# Ethics dumping in artificial intelligence

Jean-Christophe Bélisle-Pipon[1]* and Gavin Victor[2]

[1]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada, [2]Philosophy Department, Simon Fraser University, Burnaby, BC, Canada

Artificial Intelligence (AI) systems encode not just statistical models and complex algorithms designed to process and analyze data, but also significant normative baggage. This ethical dimension, derived from the underlying code and training data, shapes the recommendations given, behaviors exhibited, and perceptions had by AI. These factors influence how AI is regulated, used, misused, and impacts end-users. The multifaceted nature of AI's influence has sparked extensive discussions across disciplines like Science and Technology Studies (STS), Ethical, Legal and Social Implications (ELSI) studies, public policy analysis, and responsible innovation—underscoring the need to examine AI's ethical ramifications. While the initial wave of AI ethics focused on articulating principles and guidelines, recent scholarship increasingly emphasizes the practical implementation of ethical principles, regulatory oversight, and mitigating unforeseen negative consequences. Drawing from the concept of "ethics dumping" in research ethics, this paper argues that practices surrounding AI development and deployment can, unduly and in a very concerning way, offload ethical responsibilities from developers and regulators to ill-equipped users and host environments. Four key trends illustrating such ethics dumping are identified: (1) AI developers embedding ethics through coded value assumptions, (2) AI ethics guidelines promoting broad or unactionable principles disconnected from local contexts, (3) institutions implementing AI systems without evaluating ethical implications, and (4) decision-makers enacting ethical governance frameworks disconnected from practice. Mitigating AI ethics dumping requires empowering users, fostering stakeholder engagement in norm-setting, harmonizing ethical guidelines while allowing flexibility for local variation, and establishing clear accountability mechanisms across the AI ecosystem.

KEYWORDS

artificial intelligence, AI ethics, ethics dumping, ethical guidelines, accountability, AI governance

## Introduction

It is widely accepted that technologies are not value-neutral. This is especially true for complex and impactful technologies like artificial intelligence (AI) (Elliott, 2019; Alami et al., 2020). Beyond encoding statistical models and procedural rules, AI systems embody significant ethical baggage derived from their underlying code and the data used (Norori et al., 2021). This normative dimension shapes AI systems' recommendations, behaviors, and decision-making processes. It also influences how AI is perceived, regulated, used (or misused), and how it impacts different stakeholders, especially end-users (Berberich et al., 2020). AI's undeniable power has sparked discussions in domains such as ELSI (ethical, legal, and social implications), STS (Science and Technology Studies), and applied ethics (Dubber et al., 2020; Hagendorff, 2020; Horgan et al., 2020; Bélisle-Pipon et al., 2021; Slota et al., 2021). A recurring theme is that technologies can impose values and norms exogenous to the contexts

where they are deployed (e.g., financial technologies marginalizing those without banking access, precision agriculture conflicting with traditional practices and biodiversity, advanced medical equipment ill-suited for local healthcare infrastructures, and surveillance systems raising privacy concerns in diverse social contexts and norms). Such value imposition necessitates critical analysis of an AI's ethical underpinnings. In the wake of rising concerns over AI's societal impacts, a flourishing ecosystem of scholarship and initiatives has emerged to develop ethical frameworks and guidelines for responsible AI development and use (Hagendorff, 2020). However, this initial wave prioritized articulating high-level principles over examining practical challenges in norm design and implementation (Lauer, 2020). The field advanced cautiously, outlining a conceptual map of AI ethics imperatives which had previously been uncharted.

As the discourse matures, increasing attention is devoted to the challenges of translating ethical principles into practice and developing robust governance mechanisms (Mittelstadt, 2019; Ryan and Stahl, 2020). Key concerns include responsibly embedding ethics into design and deployment pipelines (Floridi et al., 2020), evaluating normative impacts across contexts (Eitel-Porter, 2020), and establishing accountability structures (Russell et al., 2015). Something not often considered, though, is that ethical guidelines and accountability expectations can be unfairly placed on users and impacted communities, leading to an unfair normative burden if developers and regulators offload responsibility without empowering local agency. To further investigate this dynamic, we look to the concept of "ethics dumping" in research ethics, which highlights how the export of ethical guidelines and governance from powerful actors to less-privileged contexts can impose disproportionate costs and burdens if local conditions are not accounted for (Schroeder et al. 2019a). Drawing on this framing, the present paper argues that practices surrounding the development, deployment and governance of AI can lead to different forms of ethics dumping, thereby unduly burdening, and even harming, users and host environments who may lack the capacities or agency to responsibly manage the embedded normative content and ethical implications.

The paper begins by outlining the ethics dumping concept and its significance within research ethics. It then examines four key trends illustrating how different AI actors—developers, principles/guideline authors, institutional implementors, and regulators/policymakers—can each contribute to ethics dumping through their practices. Finally, potential mitigation strategies are discussed, emphasizing approaches that empower users/host environments, foster stakeholder inclusion, harmonize ethical guidelines while allowing flexibility for local variation, and establish clear accountability structures across the AI ecosystem.

