

Why putting artificial intelligence ethics into practice is not enough: Towards a multi-level framework

Big Data & Society
April–June: 1–14
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517251340620
journals.sagepub.com/home/bds



Hao Wang¹ and Vincent Blok^{1,2}

Abstract

Artificial intelligence (AI) ethics is undergoing a practical shift towards putting principles into design practices in developing responsible AI. While this practical turn is essential, this paper highlights its potential risk of overly focusing on addressing issues at the level of individual artifacts, which can neglect more profound structural challenges and the need for significant systemic change. Such oversight makes AI ethics lose its strength in addressing some hidden, long-term harms within broader contexts. In this paper, we propose that the reflection on structural issues should be an integral part of AI ethics. To achieve this, we develop a multi-level framework to analyze socio-ethical issues of AI at both the artifact and broader structural levels. This framework can serve as a potentially transformative approach to uncover some unspoken assumptions in current AI ethics discourses and expose some blind spots in AI guidelines, policies, and regulations. Our paper paves the way to develop a practical approach that can effectively integrate this multi-level framework into real-world AI design and policymaking, ultimately bringing about transformative change.

Keywords

AI ethics, structural issues, responsible AI, trustworthy AI, value-sensitive design

Introduction

Artificial Intelligence (AI) is significantly shaping our modern world, but it also brings a series of ethical, legal, and social challenges. In response, there has been a global push to establish ethical guidelines to regulate AI's risks and adverse outcomes (Corrêa et al., 2023; Jobin et al., 2019; HLEG, 2019). However, such AI governance through ethical guidelines has received a lot of criticism. One major concern is that these guidelines and principles are often too abstract to apply in real-world situations (Bleher and Braun, 2023; Hagendorff, 2021; Mittelstadt, 2019; Veale, 2020). These principles, for instance, often do not have many real enforcement mechanisms, so there are usually no or very minor consequences for not following ethical codes—they are “toothless” (Rességuier and Rodrigues, 2020). Some companies even use such abstract ethics as a marketing tool or a way of “ethics washing” to avoid stricter legal regulations (Wagner, 2018). As a result, this principle-based AI ethics is often seen not just as “useless” but as “a dangerous distraction,” which takes significant funding and resources away from better uses (Munn, 2022).

Therefore, AI ethics is undergoing a *practical turn*, trying to translate abstract values and principles into concrete

design practices (Floridi, 2019; Hagendorff, 2022). This by-design AI ethics aims to integrate human values seamlessly into the entire lifecycle of AI development. The Ethics by Design (EbD) approach, for example, “starts with high-level values for AI, which are then translated into design requirements for AI systems, further translated into specific measures to be undertaken at specific points in the design process” (Brey and Dainow, 2023: 3). This by-design approach can also be found in some similar methodologies such as “Ethics in/for Design” (Dignum, 2018: 2), “Design for Values” (Buijsman et al., 2025), or in the application of value-sensitive design (VSD) to create “value-sensitive AI” (Sadek et al., 2023; Umbrello & van de Poel, 2021), among others. There is also a growing emphasis on building the Ethical, Legal, and Social Aspects

¹Philosophy Group, Wageningen University, Wageningen, The Netherlands

²Philosophy Group, Erasmus University, Rotterdam, The Netherlands

Corresponding author:

Hao Wang, Philosophy Group, Wageningen University, Wageningen, The Netherlands.

Email: hao2.wang@wur.nl



(ELSA) Lab approach to collaboratively design responsible, human-centric AI (Ryan and Blok, 2023), “making sure it is ethical in practice” (NLAIC, 2025). In this paper, we acknowledge that to make AI ethics useful and actionable, it may require a shift from abstract values/principles to some practical design. We believe that practical approaches like VSD, EbD, and ELSA Labs are essential for addressing ethical and social issues in specific AI development, especially in their efforts to integrate values into every aspect of AI artifacts. However, our hypothesis is that we should also be cautious about this practical turn as it tends to focus on the design of *individual artifacts*, which may ignore deeper *structural issues* that are often too abstract and complex to be addressed by technical design alone.

In this paper, structural issues refer to problems caused by patterned relations, often produced and reproduced through multiple, large-scale processes where no identifiable agent directs, controls, or intends (Haslanger, 2023; Ryan et al. 2024; Young, 2011). The structural issues are distinct from those at the artifact level, posing some different forms of harm that AI ethics should address. Consider Marilyn Frye’s metaphor of a birdcage (1983). Imagine a bird kept in a birdcage. Examining one wire individually does not explain why the bird cannot fly. It is only when considering the entire arrangement of wires—how they are connected and reinforce each other’s rigidity—that we understand why the bird’s flight is restricted. Similarly, by-design AI ethics often breaks down AI systems into individual components, like data, algorithms, and generated content. The assumption is that fixing the ethical issues in each of these components will automatically make the whole AI system responsible and trustworthy. However, this approach is like focusing on one wire at a time in Frye’s birdcage metaphor—it often gets lost in the details of individual AI artifacts while neglecting the systemic effects caused by “structures.” To truly develop responsible AI, it is *not sufficient* to just address artifact-level issues, while ignoring the structural harms that AI systems produce and reproduce.

Without considering underlying structural issues, AI ethics may risk losing its ability to address some relevant harms in larger contexts, missing opportunities for transformative change. They may end up only treating the “symptoms” rather than addressing the deep structures and root causes that give rise to these symptoms. They will thus fix bias in datasets, for instance, but fail to challenge the racist and patriarchal structures that create such biases. Also, by ignoring structural issues, by-design ethics could inadvertently become a form of “ethics washing” (Van Maanen, 2022). Some AI developers may claim their technologies are trustworthy because they have addressed all possible issues like privacy, transparency, and biases. However, they may have just solved some narrow concerns within their AI technologies, while marginalized groups keep suffering from broader structural injustices

that AI reinforces within existing unequal power systems. This creates the illusion that the problem is “solved” (in a narrow sense) and that AI appears to be “trustworthy,” but in reality, it masks underlying power imbalances, further marginalizing those already affected by structural oppression (Birhane et al., 2022). To address harms and promote an equitable society, we propose that reflecting on structural issues should be *an integral part* of any by-design AI ethics.

Highlighting the necessity to consider structural issues is not new. In the literature on AI and data ethics, many authors have already mentioned different structural issues, either directly or indirectly. Several scholars have argued the importance of practices that go beyond focusing only on artifacts and individual harms (Attard-Frost et al., 2023; Crawford, 2021; Hagendorff, 2022; Smuha 2021). Some have critically pointed out that current ethical AI discourse often overlooks *relational* aspects of AI ethics (Metcalf et al., 2023), instead “co-opting the language of critics and folding them into a limited, technologically deterministic, expert-driven view of what ethical AI/ML means and how it might work” (Greene et al., 2019: 2122). Some other studies explore how AI mirrors or perpetuates broader socio-political structures, such as capitalism (Zuboff, 2019), colonialism (Birhane, 2020; Couldry and Mejias, 2018; Muldoon and Wu, 2023), sexism (Rafanelli, 2022), racism (Benjamin, 2019), and caste systems (Sambasivan et al., 2021). Concerns have also been raised about AI’s fundamental implications caused by ontological structures for human existence in the digital age, such as dataism (Blok, 2023b), AI alienation (Haga, 2022), anthropocentrism (Hagendorff, 2022), AI instrumentalism (Gill, 2020), de-humanization (Fritts and Cabrera, 2021), and objectification (Wang, 2023).

