary Problems* 84 (3): 1–17.

Jobin, A., M. Ienca, and E. Vayena. 2019. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1 (9): 389–399. https://doi.org/10.1038/s42256-019-0088-2.

Keane, W. 2015. *Ethical Life: Its Natural and Social Histories*. Princeton: Princeton University Press.

Knobel, C., and G. C. Bowker. 2011. "Values in Design." *Communications of the ACM* 54 (7): 26–28. https://doi.org/10.1145/1965724.1965735.

Knorr-Cetina, K. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, Massachusetts: Harvard University Press.

Koch, G., G. Bella, P. Helm, and F. Giunchiglia. 2024. "Layers of Technology in Pluriversal Design Decolonising Language Technology with the Live Language Initiative." *CoDesign* 20 (1): 77–90. https://doi.org/10.1080/15710882.2024.2341799.

Latour, B., and P. Weibel, eds. 2005. *Making Things Public: Atmospheres of Democracy*. Cambridge, Massachusetts: MIT Press.

Law, J. 1999. "After ANT: Complexity, Naming and Topology." *The Sociological Review* 47 (S1): 1–14. https://doi.org/10.1111/j.1467-954X.1999.tb03479.x.

Luitse, D., T. Blanke, and T. Poell. 2025. "AI Competitions as Infrastructures of Power in Medical Imaging." *Information, Communication & Society* 1–22. https://doi.org/10.1080/1369118X.2024.2334393.

Luitse, D., and W. Denkena. 2021. "The Great Transformer: Examining the Role of Large Language Models in the Political Economy of AI." *Big Data and Society* 8 (2): 20539517211047734. https://doi.org/10.1177/20539517211047734.

Lum, K., and W. Isaac. 2016. "To Predict and Serve?" *Significance* 13 (5): 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x.

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

Mackenzie, A. 2005. "The Performativity of Code: Software and Cultures of Circulation." *Theory, Culture & Society* 22 (1): 71–92. https://doi.org/10.1177/0263276405048436.

Mackenzie, A. 2017. *Machine Learners: Archaeology of a Data Practice*. Cambridge, Massachusetts: The MIT Press. https://doi.org/10.7551/mitpress/10302.001.0001.

Mager, A., M. Eitenberger, J. Winter, B. Prainsack, C. Wendehorst, and P. Arora. 2025. "Situated Ethics: Ethical Accountability of Local Perspectives in Global AI Ethics." *Media, Culture & Society* 47 (5): 1028–1041. 01634437251328200. https://doi.org/10.1177/01634437251328200.

Marres, N., M. Castelle, B. Gobbo, C. Poletti, and J. Tripp. 2024. "AI as Super-Controversy: Eliciting AI and Society Controversies with an Extended Expert Community in the UK." *Big Data and Society* 11 (2): 20539517241255103. https://doi.org/10.1177/20539517241255103.

Miceli, M., and J. Posada. 2022. "The Data-Production Dispositif." *Proceedings of the ACM on Human-Computer Interaction* 6 (CSCW2): 460:1–460:37. https://doi.org/10.1145/3555561.

Miceli, M., J. Posada, and T. Yang. 2021. "Studying Up Machine Learning Data: Why Talk About Bias When We Mean Power?" 6 (34): 1–14. https://doi.org/10.1145/3492853.

Mittelstadt, B. 2019. "AI Ethics – Too Principled to Fail?" *SSRN Electronic Journal*, ahead of print. https://doi.org/10.2139/ssrn.3391293.

Morley, J., L. Kinsey, A. Elhalal, F. Garcia, M. Ziosi, and L. Floridi. 2023. "Operationalising AI Ethics: Barriers, Enablers and Next Steps." *AI and Society* 38 (1): 411–423. https://doi.org/10.1007/s00146-021-01308-8.

Morozov, E. 2014. *To Save Everything, Click Here: The Folly of Technological Solutionism*. Reprint Edition. New York: Public Affairs.

Nadar, S. 2014. ""Stories are Data with Soul" – Lessons from Black Feminist Epistemology." *Agenda* 28 (1): 18–28. https://doi.org/10.1080/10130950.2014.871838.

Nyholm, S. 2022. *This Is Technology Ethics: An Introduction*. Hoboken: Wiley-Blackwell.

Phan, T., J. Goldenfein, M. Mann, and D. Kuch. 2022. "Economies of Virtue: The Circulation of "Ethics" in Big Tech." *Science as Culture* 31 (1): 121–135. https://doi.org/10.1080/09505431.2021.1990875.

van de Poel, I. 2013. "Translating Values into Design Requirements." In *Philosophy and Engineering: Reflections on Practice, Principles and Process*, edited by D. P. Michelfelder, N. McCarthy, and D. E. Goldberg, 253–266. Dordrecht: Philosophy of Engineering and Technology. Springer Netherlands. https://doi.org/10.1007/978-94-007-7762-0_20.

Pols, J. 2015. "Towards an Empirical Ethics in Care: Relations with Technologies in Health Care." *Medicine, Health Care, and Philosophy* 18 (1): 81–90. https://doi.org/10.1007/s11019-014-9582-9.

Powell, A. B., F. Ustek-Spilda, S. Lehuedé, and I. Shklovski. 2022. "Addressing Ethical Gaps in "Technology for Good": Foregrounding Care and Capabilities." *Big Data and Society* 9 (2): 20539517221113774. https://doi.org/10.1177/20539517221113774.

Raji, I. D., E. M. Bender, A. Paullada, E. Denton, and A. Hanna. 2021. "AI and the Everything in the Whole Wide World Benchmark." arXiv:2111.15366. Preprint, arXiv, November 26. https://doi.org/10.48550/arXiv.2111.15366.