## Ethics dumping: a borrowed concept

In its simplest conception, ethics dumping refers to the export of ethically dubious practices from a privileged context to one with weaker governance mechanisms (Schroeder et al., 2016). Originating in research ethics, the concept highlights how problematic behaviors occur more frequently when studies are conducted in resource-poor settings "with weaker compliance structures or legal governance" (Schroeder et al., 2019b). Classic examples include drug trials that exploit regulatory gaps or institutional deficiencies in developing countries, undermining robust ethical protocols, and imposing disproportionate risks on local populations who may not share in the eventual benefits. Beyond

physical and economic asymmetries, the ethics dumping concept also captures how ethical principles and accountability norms flow from powerful actors to recipient communities without their participation or ability to shape those norms. At its core, ethics dumping emerges from skewed power dynamics, where certain downstream populations or environments are unable to negotiate or uphold ethical practices aligned with their values and contexts (Schroeder et al., 2019b; Samuel and Derrick, 2020).

The European Union (EU)-funded TRUST project developed guidance for addressing and countering ethics dumping in international research partnerships (Andanda et al., 2014). They trace the concept of ethics dumping to a more general phenomenon, "dumping." In the global economic and trade sectors, dumping is a form of predatory pricing. Dumping refers to cases where "large entities can afford to undercut local competitors for a given period, to drive them out of the market" (Andanda et al., 2014). By doing so, the companies "dump" their product into a new environment in a manner that harms said environment by predating on local competition. Viewed as a form of dumping, though ethics dumping tends to be described in terms of international research practices, it may also be understood as a more general predatory phenomenon: placing ethical responsibility in a domain which lacks the ability to incorporate that responsibility into its operations.

In the context of biomedical research, the concept of ethics dumping is further elaborated by Liao et al. (2023) in their analysis of its occurrence within China. They define ethics dumping as the practice where researchers, often from countries with stringent ethical regulations, conduct ethically questionable research in regions with less rigorous oversight, exploiting local vulnerabilities. The authors highlight how China, with its relatively weaker ethical governance structures, becomes a target for such practices. They detail cases like the CRISPR baby scandal and the Golden Rice incident to illustrate how ethics dumping manifests in biomedical research, where the pursuit of scientific advancement or personal gain overrides ethical considerations, often at the expense of local populations. The authors emphasize that ethics dumping is not merely a transfer of unethical practices but often a deliberate circumvention of ethical norms, facilitated by gaps in local oversight and regulation, leading to significant ethical breaches and harm. However, they acknowledge that it may sometimes occur unintentionally, as individuals operating in unfamiliar contexts might unknowingly engage in inappropriate practices due to insufficient understanding or awareness of relevant ethical considerations (Liao et al., 2023).

The ethics dumping concept provides a useful lens for examining dynamics within AI ethics. Although the AI context differs from research ethics, similar power asymmetries and risks of offloading normative content exist between developers of increasingly autonomous and opaque systems, and end-users expected to responsibly deploy those systems. The following section outlines four key trends illustrating how practices from different actors across the AI ecosystem can contribute to ethics dumping.

## AI ethics dumping: four concerning trends

Like other technologies where ethical implications cut across domains, defining responsible practices for AI has catalyzed a "race"

by numerous organizations aiming to shape the normative discourse and establish ethical guardrails (Whittlestone et al., 2019; Keller and Drake, 2020). However, despite this flurry of activity articulating principles and guidelines, concerns remain over how well AI ethics translates into practice (Ho, 2020). Local norms, constraints, and stakeholder values risk being overridden by one-size-fits-all frameworks (Chaudhary, 2020). More insidiously, skewed power dynamics mean that those developing ethical guidelines or deploying AI systems may effectively dictate normative standards that impacted communities must then bear primary responsibility for upholding, even if poorly equipped or excluded from shaping those standards. This section identifies four key trends (see Table 1) that can enable ethics dumping in the AI context, highlighting how different actors—developers, ethics principle-authors, institutional implementors, and regulators—may each facilitate the unfair transfer of ethical burdens and responsibilities.

## Trend 1: ethics dumping by AI developers

Like all technologies, AI systems have profound impacts on users and their environments that go beyond just technical functionality (Alami et al., 2020; ÓhÉigeartaigh et al., 2020). What fuels AI's transformative promise and "hype" is precisely its potential to significantly reshape practices, beliefs, and decision-making processes (Matheny et al., 2019; Char et al., 2020). However, this transformative capacity means users are not just adopting a tool, but also the normative priors embedded within it. As highly opaque and autonomous systems, developers often have disproportionate influence in shaping AI's ethical dimension by encoding inferences, assumptions, and values into algorithms and training datasets (Rahwan, 2018; ; McDermid et al., 2021). AI does not neutrally reflect the world but actively frames and constructs it through the choices and constraints inherent in its design (Barocas and Selbst, 2016; Mittelstadt et al., 2016). Datasets skewed by historical biases and human developer assumptions about relevant features mean AI outputs can reflect and perpetuate specific cultural viewpoints, hierarchies, and injustices (Buolamwini and Gebru, 2018).