There seems to be, however, a lack of in-depth analysis of structural issues in AI ethics. The term “structural issue” is often used broadly, but what exactly is a “structure,” and what makes an issue “structural”? If it is a matter of structure, how do structural issues differ from non-structural ones, and how are they related? Why should we morally care about structural issues if they are not intended and caused by any individual agent? These questions have not yet been theoretically clarified in the context of AI ethics. In addition, while existing AI ethics literature rightly highlights the importance of addressing structural issues, it tends to treat them as if all structural issues exist at the same level. However, a closer look will show that AI’s structural effects stretch across a *spectrum of levels* with different degrees of complexity and focus. Understanding these nuances of structural issues is crucial for AI ethics, as it helps clarify how different structural effects relate to different responsibilities in developing responsible AI. Therefore, this paper will make a detailed analysis of structural issues in AI ethics and propose a multi-level approach grounded in social, political, and philosophical literature. We will then perform a first try to explore how to implement this approach in responsible AI design.

This paper starts with a review of AI ethics literature to illustrate the structural impacts of AI that are often neglected in by-design AI ethics. To guide this review, we applied the input/throughput/output analytical framework, originally proposed, and proven effective in uncovering underlying assumptions in the responsible innovation and research literature (Blok & Lemmens, 2015). The framework views the responsible development of AI as a process with three analytical stages: input, throughput, and output. Input refers to the goals and values that shape the starting point of AI design. Throughput looks at how these elements are integrated into actual design process by engaging with stakeholders. Output focuses on the outcomes of the design. In this paper, the first three sections critically analyze a series of issues emerging during the phases of input (values identification) (Section Challenging the input of by-design AI ethics: translating uncritical values into actions), throughput (actual design) (Section Challenging the throughput of by-design AI ethics: participation without structural change), and output (outcomes evaluation) (Section Challenging the output of by-design AI ethics: unpredictable outcomes). We then show that most of these issues are structural in nature, yet they are often ignored in AI ethics. To fill this gap, we develop a multi-level framework to clearly analyze AI impacts, from specific artifact-level concerns to broader structural implications (Section The development of a multi-level framework for AI ethics). Moving further, we propose a practical method to implement this multi-level approach in responsible AI designs (Section Implementing multi-level framework as a diagnosis of socio-ethical issues).

Challenging the input of by-design AI ethics: translating uncritical values into actions

By-design AI ethics often begins with abstract inputs, such as ideal goals (e.g. developing trustworthy and human-centered AI) or related values and principles that should be integrated into the design process. However, when translating these goals, values, and principles into real-world design, they are often treated as self-evident, as if they just need to be implemented. In reality, these inputs are more complex than they appear. In this section, we will critically examine how by-design AI ethics tends to overlook the broader effects of techno-solutionism, power dynamics, and the disruption of everyday life when translating these values into design practices.

Techno-solutionism

By-design AI ethics are often driven by a pragmatic approach to values, focusing on values that can be operationalized and incorporated into specific AI designs. While this

approach is practical, it can oversimplify complex social issues that require broader political engagement and actions within a larger social context, reducing them to technical attributes that can be fixed (Hagendorff, 2020; Greene et al., 2019). This techno-solutionism in selecting and integrating values can lead to the neglect of deeper structural problems, which are often seen as too abstract or complex for AI designers to address.

The goal of trustworthy AI, for example, is often reduced to a few basic values like privacy, transparency, security, and accountability (HLEG, 2019). While these aspects are essential for ethical AI development, they are also the ones that “are most easily operationalized mathematically and thus tend to be implemented in terms of technical solutions” (Hagendorff, 2020: 103). Even when values with significant social and political weight are integrated, they are often simplified into technical issues that can be fixed within the design of a specific AI product. Take “fairness” and “justice,” for instance. In by-design ethics, they are often flattened to technically fixable issues of “purely objective” datasets and algorithms (Benjamin, 2019). This leads to attempts to fix the problem, such as debiasing datasets using tools like “AI Fairness 360” (Birhane et al., 2022). However, this reductionist approach overlooks the fact that data bias is not just a technical issue but is embedded in broader structures of historical inequality (e.g. racism, sexism, post-colonialism) (Madaio et al., 2022). When social injustice is framed narrowly as unfairness in data and algorithms, AI ethics fails to address its root causes and neglects the larger social context and power structures that contribute to the marginalization of certain groups.

Power dynamics

When translating values into practice, by-design ethics tend to assume that the values are widely shared and ready to implement. However, selecting which values to prioritize in AI development is often shaped by power dynamics. It might be true that choosing those values is not random, as it is often claimed to be based on a broad international consensus or in line with basic human rights (HLEG, 2019; Brey and Dainow, 2023). But it is still up for debate whether there are universal human rights accepted by all cultures, especially since this idea of value universalism has been criticized for being West-centrism (Zwart et al., 2024). For instance, the value of autonomy, which frequently appears in AI ethics guidelines, is deeply rooted in Western liberalism and individualism. In cultures that emphasize collectivism, autonomy may hold a different meaning or less prominence (Nakada and Tamura, 2005). Also, ethical guidelines for AI, particularly those from governments and private companies, are predominantly shaped by Western countries and corporations leading the AI industry (Jobin et al., 2019). Voices from the global South are often underrepresented in these discussions (Birhane,

2020; Wakunuma et al., 2021). So, it is somewhat naïve to claim a widespread consensus on current AI ethical values, especially given the limited participation and influence of the global South in shaping these values in AI design.

The disruption of everyday life

In AI ethics literature, the ideals of human-centered AI and trustworthy AI are often portrayed as inherently good things without much questioning. However, this optimistic framing can obscure the complex lived experience of individuals who may struggle within AI systems every day. To fully understand trust in AI, for example, we should move beyond a list of abstract values like transparency, privacy, and non-discrimination, to consider the emotions, fears, frustrations, and daily struggles that fuel distrust. People may fear losing control, feel anger over political manipulation, or worry about being replaced by machines. This erosion of trust does not happen overnight—it builds over time, driven by scandals, irresponsible practices by tech companies, and insufficient AI regulations. Distrust, therefore, is not just a reaction to specific issues like privacy violations or lack of transparency; it reflects a broader disruption of people's everyday experiences with AI.

Given this complexity, promoting trust in AI may not always be the best approach. In some cases, fostering a healthy skepticism or even justified distrust might be necessary. Some have argued that for marginalized groups, like Black people facing systemic racism, social distrust is justified (Davidson and Satta, 2021). This justified skepticism can also extend to AI systems, particularly in the context of pervasive surveillance capitalism and the relentless datafication of human lives into commercial products (Zuboff, 2019; Crawford, 2021). Such distrust does not aim to erode societal trust but rather to push for the creation of conditions that genuinely merit it (Davidson and Satta, 2021: 23). In this way, distrust becomes a way of challenging systemic issues, advocating for accountability, and ultimately making our algorithmic society more trustworthy (Wang, 2022a).

Overall, this section criticizes that by-design AI ethics often overlook some important issues like techno-solutionism, power dynamics, and the disruption of everyday life during the input phase of AI design. They often focus too much on translating abstract values into implementable design actions without taking a closer look at the values themselves.

Challenging the throughput of by-design AI ethics: participation without structural change

The throughput phase of by-design AI ethics involves the actual AI design process in collaborating with various stakeholders. This co-design process emphasizes participation and its potential for transformation. As a “reflection-in-action,” it enables the

public to surface relevant issues, reveal hidden power dynamics, and advocate for meaningful institutional change (Robertson and Simonsen, 2012: 5). However, in this section, we will show that the transformative potential of participation can be undermined and distorted if AI ethics only focus on designing specific AI artifacts while failing to consider important structural change. This oversight runs the risk of reducing participation to mere labor extraction, the manipulation of legitimacy, and a disconnection from people's lived experiences.