Ratto, M. 2011. "Critical Making: Conceptual and Material Studies in Technology and Social Life." *The Information Society* 27 (4): 252–260. https://doi.org/10.1080/01972243.2011.583819.

Seaver, N. 2017. "Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems." *Big Data and Society* 4 (2): 2053951717738104. https://doi.org/10.1177/2053951717738104.

Sharon, T., and B. J. Koops. 2021. "The Ethics of Inattention: Revitalising Civil Inattention as a Privacy-Protecting Mechanism in Public Spaces." *Ethics and Information Technology* 23 (3): 331–343. https://doi.org/10.1007/s10676-020-09575-7.

Siffels, L. E., and T. Sharon. 2024. "Where Technology Leads, the Problems Follow. Technosolutionism and the Dutch Contact Tracing App." *Philosophy and Technology* 37 (4): 125. https://doi.org/10.1007/s13347-024-00807-y.

Sismondo, S. 2017. "Science and Technology Studies and an Engaged Program." In *The Handbook of Science and Technology Studies*, edited by U. Felt, R. Fouché, C. A. Miller and L. Smith-Doerr, and Society for Social Studies of Science, 13–32, 4th ed. Cambridge, Massachusetts London, England: The MIT Press.

Skirgård, H., H. J. Haynie, D. E. Blasi, H. Hammarström, J. Collins, J. J. Latarche, J. Lesage, T. Weber, A. Witzlack-Makarevich, S. Passmore, A. Chira, L. Maurits, R. Dinnage, M. Dunn, G. Reesink, R. Singer, C. Bowern, P. Epps, and J. Hill. 2023. "Grambank Reveals the Importance of Genealogical Constraints on Linguistic Diversity and Highlights the Impact of Language Loss." *Science Advances* 9 (16): eadg6175. https://doi.org/10.1126/sciadv.adg6175.

Suchman, L. 2023. "The Uncontroversial "Thingness" of AI." *Big Data and Society* 10 (2): 20539517231206794. https://doi.org/10.1177/20539517231206794.

Tacheva, J., and S. Ramasubramanian. 2023. "AI Empire: Unraveling the Interlocking Systems of Oppression in Generative AI's Global Order." *Big Data and Society* 10 (2): 20539517231219241. https://doi.org/10.1177/20539517231219241.

Taddeo, M., and L. Floridi. 2018. "How AI Can Be a Force for Good." *Science* 361 (6404): 751–752. https://doi.org/10.1126/science.aat5991.

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press

Taffel, S., L. Bedford, and M. Mann. 2022. Ecocide isn't ethical: Political ecology and capitalist AI ethics. *Economies of Virtue: The Circulation of 'Ethics' in AI*, edited by T. Phan, J. Goldenfein, M. Mann, and T. Kuch, 58–68. Vol 46, 1 edn. Amsterdam: Institute of Network Cultures.

Tronto, J. C. 1998. "An Ethic of Care." *Generations: Journal of the American Society on Aging* 22 (3): 15–20.

van Maanen, G. 2022. "AI Ethics, Ethics Washing, and the Need to Politicize Data Ethics." *Digital Society* 1 (2): 9. https://doi.org/10.1007/s44206-022-00013-3.

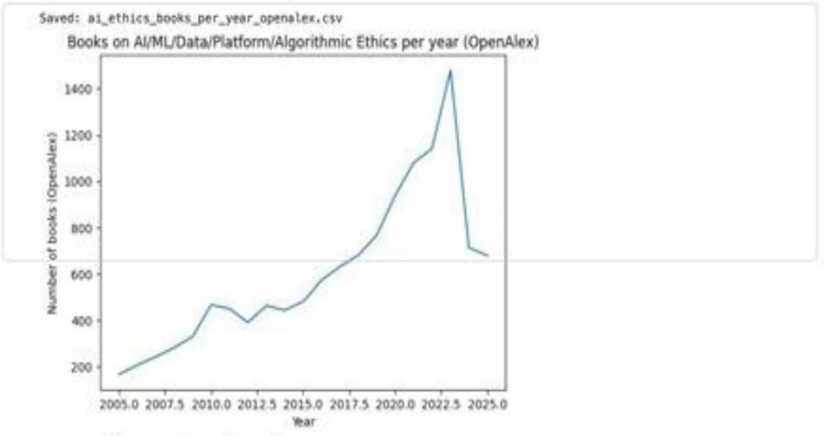
Verbeek, P. P. 2006. "Materializing Morality: Design Ethics and Technological Mediation." *Science, Technology, & Human Values* 31 (3): 361–380. https://doi.org/10.1177/0162243905285847.

Widder, D. G., S. Gururaja, and L. Suchman. 2024. "Basic Research, Lethal Effects: Military AI Research Funding as Enlistment." arXiv:2411.17840. Preprint, arXiv. November 26. https://doi.org/10.48550/arXiv.2411.17840.

Willow, A. J. 2023. "Introduction: Why Optimism?" In *Anthropological Optimism*, 1–30. New York/London: Routledge.

Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. Main edition. London: Profile Books.

Appendix



	year	books_count_openalex
11	2016	573
12	2017	632
13	2018	681
14	2019	768
15	2020	939
16	2021	1080

Cite this article: Helm, P. and Gerlek, S. 2026. Empirical AI Ethics: Grounded, relational, transformative. *Cambridge Forum on AI: Culture and Society*, **2**, e13, 1–15. https://doi.org/10.1017/cfc.2026.10023

https://doi.org/10.1017/cfc.2026.10023 Published online by Cambridge University Press