While typically undertaken with no malicious intent, allowing values to creep into algorithms in this manner facilitates ethics dumping in several ways. First, by implicitly encoding norms within AI systems' architecture (designed in a specific cultural/institutional milieu), users adopt this "black boxed" normative baggage upon deployment even if antithetical to local values or accountability customs (Ananny and Crawford, 2018; Crawford, 2021; Bélisle-Pipon et al., 2021). That is, users might be interacting with an algorithm that embodies values which do not align with their own, but AI opacity prevents them from recognizing whether this is even the case. Second, despite the vast influence of developer ethics-by-design choices (such as enhancing privacy protection and data security, mitigating bias, and seeking to improve fairness in algorithms), the opacity and inscrutability of most AI systems leave users with little insight into the rationale or assumptions driving system behaviors, and leave them with little means of addressing the normative considerations baked into the algorithms (Burrell, 2016; Akhai, 2023; Stafie et al., 2023). Finally, the very distance between developers and where impacts manifest means developers are not the ones who will have to manage the impacts of their technology (Crawford and Calo, 2016). This misalignment between influence over normative design choices and accountability for downstream effects enables ethics dumping—it allows AI use contrary to exemplary or even local standards if communities lack resources to investigate and intervene.

As an example, consider OpenAI's recent controversy surrounding watermarking ChatGPT text outputs. Though the company has been developing a watermarking feature for about a year, it has decided against implementing the feature for fears that it will result in less usage (Davis, 2024). By not instituting a watermark, OpenAI places responsibility on downstream stakeholders to find mechanisms of accountability for users. Importantly, OpenAI bases its decision not on increasing calls for responsible usage of large language models, but with surveyed users, almost 30% of whom would be less likely to use ChatGPT with a watermark (Davis, 2024). The company therefore has the means to act, but faced with a strategic loss, has decided not to make the use of its product more transparent and traceable. In

TABLE 1 Summary of ethics dumping trends in artificial intelligence.

| Trend | Description | Key Features |
|---|---|---|
| Trend 1: ethics dumping by AI developers | AI developers exert disproportionate influence over ethical norms through design choices and embedded values in algorithms. This can result in users unknowingly adopting biased or misaligned norms due to the opaque nature of AI systems. | Norms encoded in AI design<br>Lack of transparency<br>Users adopt technology without the means to implement ethical safeguards |
| Trend 2: ethics dumping by ethics guideline authors | Broad and vague ethical guidelines, developed without local stakeholder engagement, place the burden on companies/developers or end-users to interpret and apply these principles within their specific contexts, leading to ethics dumping. | Vague principles difficult to translate into actionable items<br>Disconnect from operational realities<br>Users bear responsibility for implementation |
| Trend 3: ethics dumping by institutional implementors | Institutions adopting AI systems often prioritize deployment over ethical evaluation, resulting in a shift of responsibility for ethical and legal compliance to end-users, who may lack the resources to manage these responsibilities. | Context of rapid AI deployment<br>Inadequate ethical evaluation<br>Responsibility shifted to end-users |
| Trend 4: ethics dumping via governance frameworks | Governance for large-scale AI solutions often lack inclusive processes, imposing norms without considering local contexts, and shifting compliance burdens onto users. | Top-down governance<br>Lack of thorough stakeholder inclusion<br>One-size-fits-all ethical norms<br>Shifted compliance responsibilities to users |

doing so, ChatGPT is "dumped" into the market—a market which is not clearly prepared to deal with the ramifications of the product—while also entrusting these users with responsibilities for accountability attribution, even if it is far more difficult for those outside OpenAI to institute accountability structures.

Some suggest that ethics-by-design can resolve these tensions by hardwiring ethical guidelines into algorithms from inception (Dignum, 2018). However, self-regulation is an insufficient solution if the processes around defining and embedding ethical norms lack meaningful stakeholder participation and oversight. Without inclusive and thorough norm-setting structures, ethics-by-design risks amplifying ethics dumping by concentrating normative influence in the hands of developers and institutions least accountable to impacted populations (Hasselbalch, 2021; Fraenkel, 2024; Umbrello, 2024).

## Trend 2: ethics dumping by AI ethics guideline authors

The recent profusion of AI ethics guidelines and principles represents a shared aspiration to normatively constrain the development and application of transformative AI systems. However, their very generality and lack of context-specific stakeholder engagement, counterintuitively, can *enable* ethics dumping. Most ethical AI principles articulate high-level injunctions like transparency, fairness, accountability, and respect for human rights that few would disagree with (Jobin et al., 2019). Well-intentioned as ethical principles are, their vagueness and distance from operational realities risk creating an "ethics buffer" (or ethics washing) that signals virtuous intent while doing little to guide on-the-ground implementation by those bearing responsibility for upholding such principles (Vakkuri et al., 2019). This "ethics buffer" dynamic engenders ethics dumping because broad principles developed without local stakeholder buy-in place disproportionate burdens on users (Criado-Perez, 2019).

Consider, for example, the White House "Blueprint for an AI Bill of Rights" and its call for notices informing those impacted by AI systems (The White House, 2022). The Blueprint states that

> Designers, developers, and deployers of automated systems should provide generally accessible plain language documentation including clear descriptions of the overall system functioning and the role automation plays, notice that such systems are in use, the individual or organization responsible for the system, and explanations of outcomes that are clear, timely, and accessible (The White House, 2022).