Participation as an extraction of labor

Ideally, by participation, AI ethics would include feedback from the public or affected users in the AI design process (Hansson, 2017). This allows the public to serve as an additional resource for AI designers to identify ethical issues and improve ethical compliance in the design process. However, this idealized view of participation may ignore the economically extractive structure underlying real-life AI designs. For example, in many interdisciplinary projects, ethical/legal experts and citizens may act in consultation roles to identify the potential ethical, legal, and social issues in particular AI systems. However, this consultation might be limited to some particular forms that only align with business objectives. On the one hand, companies may use stakeholder input to extract business information that primarily serves their business objectives (Brand and Blok, 2019). On the other hand, when social values conflict with companies' focus on economic value creation and profit-seeking, social values are often dismissed as less relevant in AI designs (Blok and Lemmens, 2015: 22). As a result, this participation may reduce the consultation role of ethical/legal experts and citizens to some free labor, indirectly helping AI companies maximize their profits.

This form of participation exemplifies what Mona Sloane calls “extractive participation,” where human labor is utilized to improve existing AI systems without proper acknowledgment or compensation (Sloane, 2024). This extractive participation extends beyond the consultation role of experts and citizens. A more direct example is the employment of low-paid workers to carry out repetitive tasks, such as labeling and flagging unethical content (Perrigo, 2023). Additionally, this extraction occurs in everyday life, where users' interactions with AI products like ChatGPT generate continuous feedback and data, enabling companies to refine their technologies without recognizing or compensating users for their contributions (Sloane, 2024).

Participation as a manipulation of legitimacy

In an ideal scenario, participation acts as a bridge between AI designers and the public, fostering trust and ensuring democratic legitimacy in AI development. In this vision, participation is not just about identifying ethical challenges

but about embedding inclusion and legitimacy into the development process (Ten Holter, 2022). However, this idealized picture often overlooks the power dynamics inherent in the co-design process, where participation can be manipulated by the powerful to acquire legitimacy. This distorted form of participation is what Sherry Arnstein calls “manipulation” in her famous participation ladder model (2019). Arnstein argues that when power imbalances exist, participation can become illusory: citizens may believe they have a say, but the process is structured to deny them any real power. Such forms of participation, disguised as inclusive and democratic, often serve to pacify dissent by offering symbolic rather than substantive involvement (Gilman, 2022).

In real-world AI design, these power imbalances are prevalent, where participation processes are often overdetermined by those with the most power of influence (Blok and Lemmens, 2015). This asymmetrical structure allows powerful entities to manipulate the design process to “acquire societal and ethical legitimacy” (2015: 25). For instance, interdisciplinary participation is frequently criticized for treating ethics as an afterthought or an add-on to innovation programs—used to validate existing practices rather than genuinely address social challenges (Zwart and Nelis, 2009; Ryan and Blok, 2023). This manipulation is especially harmful to those already marginalized groups, whose active participation may only be seen as mere data points or statistical abstractions for AI companies seeking ethical proof. After participation, their marginalized situation is still unchanged. This manipulative dynamic can even create a vicious cycle: when participants see that their involvement fails to challenge existing power structures, they become more skeptical of the process, making meaningful and transformative engagement even harder to achieve (Wang, 2022b).

Participation as a detachment from the lifeworld

When participation is not distorted by economic and power structures, and when powerholders really want to engage citizens and other stakeholders in AI design decisions, there is still a concern about overlooking the disruptive structure of datafication on the lifeworld where participants live and act in their daily lives. For instance, when participants are asked to identify relevant ethical issues in co-design processes, they are often prompted to base their choices on existing abstract values such as privacy, transparency, accountability, fairness, and so on (Boenink and Kudina, 2020). However, humans’ everyday concerns are not just abstract or theoretical concepts, but rather involve real, subjective feelings experienced in daily life. These feelings are easily dismissed as individual complaints and seen as irrelevant in AI design, but they might actually reflect a bigger structural issue in the widespread use of AI in our societies like *the datafication of the lifeworld*. This datafication involves the process of

transforming every aspect of our daily lives into data that can be quantified, analyzed, and used for decision-making, prediction, and optimization (Coudry and Mejias, 2018; Zuboff, 2019). That means when we design AI technologies, we are not just designing individual artifacts but contributing to the broader process of datafication that shapes and disrupts the meaning of the world.

During stakeholder meetings, for example, some farmers (especially smallholders) often express a feeling of powerlessness when it comes to using AI (van der Burg et al., 2022). They have to decide whether to adapt, learn new skills, or risk falling behind. This feeling that AI is an unstoppable force is often considered a personal issue, not directly tied to any specific AI technology or values outlined in existing ethical frameworks, so it is often disregarded in AI designs. However, this personal feeling of being compelled to adopt and integrate AI on their farms may actually reflect a larger issue: the disruptive impact of datafication on farmers’ traditional farming practices. Datafication is not just about creating a digitalized environment; it is about redesigning the meaning of everyday life (Blok, 2023b). It cultivates the new assumption that by transforming all aspects of daily life into data, we can gain valuable insights and optimize processes for improved efficiency (Korenhof et al., 2021). But for smallholders, they may not necessarily embrace these efficiency-focused and optimization-driven approaches. Instead, they may prioritize their connection to the community and practice eco-friendly methods, even if they are not as efficient. This broader concern prompts AI designers to think carefully about some different requirements in their design practices, like considering potential disruptions to local farming practices in developing AI technologies, and ensuring that the use of AI in farming respects the way people already do things and their community traditions.

Overall, by-design AI ethics often highlights the importance of active and inclusive participation in the AI design process. However, as we have elaborated, if the unequal power structure does not shift and the disruptive structure of datafication is not properly considered, participation may fall into a trap of extraction, manipulation, or detachment from people’s everyday experiences.

Challenging the output of by-design AI ethics: unpredictable outcomes

The output of by-design AI ethics is often assumed to produce predictable results, such as mitigating the negative impacts of AI and promoting social benefits by integrating values into AI design. In this section, however, we challenge this belief of predictable AI outcomes by critically examining its underlying assumptions of linear reductionism, anthropocentric designer agency, and immaterialism.

Linear reductionism

By-design AI ethics often assume a linear reductionism that an AI system can be breakable into different components of AI products like data, algorithm, and generated content, and if we fix ethical and social issues associated with each component (Dignum, 2018; Brey and Dainow, 2023), the entire AI system will automatically produce responsible outcomes. However, AI is more than just an individual artifact/system but constitutes a complex socio-technical ecosystem marked by nonlinearity and uncertainties. This intricate entanglement with larger social dynamics and structures makes AI inherently unpredictable.

AI is deeply intertwined with socio-political structures, forming a complex system that cannot be directly broken down into independent parts or linear processes. As Carsten Stahl shows, this AI as a complex ecosystem has the intricate interplay among its various components, often exhibiting highly *nonlinear* relationships (2021). As a result, the intervention of one aspect of the ecosystem can often stir up unexpected outcomes that do not match what was initially planned. A typical example is the Jevons Paradox or rebound effect in the context of AI: AI's boosted efficiency in performing tasks may actually trigger more demand for those tasks, resulting in consuming more resources (York and McGee, 2016). This nonlinear effect can be even stronger when assessing AI's long-term impacts. Imagine trying to gauge the impact of Gutenberg printing his first Bible in 1455. Back then, no one could foresee that it would challenge the authority of the Catholic Church and pave the way for modern science and new industries (Naughton, 2023). If AI technology is as transformative as claimed, its long-term impacts can be even more profound and unpredictable.

Anthropocentric designer agency

By-design AI ethics often assume that humans have ultimate control over how AI systems are developed and used, so humans can steer them toward social benefits. They believe that humans can shape AI technologies and control the outcomes in ways that align with ethical, moral, and social values, ensuring that AI serves the greater good (van den Hoven, 2013). This belief reflects the high degree of *anthropocentric* designer agency (Donia and Shaw, 2021), where technologies are often seen as passive objects awaiting (re)design by human designers, while humans are viewed as the exact opposite of objects, characterized as active, thinking subjects who can control technologies (Rosenberger, 2014). However, the historical trajectory of innovation reveals that design is not solely dictated by human actors but sometimes follows its own course or inherent trends in technological progress.

Gilbert Simondon's notion of the "technological milieu" can help explain this idea of inherent trends more clearly.

He suggests that innovations do not materialize in isolation but are molded by the broader context of technological advancements over time (2017). The design of a new smartphone, for example, is not simply determined by the intentions of its designers but is also influenced by advancements in materials science, semiconductor technology, and wireless communication protocols. This example suggests that AI design decisions are not totally determined or controlled by designers' intentions but sometimes are shaped by the cumulative effects of past innovations, ongoing research, and technological infrastructure. Our intention is not to imply that AI technology develops autonomously or independently. Instead, we highlight the co-evolutionary relationship between humans and AI technologies, where design decisions are influenced by both human agency and inherent technological tendencies (Simondon, 2017). This relationship means that AI technology cannot be entirely directed and designed by human desires or intentions. The ethical and social outcomes of AI are often not predicated and controlled only by designers.

Immateriality

By-design AI ethics often focus on the digital and immaterial aspects of AI, such as data, algorithms, and generated content, while they tend to ignore the material production relationships and underlying physical infrastructure that make all these digital elements possible. They often see AI as something that emulates human cognitive abilities and exists only in the "cloud" (Wyatt, 2021), resulting in a misleading impression that AI is only about digital, virtual, and immaterial aspects (Newlands, 2021; Brevini, 2023). This myth of immaterialism obscures some important ethical and social impacts of AI, making them often invisible to the public. Three often overlooked material issues of AI include its labor exploitation, environmental costs, and the extraction of everyday life.

Research has revealed that so-called AI systems often depend on the labor of millions of low-paid "ghost" workers worldwide. These workers perform repetitive tasks under precarious conditions, yet their contributions are deliberately obscured from public view (Gray and Suri, 2019; Perrigo, 2023). By concealing this human labor, AI companies create the illusion that AI is highly advanced, fully automated, and objective, which helps them evade regulation and attract greater commercial investment (Newlands, 2021: 8).

Beyond hidden labor, the environmental costs of AI are also often shrouded in secrecy by corporations (Crawford, 2021). Recent studies, however, have shown that AI systems consume significant amounts of energy and emit substantial greenhouse gases during training, fine-tuning, and usage (Luccioni et al., 2023). Also, they rely on industrial processes such as mining natural resources, manufacturing technical components, and managing electronic waste (Wynsberghe et al., 2023).

Lastly, AI is not merely about running abstract data and algorithms; it fundamentally depends on extracting human experiences from everyday life and converting them into data commodities. Critics argue that this commodification of life data should not occur arbitrarily, as these data are deeply tied to human experiences and emotions (Zuboff, 2019). These experiences are essential for shaping individuals' sense of identity and their meaningful relationships with others (Roessler, 2015). Despite this, current design-based AI ethics often fail to address the profound impact of continuous real-time surveillance on people's understanding of themselves and the world.

In summary, this section challenges the widespread belief in current design-based AI ethics that AI impacts are predictable and can be directed toward social benefits. To do this, we critically examine three core assumptions that underlie the belief in predictability. Along with the previous two sections, we have explored concerns arising from the input (section Challenging the input of by-design AI ethics: translating uncritical values into actions), throughput (section Challenging the throughput of by-design AI ethics: participation without structural change), and output (section Challenging the output of by-design AI ethics: unpredictable outcomes) phases of by-design AI ethics. Many of these concerns center on structural issues that are often neglected in AI ethics practices (see Table 1). For the remainder of the paper, we will focus on clearly understanding the structural features of these issues.

The development of a multi-level framework for AI ethics

The previous sections provided a review of AI ethics in the practical turn. We identified two types of socio-ethical issues: one relates to specific AI artifacts, which by-design ethics often focus on, while the other concerns broader challenges that are often neglected in AI ethics. Most of these broader issues are structural in nature. In this section, we will explore what makes issues structural, how structural issues relate to artifact-level ones, and why AI ethics should integrate both in developing responsible AI. To achieve this, we turn to social, political, and philosophical theories to build a multi-level approach to AI ethics.

We start with clarifying what "structure" means. As Sally Haslanger (2023) points out, in social and political philosophy, "structure" is often used interchangeably with "system." A structure or system roughly means a whole that is greater than the sum of its individual parts. For example, a forest is more than just a collection of trees, and a band is more than individual musicians. In the case of a band, it includes not only the members but the formal roles (like vocalist or drummer) and the informal elements, such as shared beliefs, tastes in music, and playing styles, which

collectively define what the band is. In this context, a structure can be seen as an abstract representation of patterned relations, or, as William Sewell describes it, a "schema" that defines what it means to form a cohesive whole (Sewell, 2005). However, Iris Young (2001, 2011) and Haslanger (2023; 2012) argue that structures are *not just passive, abstract representations but actually actively shape the realities that we live and act*. For instance, these structures "fundamentally condition the opportunities and life prospects of the persons located in those positions" (Young, 2001: 14). So, following Young and Haslanger, we view a structure as the combination of patterns and related real-world actions, which exerts an active force to influence resource distribution and lived experiences.

As for Young, an issue is "structural" only because it is caused by the existence of structures, rather than from the actions or wrongdoing of any single individual. She argues that individuals are not responsible for the initial production of these structures, because structures often persist long before they live and after they die. That means the production of unjust structures, such as racial hierarchies, is often unintentional or "even run(s) counter to" the individual intentions (Young 2011: 63). However, individuals do play a role in the ongoing *reproduction* of these structures. Young draws on Anthony Giddens's theory of "structuration" (1979) to explain this dynamic: "When individuals act, they are doing two things at once: 1) They are trying to bring about a state of affairs that they intend, and 2) they are reproducing the structural properties, the positional relations of rules and resources, on which they draw for these actions" (Young, 2011: 60). For example, the structural injustice of sweatshop labor exists independently of any single person. However, when someone buys clothes from a high-street shop, they inadvertently perpetuate the practices that sustain this injustice. Structures act as the preconditions for individual actions, yet those actions, in turn, help reproduce these structures.

Importantly, this reproduction does not mean individuals are the cause of the structural problem itself. Instead, structural issues are the result of complex, large-scale processes that involve many actors and systems over time. Let's revisit the example of sweatshop labor as a structural issue. At its core, sweatshop labor is maintained by the structural dynamics of production and consumption in the global capitalist economy. These dynamics are part of multiple interconnected systems. As Young (2006) explains, workers in developing countries occupy a structural class position relative to the small entrepreneurs who employ them. These factory owners and managers, in turn, are dependent on contracts from multinational clothing companies, which are themselves locked in fierce competition within the global garment industry. Consumers are similarly entangled, influenced by the fashion industry's creation of a "need" for new clothes. This desire for up-to-date fashion becomes internalized and perpetuated through habits and social

Table 1. Overlooked socio-ethical issues in by-design AI ethics.