The Blueprint does well to specify what is meant by a notice, so that the notice is clear, informative, and relevant. Important as disclosures and notices may be, this call transfers ethical responsibility to "designers, developers, and deployers" (The White House, 2022). While this may seem normal when it comes to such high-level guidance, the problem is this: the guidelines do not seem to specify *when developers* are under the obligation to provide a notice versus when the *deployers* are under the obligation to provide a notice. This leaves a vacuum of guidance, and the expectation that those involved in the creation and rollout of the algorithm will fill this vacuum. But, designers, developers, and deployers do not have ethical standards by which they can arbitrate between each other's

role in the process of forming a notice. In the absence of more clear responsibilities held by parties involved, guidance like the Blueprint takes a step in the right direction, but also passes ethical responsibilities to parties lacking the structures by which to embody and carry out these responsibilities.

Responsibility and accountability flow towards practitioners although they lacked agency and authority in defining the principles expected of them. Care is a central aspect of responsibility, especially toward people in vulnerable situations, such as those in resource-poor areas. When those tasked with supporting these individuals fail to provide adequate care, there is a risk of exploitation or "dumping." This can occur when limited capacity and expertise hinder effective governance, leading to insufficient protections for those in need (Schroeder et al., 2019b). Some AI ethics guidelines do acknowledge the importance of stakeholder engagement and responsiveness to local norms (Fjeld et al., 2020). However, most guideline development processes have been criticized for marginalizing or excluding key stakeholders like minority groups and civil society (Bélisle-Pipon et al., 2022). Without inclusive representation at a principle's conception, enforcing guidelines across diverse cultural contexts enables ethics dumping by disconnecting normative ideals from local realities. Importantly, this dumping dynamic extends beyond guidelines produced by industry or multi-stakeholder bodies. Well-intentioned guidance from respected institutions, if developed through insular processes insensitive to on-the-ground complexities, paradoxically facilitates dumping by advocating one-size-fits-all accountability checklists that practitioners must then struggle to reinterpret and implement (Rakova et al., 2021; Khosravi et al., 2022; Bennett et al., 2023). Ethics dumping here entails burdening frontline actors with operationalizing academic philosophies abstracted away from pragmatic realities and without accounting for contextual variation.

## Trend 3: ethics dumping by institutional AI implementors

A third manifestation of AI ethics dumping emerges from the practices of institutions acquiring and deploying AI systems. Well-intentioned efforts to leverage transformative AI capabilities can prompt adoption of systems whose embedded normative logics and accountability implications were not robustly evaluated from an ethical lens (Taddeo and Floridi, 2018).

For example, in healthcare settings, initiatives to deploy AI predictive and diagnostic tools, clinical decision support systems (CDSS), AI-generated in-basket responses, or patient monitoring tools tout promises of increased efficiency and optimized resource allocation (Chen et al., 2021; Elhaddad and Hamam, 2024; Garcia et al., 2024; Khosravi et al., 2024). However, ethical issues like eroding human decision autonomy, entrenching biases from historical data, and disrupting professional role boundaries often remain underexplored during acquisition and piloting phases (Char et al., 2018). Pressing user needs and institutional enthusiasm catalyze a "push" dynamic where normative downsides are discounted in the rush towards AI deployment (Lindgren, 2023; Gray and Shellshear 2024). Medical institutions' policies can negatively impact caregivers, administrative staffs, patients, relatives, and stakeholders by shifting the responsibility to them to supervise the use of institution-approved

or promoted tools. This often entails failing to provide necessary frameworks, usage agreements, best practices, or terms of reference to guide the implementation, evaluation, and eventual decommissioning of medical AI tools. If the institution does not take into account the complete lifecycle of a medical AI, and does not act as appropriate gatekeepers, this will have dumping impacts on downstream uses within the institution, and carry not only health, but also ethical and legal risks.

Even when ethical risks and implications are assessed, implementors' incentives frequently align more with signaling ethical conduct through procedural checklists rather than committing resources to substantive mitigations (Raji et al., 2020). As with ethics guidelines, the push dynamic stems from implementor institutions lacking robust governance structures ensuring inclusive participation of stakeholders impacted by new AI deployments. When affected communities are not involved in co-defining metrics, risk thresholds, and ethical guidelines, institutions deploying AI systems may inadvertently engage in practices resembling ethics dumping. Such approach often centralizes control over data systems among institutional actors, while limiting transparency, accountability, and public engagement, thereby leaving communities with little agency over impactful decisions (Dencik et al., 2018). This asymmetry between normative authority and experiential burden reflects a central dynamic underlying ethics dumping.