By-design AI Ethics	Explanation	Assumption	Neglected Issue
Input	The selection of which values should be translated into AI artifacts.	Those values and principles are treated as pre-given entities, ready-made for translation.	<ul style="list-style-type: none"> • Technical/AI solutionism: • Power asymmetries • Dataism/Pan-computationalism
Throughput	The actual AI design process in which stakeholders actively participate.	The participation of diverse stakeholders in co-design is equally shaping AI development.	<ul style="list-style-type: none"> • Extraction of labor • Legitimacy manipulation • Detachment from the lifeworld
Output	Evaluation of the ethical, legal, and social outcomes of AI design.	The outcomes can be anticipated and steered by human agents.	<ul style="list-style-type: none"> • Anthropocentric designer agency • Linear reductionism • Invisible labor, environmental cost, and extraction of everyday life

AI: artificial intelligence.

norms—what Young refers to as “habitus” (Young, 2011: 61). So, it may seem that individuals can avoid contributing to these unequal relationships by making different choices, such as buying only fair-trade or second-hand clothing, but such actions by individuals *alone* cannot disrupt the underlying structural forces that perpetuate the inequality.

Based on the above analysis, we can define a structural issue as follows: *An issue is structural when it is caused by patterned relations or conditioning effects that are often produced and reproduced through multiple, large-scale processes, with no single agent directing, controlling, or intending the outcome.*¹

Our focus on defining and addressing structural issues is not meant to replace the consideration of individual issues but rather to call for a multi-level approach to analyzing them. Young, for instance, has already highlighted the need for a “two-level” analysis in understanding the injustice of sweatshop labor (2004: 377). On the individual level, factory managers and owners can be held morally or legally accountable for specific violations of human rights or labor laws (2006: 116). On the structural level, however, all individuals whose actions reproduce the practice *share* a political responsibility to work toward structural change. These individuals are not liable for the initial production of these structures but have a forward-looking obligation to participate in collective action to transform unjust practices. So, according to her, the concept of political responsibility, resulting from the reproduction of unjust structures, should complement, not replace, the liability model of responsibility. This two-level approach can also be found in other social and philosophical literature. For instance, in Wright Mills’ influential book, *The Sociological Imagination* (1959), he criticizes how social problems are often wrongly addressed as isolated individual issues rather than being seen in the broader structures that shape them. Similarly, as noted in the Introduction, Frye (1983) uses the birdcage metaphor to illustrate how structural factors interconnect to create barriers.

Now, we can see how this multi-level analysis can be applied to AI ethics. In the previous sections, we identified some overlooked socio-ethical issues, such as technical/AI solutionism, power asymmetries, and dataism/datafication. Most of these are structural issues in the sense that they are not caused by any individual designers or companies. Instead, they are produced and reproduced through large-scale, interconnected social, economic, and technological processes. For example, “AI solutionism”—the tendency to treat complex social problems as solvable through technical solutions—emerges not from a single designer’s intent but from the broader culture of technological optimism (technophilia) and market-driven innovation. Similarly, “power asymmetries” in AI systems are produced by widespread socioeconomic inequalities, regulatory gaps, and the concentration of decision-making power in a few Big Tech companies. “Dataism,” the ideology that prioritizes data as the ultimate solution to social issues, is reinforced by academic trends, commercial incentives, and political agendas that emphasize data-driven approaches. These structural issues cannot be adequately addressed by focusing only on individual accountability or isolated policy changes. Instead, they require a broader, multi-level analysis that considers how these problems are embedded within and perpetuated by systemic patterns.

Moving further, we want to enhance this multi-level analysis by distinguishing four levels of issues rather than relying solely on the dichotomy of individual versus structural. AI’s structural influence reveals a *spectrum* of structural challenges spanning different degrees of complexity and scope. For instance, some structural issues are socio-political, such as data racism, data sexism, and algorithmic colonialism. These are rooted in existing unequal social and political structures. On the other hand, some other issues like dataism or the loss of meaning are more ontological in nature. They relate to our basic understanding of the world and reality—our beliefs, mindsets, and worldviews. These ontological structures are more stable and persistent,

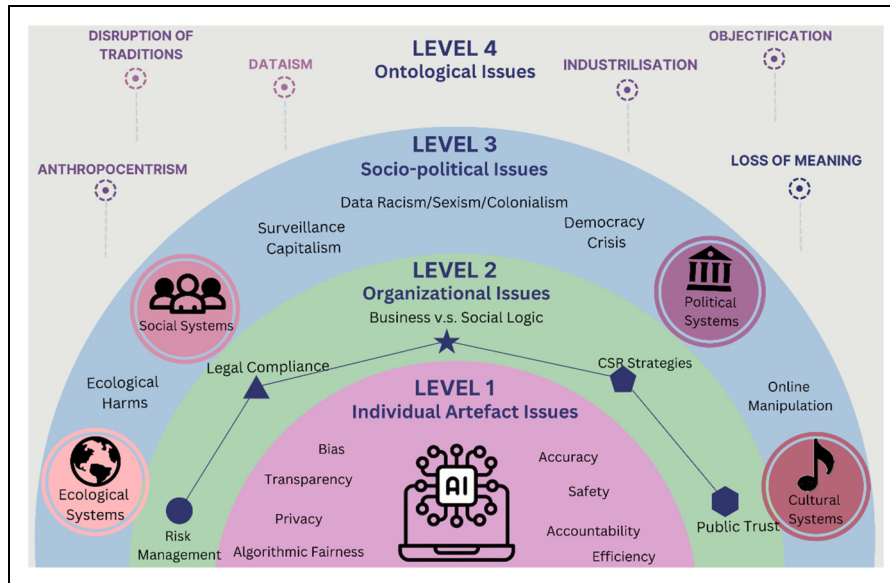


Figure 1. Relevant socio-ethical issues of AI across multiple levels.
AI: artificial intelligence.

reflecting larger, more foundational processes that are even more “structural” than socio-political structures. Also, we can identify an intermediate level between the individual and structural: the organizational level. This level concerns digital platforms or AI companies that may not create structural outcomes but can cause group effects that go beyond individual actions. The organizational issues include, for example, tensions between profit-making and social objectives, or the challenge of building public trust. These issues represent a transitional layer between individual actions and structural forces.

Hence, we can develop a refined multi-level framework in understanding the AI’s ethical impacts, ranging from individual artifact level, to organizational, socio-political, and ontological levels (see Figure 1). If an issue is *primarily* caused by the development and use of specific AI technologies, we categorize it as an **artifact-level** issue. For example, the Facebook app may cause data privacy issues, as seen in the Cambridge Analytica scandal (Hu, 2020). The FICO Score (a credit scoring system in the US) may create transparency issues if users are unclear about why their credit applications are rejected (Wang, 2022b). By addressing these specific issues, the responsibility of this particular AI artifact can be largely improved. At the **organizational level**, socio-ethical issues focus on how organizations (companies or institutions) integrate, manage, and govern AI systems. In the case of the Cambridge Analytica scandal, for example, the issue at this level is not just about addressing privacy concerns on a technical level. More importantly, it involves gaining public trust, changing the organization’s structure or culture, and thinking about how to make profits while complying with ethical values.

At the **socio-political level**, socio-ethical issues deal with the broader social, ecological, and political impacts of AI systems. We will take a bigger look at for instance how Facebook, embedded in the wider structure of surveillance capitalism, disrupts the socio-political landscape by turning private human experiences into a commodity for exchange and undermining democracy (Zuboff, 2019). The issues at the **ontological level** address deeper philosophical and existential concerns about how AI reshapes the nature of human existence, identity, and our basic understanding of reality. For instance, when applying digital technologies in smart cities, it is doing two things at once: i) it is using sensors and collecting data to create individual artifacts, and ii) it reproduces an *ontological structure of datafication*, where interconnected technologies form an ensemble of digital ecosystem where humans live and act (Blok, 2023a: 5). This datafication structure on the one hand creates a new infrastructure for daily life, where all parts of lives are datafied or reduced to data. On the other hand, it reshapes our basic assumptions and beliefs that data is the core building block of reality and that everything, including human and non-human beings’ behavior, can be quantified and understood through data analysis. These beliefs fundamentally alter how we perceive and interact with the world, which in turn affects how we carry out real-life practices in AI ethics, governance, and design.

By adopting this refined multi-level analysis, we can specify which types of issues belong to each category and recognize the diversity within structural issues themselves. This nuanced understanding is crucial for AI ethics, as identifying these levels more precisely can clarify the responsibilities involved in developing responsible AI. However, it is important to avoid “reify(ing) the

Table 2. A multi-level impact matrix to identify socio-ethical issues of AI.

Level	Impact	Issue	Diagnostic question
1. Artifact	1.1 Input Data	<ul style="list-style-type: none"> Bias or lack of representation in data 	<i>How do AI developers ensure that the input data for the AI systems is representative, unbiased, and ethically sourced?</i>
	1.2 Design Process	<ul style="list-style-type: none"> Privacy and data security Lack of stakeholder consultation 	<i>How are stakeholders engaged in the design process, and what mechanisms ensure their voices are heard?</i>
	1.3 Unintended Effects	<ul style="list-style-type: none"> Discrimination Inaccuracy Non-transparency 	<i>In what ways might this AI technology lead to unintended effects?</i>
2. Organizational	2.1 Responsibility Goal	<ul style="list-style-type: none"> Business v.s. social logic 	<i>How does the organization balance profit-making with ethical and social responsibilities when developing the AI technology?</i>
	2.2 Risk Management	<ul style="list-style-type: none"> Lack of accountability; Insufficient ethical concerns 	<i>When the AI causes harm, who will be held accountable, and what processes are in place to ensure its compliance with regulations and ethical norms?</i>
	2.3 Post-deployment	<ul style="list-style-type: none"> Negative social or environmental effects Loss of public trust 	<i>How does the organization plan to rebuild trust when negative issues arise with its AI technology?</i>
3. Socio-political	3.1 Existing Power Structure	<ul style="list-style-type: none"> Socioeconomic inequalities Power asymmetries 	<i>What are existing marginalized groups related to the design and application of this AI system?</i>
	3.2 Reproducing Injustice	<ul style="list-style-type: none"> Injustice Unequal distribution of benefits 	<i>When AI is used, does it benefit all communities equally, or does it exclude any groups from its benefits?</i>
	3.3 Structural Change	<ul style="list-style-type: none"> Relational responsibility 	<i>How can this AI system actively contribute to a fairer socio-political landscape?</i>
4. Ontological	4.1 Mindset	<ul style="list-style-type: none"> Dataism AI solutionism Anthropocentrism 	<i>What are the underlying assumptions driving the design of this AI technology?</i>
	4.2 Lifestyle Influence	<ul style="list-style-type: none"> Disruption of traditions Detachment Datafication Loss of meaning 	<i>How might this AI technology impact humans' ways of life, and when does this influence might become problematic?</i>
	4.3 Reimagining Solutions	<ul style="list-style-type: none"> Design for marginal groups Community-based design Social innovations 	<i>What are alternative (non-AI) solutions that may be more appropriate to address the problem?</i>

AI: artificial intelligence.

metaphor of structure” (Young, 2001: 18) and treating structures as independent entities from the artifacts. In fact, there are interconnections across levels. For example, a biased algorithm (artifact level) may lead to discriminatory hiring practices within a company and affect public trust in that organization (organizational level). This, in turn, may reinforce existing unequal power structures and perpetuate systemic biases (socio-political level), which may further influence broader beliefs and narratives about justice and meritocracy (ontological level). Considering these interconnections, when we attribute an issue to a particular level, it does not mean that the issue exists *solely* within that level. Rather, each level has its

primary focus, ranging from the immediate impact of AI technologies (artifact level), to the dynamics of organizations (organizational level), to systemic inequalities (socio-political level), and finally to broader worldviews and ideologies (ontological level). This layered approach allows for a more nuanced understanding of how issues manifest and interact across different domains.

Implementing multi-level framework as a diagnosis of socio-ethical issues

In this section, we perform a first try to implement the multi-level framework. While a detailed explanation is beyond the

scope of this article, we propose here a multi-level impact matrix to briefly illustrate how this approach can help identify those issues. We categorize the social-ethical impacts into four levels: artifact, organizational, socio-political, and ontological. At each level, we analyze three types of impacts based on the input, throughput, and output of AI influence. For each type of impact, we use a prompt question to diagnose potential socio-ethical issues in AI. As defined in the Introduction, the input/throughput/output framework provides a structured analysis. The input refers to the goals or resources that need to be considered before the design process. The throughput involves integrating these goals or resources into actual design or broader social processes. The output focuses on the intended or unintended effects of those processes and potential ways to address them.

We thus apply the input/throughput/output framework to make a structured analysis of AI's impacts across different levels. **At the artifact level**, the input is about what the goals the AI is going to achieve and what challenges it is meant to address. The throughput involves how stakeholders are engaged in the design process. The output focuses on the unintended effects of AI. **For the organization level**, the input is about how economic and social goals are defined and how they may come into tension. The throughput addresses how organizations manage the ethical, legal, and social risks of AI and how to make it compliant with regulations and ethical norms. The output is about building and maintaining trust when issues arise. **At the socio-political level**, the input concerns existing power structures, injustices, and marginalized groups. The throughput explores how AI may mirror or reproduce these inequalities. The output focuses on how to responsibly contribute to a more just social-political landscape. **At the ontological level**, the input concerns the existing hidden assumptions in AI developers' thinking when they start to design the technology. The throughput examines how these assumptions are baked into AI systems and how they influence human ways of life. The output challenges us to reimagine alternative solutions or beliefs.

All the issues associated with AI's impacts at different levels, along with the diagnostic questions to identify them, are summarized in Table 2. This matrix does not provide a full list of issues that AI developers can directly address. Instead, we intentionally design it not as a checklist for AI developers, but to be more suitable for workshop settings or focus groups, where various stakeholders can engage in meaningful discussions. A workshop setting is more suitable here because, as we have argued throughout the paper, many socio-ethical issues, especially some structural ones, cannot be fixed by design alone and need diverse stakeholders to negotiate solutions. So, the questions are carefully tailored to make sure that they can effectively identify relevant issues across impacts at different levels, but at the same time being open-ended and generic enough to capture nuances, diverse perspectives, and foster debate on the issues.

This multi-level impact matrix can be used during the early-stage design phase when AI technologies are still adaptable, so AI developers can work with stakeholders to identify potential issues and adjust AI in a responsible way. For example, consider the well-known case of Amazon's automated recruitment system, which discriminated against women applicants (Kodiyar, 2019). In the early stages of development, diverse stakeholders—such as AI developers from Amazon, policymakers, job applicants, and ethical and legal experts—could use this matrix and its set of questions to diagnose socio-ethical issues. This matrix would help reveal that bias is not just a problem of biased datasets (BBC, 2018), but how it is embedded in the input, throughput, and output of the AI design process, and how it is linked to broader organizational cultures, historical patriarchal inequalities, and cultural beliefs. Also, it could prompt reflection on AI solutionism, encouraging Amazon and stakeholders to consider whether such automated hiring systems should even be developed in the first place, or explore alternative design philosophies that prioritize applicants over profits and efficiency.