## Trend 4: ethics dumping via governance frameworks

Many argue that to ensure responsible and trustworthy AI, ethical norms need to be considered from the outset (and not be an afterthought) of the development and deployment of an AI solution, as well as adopting an inclusive, open and transparent approach, especially when it affects a very large number of people or is a public/governmental solution (Dobbe et al., 2021; Couture et al.,2023). This is particularly crucial for the responsible development and deployment of large-scale AI solutions aimed at population-wide or population-specific impacts. Centralized legal/regulatory governance frameworks have advantages of uniformity, reducing ethical hazards from unconstrained corporate self-interest, and institutionalizing public accountability mechanisms. However, many governance frameworks for large-scale AI solutions present a concerning lack of sustained stakeholder inclusion and bottom-up norm-setting, at the risk of neither considering nor responding to the real needs of end-users, but above all, in our case, of making them bear the burden of accountability for the implementation and responsible, trustworthy use of said AI solution. Their development is too frequently steered by regulators, legislators, and entrenched dominant (often coming from industry) incumbents leveraging their advantageous position in the AI sector to guide the development of governance (Bélisle-Pipon et al., 2022). Even well-intentioned civil society voices often find their perspectives marginalized in the closed-door negotiations that shape these rules.

The AuroraAI program in Finland serves as a notable example of how AI ethics dumping can manifest through inadequate governance. AuroraAI was an ambitious initiative aimed at integrating AI into public service delivery to enhance efficiency and provide personalized, data-driven support to citizens (Finnish Center for Artificial Intelligence, 2023). This was set to be the first independent AI assistant

dedicated to public services (ODSC-Open Data Science, 2018). While the program's goal of improving individual well-being through seamless, AI-facilitated interactions is laudable, the ethical oversight of the project has been inadequate. The governance model for AuroraAI has been largely top-down, with critical decisions and ethical considerations being handled by a central Ethics Board established only after significant progress had already been made on the project and which had limited influence over the pre-existing goals and decisions (Leikas et al., 2022).

This delayed and centralized approach to ethical oversight is emblematic of AI ethics dumping, where the burden of managing ethical implications is shifted away from the developers and policymakers who designed the system onto local users and communities who interact with it daily. The limited public engagement in the ethical deliberation process, coupled with the program's broad and generalized ethical guidelines, has resulted in a scenario where the responsibility for addressing complex ethical challenges is disproportionately placed on those least equipped to manage them (Algorithm Watch, 2020). Some raised concerns about "ethics washing" (Leikas et al., 2022). This dynamic not only sidelines broader societal input but also risks embedding ethical standards that may not align with local values or contexts. As a result, the AuroraAI program, despite its human-centric objectives, exemplifies how a governance framework can facilitate ethics dumping by imposing a one-size-fits-all model that overlooks the need for localized ethical consideration and sustained stakeholder engagement.

When governance frameworks for large-scale AI solutions are developed without inclusive processes and are disconnected from the specific realities of end-users, they can exacerbate AI ethics dumping at least three ways. First, by encoding ethical principles favored by developers and policymakers insensitive to minority group needs, governance creates downstream burdens for users to uphold norms shaped without their input (Tallberg et al., 2023). Second, prescriptive top-down design and ruling force a "one-size-fits-all" paradigm on diverse contexts, overriding local values and accountability customs (Cinnamon, 2019). Finally, such governance frameworks frequently concentrate liability risk and regulatory compliance responsibilities on frontline users rather than on developers and institutional procurers of AI systems (Duffourc et al., 2023). Even robust public engagement processes often suffer from tokenism, with participatory mechanisms ill-equipped to translate marginal voices into pragmatic reforms. Consequently, despite progress in AI ethics since revelations of bias, discrimination, and lack of accountability several years ago, there remains an urgent need to proactively mitigate against ethics dumping dynamics that replicate the same power imbalances and normative exclusions the field was meant to redress (Powles and Nissenbaum, 2018).

## Discussion

Research ethics dumping and AI ethics dumping, while both involving the unfair transfer of ethical responsibilities, occur in distinct contexts with differing mechanisms and implications. Research ethics dumping refers to the practice of exporting ethically questionable research practices from a setting with strong governance and ethical oversight to one with weaker regulations, often in developing countries. This often results in local populations bearing

the brunt of risks without corresponding benefits, as researchers exploit regulatory gaps and institutional deficiencies. In this context, ethics dumping is a product of power imbalances where ethical standards are dictated by privileged actors without consideration for the local context, leading to exploitation and harm.

AI ethics dumping, on the other hand, involves the offloading of ethical responsibilities from AI developers, guideline authors, institutional implementors, and policymakers onto end-users and communities that may not have the capacity to manage the embedded normative content of AI systems. This form of dumping arises from the complex and opaque nature of AI technologies, where the ethical implications are often encoded into the algorithms and systems themselves. The lack of transparency and the centralized development of ethical guidelines disconnected from local realities exacerbate the issue, leaving users to grapple with ethical dilemmas without adequate support or resources. Unlike research ethics dumping, which typically involves the physical relocation of research practices, AI ethics dumping is more about the diffusion of ethical burdens across a digital and global landscape, where the consequences of ethical misalignments are often invisible yet profound.

The risks of AI ethics dumping pervade AI development, ethics guidelines enunciation, institutional procurement/deployment, and statutory governance mechanisms. However, scattered across AI ethics discourse are proposals and principles that could serve as guideposts for mitigating dumping dynamics, if embraced more systematically. These value-commitments, processes, and structures emphasize empowering users/host environments, committing to inclusive stakeholder participation, enabling ethical localization while pursuing governance harmonization, and establishing clear accountability attribution across the AI lifecycle.