Conclusion

This paper critically assesses the landscape of AI ethics, with a particular focus on its practical tendency to integrate ethical values into AI design practices. While this by-design AI ethics can be beneficial in addressing explicit ethical and social issues related to specific AI artifacts, it may overlook relevant structural issues and fail to address hidden, long-term harms. We start with a detailed review of the input (Section Challenging the input of by-design AI ethics: translating uncritical values into actions), throughput (Section Challenging the throughput of by-design AI ethics: participation without structural change), and output (Section Challenging the output of by-design AI ethics: unpredictable outcomes) phases of by-design AI ethics, revealing its strong emphasis on individual artifacts while neglecting broader structural concerns. To address this gap, we propose a multi-level framework to identify socio-ethical issues in AI at both the artifact level and the broader structural level (Sections The development of a multi-level framework for AI ethics, and Implementing multi-level framework as a diagnosis of socio-ethical issues). This framework can help ethical experts, AI designers, and policymakers develop responsible AI technologies by critically examining hidden assumptions and blind spots within AI guidelines, policies, and regulatory frameworks. Future research will explore how to integrate this multi-level framework into real-world AI design and policymaking to drive structural and transformative change.


Acknowledgments

We would like to thank Victor Betriu, Georgios Tsagdis, Luuk Stellinga, Zoe Robaey, Norbert Peeters, Alessio Gerola, and

Thijs Loonstra for their feedback on an early version of this paper. We are also grateful for the support of ELSA Lab team and Philosophy Group at Wageningen University.

ORCID iD

Hao Wang  <https://orcid.org/0000-0002-8107-6095>

Vincent Blok  <https://orcid.org/0000-0002-9086-4544>

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This work was supported by Nederlandse Organisatie voor Wetenschappelijk Onderzoek [grant number Nwa.1332.20.002].

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Note

1. This definition should hold for both ontology and social-political structures. So, the “large-scale processes” means a social process as well as an ontological process.

References

- Amstein SR (2019) A ladder of citizen participation. *Journal of the American Planning Association* 85(1): 24–34.
- Attard-Frost B, De los Ríos A and Walters DR (2023) The ethics of AI business practices: A review of 47 AI ethics guidelines. *AI and Ethics* 3(2): 389–406.
- BBC (2018, October 10) *Amazon scrapped ‘sexist AI’ tool*. BBC News. <https://www.bbc.com/news/technology-45809919>
- Benjamin R (2019) Assessing risk, automating racism. *Science* 366(6464): 421–422.
- Birhane A (2020) Algorithmic colonization of Africa. *SCRIPT-ed* 17(2): 389–409.
- Birhane A, Ruane E, Laurent T, et al. (2022) The forgotten margins of AI ethics. In: 2022 ACM Conference on Fairness, Accountability, and Transparency, pp.948–958.
- Bleher H and Braun M (2023) Reflections on putting AI ethics into practice: How three AI ethics approaches conceptualize theory and practice. *Science and Engineering Ethics* 29(3): 21.
- Blok V (2023a) The ontology of technology beyond anthropocentrism and determinism: The role of technologies in the constitution of the (post)Anthropocene world. *Foundations of Science* 28(3): 987–1005.
- Blok V (2023b) *Philosophy of Technology in the Digital Age: The Datafication of the World, the Homo Virtualis, and the Capacity of Technological Innovations to Set the World free*. Wageningen: Wageningen University & Research.
- Blok V and Lemmens P (2015) The emerging concept of responsible innovation. Three reasons why it is questionable and calls for a radical transformation of the concept of innovation. In: Koops B, Oosterlaken R, Swierstra T and van den Hoven J (eds) *Responsible Innovation 2*. Cham: Springer International Publishing, pp.19–35.
- Boenink M and Kudina O (2020) Values in responsible research and innovation: From entities to practices. *Journal of Responsible Innovation* 7(3): 450–470.
- Brand T and Blok V (2019) Responsible innovation in business: A critical reflection on deliberative engagement as a central governance mechanism. *Journal of Responsible Innovation* 6(1): 4–24.
- Brevini B (2023) Myths, techno solutionism and artificial intelligence: reclaiming AI materiality and its massive environmental costs. In: *Handbook of Critical Studies of Artificial Intelligence*. Regency: Edward Elgar Publishing, pp.869–877.
- Brey P and Dainow B (2023) Ethics by design for artificial intelligence. *AI and Ethics* 4(4): 1265–12777.
- Buijsman S, Klenk M and Van den Hoven J (2025) Ethics of AI: Toward a “Design for Values” Approach. In: Smuha N (ed) *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence edit*. Cambridge: Cambridge University Press, pp.59–78.
- Corrêa NK, Galvão C, Santos JW, et al. (2023) Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns* 4(10): 1–14.
- Couldry N and Mejias UA (2018) Data colonialism: Rethinking big data’s relation to the contemporary subject. *Television & New Media* 20(4): 336–349.
- Crawford K (2021) *The Atlas of AI Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.
- Davidson L and Satta M (2021) Justified social distrust. In: Weber K (eds) *Social Trust*. New York: Routledge, 123.
- Dignum V (2018) Ethics in artificial intelligence: Introduction to the special issue. *Ethics and Information Technology* 20(1): 1–3.
- Donia J and Shaw JA (2021) Ethics and values in design: A structured review and theoretical critique. *Science and Engineering Ethics* 27(5): 57.
- Floridi L (2019) Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy & Technology* 32(2): 185–193.
- Fritts M and Cabrera F (2021) AI recruitment algorithms and the dehumanization problem. *Ethics and Information Technology* 23(4): 791–801.
- Frye M (1983) *Oppression. The Politics of Reality*. Trumansburg, N.Y.: The Crossing Press.
- Giddens A (1979) *Central Problems in Social Theory: Action, Structure and Contradiction in Social Analysis*. London: MacMillan.
- Gill KS (2020) Prediction paradigm: The human price of instrumentalism. *Ai & Society* 35(3): 509–517.
- Gilman ME (2022) Beyond window dressing: Public participation for marginalized communities in the datafied society. *Fordham Law Review* 91(2): 556.
- Gray M and Suri S (2019) *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. New York: Houghton Mifflin Harcourt.