## Empowering users and host environments

A core ethics dumping dynamic stems from the imposition of AI systems designed elsewhere through opaque processes onto local users without substantive normative negotiation or adaptation. This concentrates normative and technical mastery with developers while users must unilaterally shoulder accountability burdens for unforeseen negative impacts. Mitigating this requires a *pull* rather than *push* approach to AI deployment, where systems are co-developed or iteratively customized to meet verified needs of host environments and align with *in-situ* value systems (Shilton, 2018). The technical and organizational specifics of AI rollout must be workable within existing user constraints, dialogically incorporating feedback on potential tensions and trade-offs (Leslie, 2019). Local users and community representatives must maintain agency and decision rights over whether, how and to what degree AI should be integrated rather than having transformations imposed upon them. Responsible AI development cannot presume to dictate conditions of use but must empower recipients with meaningful refusal rights and agenda-setting influence over ethical baselines (Tinnirello, 2022). This approach shifts the locus of normative control and ensures users are not burdened with compliance obligations disconnected from their lived realities and accountabilities. Rather than ethics dumping, the emphasis should be on co-design and sustained relational alignment between developers and users, balancing expert inputs with contextually-nuanced frontline norms in an inclusive co-reasoning process along the AI lifecycle (Pacia et al., 2024).

## Inclusive stakeholder engagement

The EU-funded TRUST project developed a code of conduct and two other tools that encourage freedom from ethics dumping in international research (Andanda et al., 2014). Such tools encourage active collaboration both internally to the EU and with countries outside the EU (Andanda et al., 2014) as suggested by the Directorate-General for Research and Innovation (2010). A similar approach in the context of AI ethics dumping would be to institutionalize inclusive stakeholder engagement processes across all AI norm-setting activities—ethics guidelines development, institutional procurement protocols, impact assessments and statutory governance frameworks. Currently, key voices are too often sidelined or marginalized in agenda-setting over AI ethics, from minority representatives and civil society advocates to frontline professionals and impacted community members (Bélisle-Pipon et al., 2022). Exclusion facilitates dumping by concentrating normative and technical control with a narrow set of actors furthest removed from downstream burdens of AI deployment. Inclusive processes must solicit, incorporate, and empower diverse stakeholder perspectives throughout each stage of the AI lifecycle—from initial problem formulation to ongoing sociotechnical audits. (Dobbe et al., 2021; Zhang et al., 2021; Riezebos et al., 2022; Lam et al., 2023).

## Enhancing accountability mechanisms in AI ethics

Robust accountability frameworks are critical to ensure that ethical standards are met in AI development and deployment (Wirtz, 2022). This necessitates the establishment of independent oversight bodies with the authority to enforce sanctions and conduct investigations into breaches of ethics (Schmitt, 2022). Transparency is paramount in these processes to enable public oversight and to offer redress for those impacted by AI technologies (Bélisle-Pipon et al., 2022). Beyond procedural measures, instilling a culture of ethics vigilance within the AI development community is essential. Such a culture encourages ongoing self-reflection, adherence to ethical principles, and the courage to see them materialize for stakeholders and end-users throughout the AI lifecycle. Through these mechanisms, it becomes possible to hold AI ethics dumpers accountable, thereby ensuring that AI technologies are developed and deployed in alignment with human rights and democratic values.

## Challenges

While the concept of ethics dumping serves as a valuable lens for examining power imbalances and accountability gaps within the AI ethics ecosystem, it is crucial to recognize the complexity involved in the distribution of ethical responsibility. In some contexts, the transfer of ethical responsibility may not only be unavoidable but also necessary. However, this necessary transfer raises a critical concern: what happens when the entities that *should* assume this responsibility lack the power or capacity to act ethically? In such situations, the transfer could result in a form of ethics dumping on certain stakeholders that, despite their intentions, may be ill-equipped to manage the ethical implications effectively. This highlights a challenge in the discourse around ethics dumping—differentiating between justified transfers of responsibility and those that unfairly burden entities incapable of upholding ethical

standards. Thus, while it is important to advocate for the democratization of normative control (for instance through stakeholder engagement) and to guard against ethics dumping, it is equally important to acknowledge that not all transfers of ethical responsibility are inherently negative. Some are essential for operationalizing ethical principles. The key challenge is ensuring that these transfers are accompanied by adequate support and resources, enabling the responsible parties to fulfill their ethical obligations without perpetuating power imbalances or ethical shortcomings.

## Conclusion

The concept of ethics dumping provides a revealing lens for examining power asymmetries and accountability deficits within the AI ethics ecosystem. As this analysis has shown, practices across the AI lifecycle—from developer choices embedding normative assumptions into systems, to institutional deployments imposing transformative technologies without assessing impacts on users, to governance frameworks dictating compliance obligations while marginalizing key stakeholder voices—all risk facilitating the dumping of ethical burdens onto those least equipped to negotiate or uphold externally-imposed norms.

Enabled by AI's opacity, autonomy, and transformative impacts, ethics dumping represents an insidious dynamic wherein the very actors most influential in shaping AI's normative terrain are able to distance themselves from accountability for its local manifestations and ripple effects. Those "on the ground" interfacing with AI systems daily must then unilaterally shoulder responsibility for addressing ethical quandaries and negative externalities they had little part in defining or governing from the outset. If left unaddressed, ethics dumping threatens to replicate many of the same injustices, exclusions, and power imbalances the AI ethics movement overtly sought to rectify. Virtuous rhetoric around human-centered AI risks ringing hollow if it prioritizes ethics-washing over inclusive participatory mechanisms. Centralized governance aiming for harmonized norms is liable to perpetuate digital coloniality if undertaken without sustained dialogue and norm-negotiation with peripheralized stakeholders. Even the most well-intentioned ethical principles can enable ethics dumping if stakeholder representation was lacking during their conception and formalization.

Mitigating ethics dumping therefore requires multifaceted interventions that democratize normative control over AI systems while clearly delineating responsibility across the entire ecosystem. Empowering users and host environments with agency over AI deployment decisions through *pull* models and continuous feedback loops is crucial, as is fortifying stakeholder inclusion throughout all norm-setting processes. Creating space for ethical localization and on-the-ground improvisation within harmonized governance frameworks is equally vital for avoiding one-size-fits-all imposition of exogenous norms. Ultimately, a commitment to leveling asymmetries and centering those facing the brunt of AI's impacts must be the lodestar for AI ethics going forward. Only by proactively restructuring the distribution of normative control and accountability attribution can the existential risks of ethics dumping be averted. Responsible AI innovation requires democratizing agenda-setting for an ethics of technologies in a pluralistic manner attuned to contextual realities, local priorities, and indigenous perspectives. Broad principles and centralized policies have their place, but not at the cost of perpetuating ethics dumping that undermines AI's emancipatory potential through new modes of disenfranchisement and dispossession.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

Akhai, S. (2023). From black boxes to transparent machines: The quest for explainable AI. *SSRN*. doi: 10.2139/ssrn.4390887

Alami, H., Lehoux, P., Auclair, Y., de Guise, M., Gagnon, M. P., Shaw, J., et al. (2020). Artificial intelligence and health technology assessment: anticipating a new level of complexity. *J. Med. Internet Res.* 22:e17707. doi: 10.2196/17707

Algorithm Watch (2020) Automating Society Report 2020. Available at: https://automatingsociety.algorithmwatch.org

Ananny, M., and Crawford, K. (2018). Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. *New Media Soc.* 20, 973–989. doi: 10.1177/1461444816676645

Andanda, P., Wathuta, J., Leisinger, K., and Schroeder, D. (2014) National and International Compliance Tools, a report for TRUST, Available at: https://trust-project.eu/wp-content/uploads/2017/02/TRUST-664771-National-and-International-Compliance-Tools-Final.pdf

Barocas, S., and Selbst, A. D. (2016). Big Data's Disparate Impact. *Calif. L. Rev* 104, 671–732. doi: 10.2139/ssrn.2477899

Bélisle-Pipon, J.-C., Couture, V., Roy, M.-C., Ganache, I., Goetghebeur, M., and Cohen, I. G. (2021). What makes artificial intelligence exceptional in health technology assessment? *Frontiers in Artificial Intelligence*. doi: 10.3389/frai.2021.736697

Bélisle-Pipon, J.-C., Monteferrante, E., Roy, M.-C., and Couture, V. (2022). Artificial intelligence ethics has a black box problem. *AI Soc.* doi: 10.1007/s00146-021-01380-0

Bélisle-Pipon, J.-C., Powell, M., English, R., Malo, M.-F., Ravitsky, V.Bridge2AI–Voice Consortium, and Bensoussan, Y. (2024). Stakeholder perspectives on ethical and trustworthy voice AI in health care. *Digital Health*, 10. doi: 10.1177/20552076241260407

Bennett, S. J., Claisse, C., Luger, E., and Durrant, A. (2023). Unpicking epistemic injustices in digital health: On designing data-driven technologies to support the self-management of long-term health conditions. *In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society.* doi: 10.7488/era/3891

Berberich, N., Nishida, T., and Suzuki, S. (2020). Harmonizing artificial intelligence for social good. *Philos. Technol.* 33, 613–638. doi: 10.1007/s13347-020-00421-8

Buolamwini, J., and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, Conference on Fairness, Accountability, and Transparency. 81, 1–15. Available at: https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Burrell, J. (2016). How the machine "thinks": understanding opacity in machine learning algorithms. *Big Data Soc.* 3:205395171562251. doi: 10.1177/2053951715622512

Büthe, T., Djeffal, C., Lütge, C., Maasen, S., and Ingersleben-Seip, N.Von. (2022). Governing AI – Attempting to herd cats? Introduction to the special issue on the governance of artificial intelligence. *Journal of European Public Policy*, 29, 1721–1752. doi: 10.1080/13501763.2022.2126515

Buolamwini, J., and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research* 81, 1–15. Conference on Fairness, Accountability, and Transparency. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Cave, S., and Dihal, K. (2020). The whiteness of AI. *Philos. Technol.* 33, 685–703. doi: 10.1007/s13347-020-00415-6

Char, D. S., Abramoff, M. D., and Feudtner, C. (2020). Identifying ethical considerations for machine learning healthcare applications. *Am. J. Bioeth.* 20, 7–17. doi: 10.1080/15265161.2020.1819469