- Greene D, Hoffmann AL and Stark L (2019) Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning.
- Haga T (2022) Alienation in a digitalized world. *Ai & Society* 37(2): 801–814.
- Hagendorff T (2020) The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines* 30(1): 99–120.
- Hagendorff T (2021) Blind spots in AI ethics. *AI and Ethics* 2(4): 851–867.
- Hagendorff T (2022) A virtue-based framework to support putting AI ethics into practice. *Philosophy & Technology* 35(3): 1–24.
- Hansson SO (2017) *The Ethics of Technology: Methods and Approaches*. London: Rowman & Littlefield International.
- Haslanger S (2012) *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press.
- Haslanger S (2023) Systemic and structural injustice: Is there a difference? *Philosophy* 98(1): 1–27.
- Hleg A (2019) European Commission's Ethics Guidelines for Trustworthy Artificial Intelligence.
- Hu M (2020) Cambridge Analytica's black box. *Big Data & Society* 7(2): 2053951720938091.
- Jobin A, Ienca M and Vayena E (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1(9): 389–399.
- Kodiyan AA (2019) An overview of ethical issues in using AI systems in hiring with a case study of Amazon's AI based hiring tool. *Researchgate Preprint* 12(1): 1–9.
- Korenhof P, Blok V and Kloppenburg S (2021) Steering representations—Towards a critical understanding of digital twins. *Philosophy & Technology* 34(4): 1751–1773.
- Luccioni AS, Viguier S and Ligozat AL (2023) Estimating the carbon footprint of BLOOM, a 176B parameter language model. *Journal of Machine Learning Research* 24(1): 15.
- Madaio M, Blodgett SL, Mayfield E, et al. (2022) Beyond Fairness Structural (IN)justice lenses on AI for Education. In: Holmes W and Porayska Pomsta K (eds) *The Ethics of Artificial Intelligence in Education*. New York: Routledge, pp.203–239.
- Metcalf J, Singh R, Moss E, et al. (2023) Taking algorithms to courts: A relational approach to algorithmic accountability. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, pp.1450–1462.
- Mills CW (1959) *The Sociological Imagination*. Oxford: Oxford University Press.
- Mittelstadt B (2019) Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* 1(11): 501–507.
- Muldoon J and Wu BA (2023) Artificial intelligence in the colonial matrix of power. *Philosophy & Technology* 36(4): 80.
- Munn L (2022) The uselessness of AI ethics. *AI and Ethics* 3(3): 869–877.
- Nakada M and Tamura T (2005) Japanese Conceptions of privacy: An intercultural perspective. *Ethics and Information Technology* 7(1): 27–36.
- Naughton J (2023, December 30) *For all the Hype in 2023, We Still Don't Know What AI's Long-Term Impact Will Be*. The Guardian. <https://www.theguardian.com/commentisfree/2023/dec/30/ai-artificial-intelligence-2023-long-term-impact-nvidia-h100-microsoft#:~:text=For%20all%20the%20hype%20in,be%20%7C%20John%20Naughton%20%7C%20The%20Guardian>
- Newlands G (2021) Lifting the curtain: Strategic visibility of human labour in AI-as-a-Service. *Big Data & Society* 8(1): 1–14.
- NLAIC (2025) ELSA Concept. Available at: <https://nlaic.com/en/bouwsteen/human-centric-ai/elsa-concept/> (accessed 13/04/2025).
- Perrigo B (2023) OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic. *Time*.
- Rafanelli LM (2022) Justice, injustice, and artificial intelligence: Lessons from political theory and philosophy. *Big Data & Society* 9(1): 20539517221080676.
- Rességuier A and Rodrigues R (2020) AI Ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society* 7(2): 2053951720942541.
- Robertson T and Simonsen J (2012) Challenges and opportunities in contemporary participatory design. *Design Issues* 28(3): 3–9.
- Roessler B (2015) Should personal data be a tradable good? On the moral limits of markets in privacy. In: Roessler B and Mokrosinska D (eds) *Social Dimensions of Privacy: Interdisciplinary Perspectives*. Cambridge: Cambridge University Press, pp.141–161.
- Rosenberger R (2014) Multistability and the agency of mundane artifacts: From speed bumps to subway benches. *Human Studies* 37(3): 369–392.
- Ryan M and Blok V (2023) Stop re-inventing the wheel: Or how ELSA and RRI can align. *Journal of Responsible Innovation* 10(1): 2196151.
- Ryan M, De Roo N, Wang H, et al. (2024) AI through the looking glass: an empirical study of structural social and ethical challenges in AI. *AI & Soc*, 1–17. DOI: 10.1007/s00146-024-02146-0.
- Sadek M, Calvo RA and Mougenot C (2023) Designing value-sensitive AI: A critical review and recommendations for socio-technical design processes. *AI and Ethics* 4(4): 949–967.
- Sambasivan N, Arnesen E, Hutchinson B, et al. (2021) Re-imagining algorithmic fairness in India and beyond. In: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, pp.315–328.
- Sewell W (2005) *Logics of History: Social Theory and Social Transformation*. Chicago: University of Chicago Press.
- Simondon G (2017) *On the Mode of Existence of Technical Objects*. Minneapolis: Univocal Publishing.
- Sloane M (2024) Controversies, contradiction, and “participation” in AI. *Big Data & Society* 11(1): 20539517241235862.
- Smuha NA (2021) Beyond the individual: governing AI's societal harm. *Internet Policy Review* 10(3): 1–31.
- Stahl BC (2021) *Artificial Intelligence for a Better Future*. Cham: Springer Nature. pp.91–115.
- Ten Holter C (2022) Participatory design: Lessons and directions for responsible research and innovation. *Journal of Responsible Innovation* 9(2): 275–290.
- Umbrello S and van de Poel I (2021) Mapping value sensitive design onto AI for social good principles. *AI Ethics* 1(3): 283–296.

- Van den Hoven J (2013) Value sensitive design and responsible innovation. In: Owen, R, Bessant, J, and Heintz, M (eds) *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*. The Atrium: John Wiley & Sons Ltd, pp.75–83.
- van der Burg S, Giesbers E, Bogaardt MJ, et al. (2022) Ethical aspects of AI robots for agri-food; a relational approach based on four case studies. *Ai & Society* 39(2): 541–555.
- Van Maanen G (2022) AI ethics, ethics washing, and the need to politicize data ethics. *Digital Society* 1(2): 9.
- Veale M (2020) A critical take on the policy recommendations of the EU high-level expert group on artificial intelligence. *European Journal of Risk Regulation* 11(1): e1.
- Wagner B (2018) Ethics as an Escape from regulation. In: Bayamlioğlu E, Baraliuc I, Janssens L, et al. (eds) *Being Profiled: Cogitas Ergo Sum. 10 Years of 'Profiling the European Citizen'*. Amsterdam: Amsterdam University Press, pp.84–89.
- Wakunuma K, de Castro F, Jiya T, et al. (2021) Reconceptualising responsible research and innovation from a global south perspective. *Journal of Responsible Innovation* 8(2): 267–291.
- Wang H (2022a) *Algorithmic Colonization: Automating Love and Trust in the Big Data Society*. Amsterdam: University of Amsterdam Library.
- Wang H (2022b) Transparency as manipulation? Uncovering the disciplinary power of algorithmic transparency. *Philosophy & Technology* 35(3): 69.
- Wang H (2023) Algorithmic Colonization of Love: Ethical Challenges of Dating App Algorithms. *Techné*.
- Wyatt S (2021) Metaphors in critical internet and digital media studies. *New Media & Society* 23(2): 406–416.
- Wynsberghe AV, Vandemeulebroucke T, Bolte L, et al. (2023) *Towards the Sustainability of AI: Multi-Disciplinary Approaches to Investigate the Hidden Costs of AI*. Basel: MDPI.
- York R and McGee JA (2016) Understanding the Jevons paradox. *Environmental Sociology* 2(1): 77–87.
- Young I (2001) Equality of whom? Social groups and judgments of injustice. *Journal of Political Philosophy* 9(1): 18.
- Young I (2004) Responsibility and global labor justice. *Journal of Political Philosophy* 12(1): 365–388.
- Young I (2006) Responsibility and global justice: A social connection model. *Social Philosophy and Policy* 23(1): 102–130.
- Young I (2011) *Responsibility for Justice*. Oxford: Oxford University Press.
- Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.
- Zwart H, Barbosa Mendes A and Blok V (2024) Epistemic inclusion: A key challenge for global RRI. *Journal of Responsible Innovation* 11(1): 2326721.
- Zwart H and Nelis A (2009) What is ELSA genomics? *Embo Reports*.