Char, D. S., Shah, N. H., and Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. *N. Engl. J. Med.* 378, 981–983. doi: 10.1056/NEJMp1714229

Charros-García, N. (2019). Mexico–US cooperation against the smuggling of migrants. *European Rev. Latin Am. Caribbean Stud.* 107, 233–257. doi: 10.32992/erlacs.10526

Chaudhary, M. Y. (2020). Initial considerations for Islamic digital ethics. *Philos. Technol.* 33, 639–657. doi: 10.1007/s13347-020-00418-3

Chen, I. Y., Pierson, E., Rose, S., et al. (2021). Ethical machine learning in healthcare. *Annu. Rev. Biomed. Data Sci.* 4, 285–318. doi: 10.1146/annurev-biodatasci-092820-114757

Cinnamon, J. (2019). Data inequalities and why they matter for development. *Inf. Technol. Dev.* 26, 214–233. doi: 10.1080/02681102.2019.1650244

Couture, V., Roy, M. C., Dez, E., Laperle, S., and Bélisle-Pipon, J. C. (2023). Ethical implications of artificial intelligence in population health and the public's role in its governance: Perspectives from a citizen and expert panel. Journal of Medical Internet Research, 25:e44357. doi: 10.2196/44357

Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*: Yale University Press. doi: 10.12987/9780300252392

Crawford, K., and Calo, R. (2016). There is a blind spot in AI research. *Nature* 538, 311–313. doi: 10.1038/538311a

Criado-Perez, C. (2019). Invisible women: Data Bias in a world designed for men. New York: Abrams.

Davis, W. (2024) OpenAI won't watermark ChatGPT text because its users could get caught. The Verge. Available at: https://www.theverge.com/2024/8/4/24213268/openai-chatgpt-text-watermark-cheat-detection-tool

Dencik, L., Hintz, A., Redden, J., and Warne, H. (2018). Data scores as governance: Investigating uses of civic scoring in public services: Datajustice Project. Available at: https://datajusticelab.org/wp-content/uploads/2018/12/data-scores-as-governance-project-report2.pdf

Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issue. *Ethics Inf. Technol.* 20, 1–3. doi: 10.1007/s10676-018-9450-z

Directorate-General for Research and Innovation (2010). European textbook on ethics in research. *European Commission*. doi: 10.2777/51536

Dobbe, R., Gilbert, T. K., and Mintz, Y. (2021). Hard choices in artificial intelligence. *Artificial Intelligence*, 300:103555. doi: 10.1016/j.artint.2021.103555

Dubber, M. D., Pasquale, F., and Das, S. (2020). The Oxford handbook of ethics of AI. *Oxford Academic.* doi: 10.1093/oxfordhb/9780190067397.001.0001

Duffourc, M. N., and Gerke, S. (2023). The proposed EU directives for AI liability leave worrying gaps likely to impact medical AI. *npj Digital Medicine*, 6:77. doi: 10.1038/s41746-023-00823-w

Eitel-Porter, R. (2020). Beyond the promise: implementing ethical AI. *AI Ethics.* doi: 10.1007/s43681-020-00011-6

Elhaddad, M., and Hamam, S. (2024). AI-driven clinical decision support systems: An ongoing pursuit of potential. *Cureus*, 16:e57728. doi: 10.7759/cureus.57728

Elliott, K. C. (2019). Managing value-laden judgments in regulatory science and risk assessment. *EFSA J.* 17:e170709. doi: 10.2903/j.efsa.2019.e170709

European Commission (2015) Horizon 2020 Work Programme 2014–2015: Science with and for Society. Available at: https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/main/h2020-wp1415-swfs_en.pdf

Finnish Center for Artificial Intelligence (2023) What we learned from AuroraAI: The pitfalls of doing ethics around unsettled technologies. FCAI EAB Blog. Available at: https://fcai.fi/eab-blog/2023/12/11/what-we-learned-from-auroraai-the-pitfalls-of-doing-ethics-around-unsettled-technologies

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., and Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI (Berkman Klein Center Research Publication No. 2020-1). doi: 10.2139/ssrn.3518482

Floridi, L., Cowls, J., King, T. C., and Taddeo, M. (2020). How to design AI for social good: seven essential factors. *Sci. Eng. Ethics* 26, 1771–1796. doi: 10.1007/s11948-020-00213-5

Fraenkel, N. F. (2024). Beyond principles: Virtue ethics in AI development: A developer-centric exploration of microethical challenges and empowerment (Master's thesis). *University of Helsinki.* http://urn.fi/URN:NBN:fi:hulib-202410104293

Garcia, P., Ma, S. P., Shah, S., et al. (2024). Artificial intelligence–generated draft replies to patient inbox messages. *JAMA Network Open*, 7:e243201. doi: 10.1001/jamanetworkopen.2024.3201

Gray, D., and Shellshear, E. (2024). Why data science projects fail: The harsh realities of implementing AI and analytics, without the hype: CRC Press.

Hacker, P. (2018). Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law. *Common. Mark Law. Rev.* 55, 1143–1185. Available at: https://ssrn.com/abstract=3164973

Hagendorff, T. (2020). The ethics of AI ethics: an evaluation of guidelines. *Mind. Mach.* 30, 99–120. doi: 10.1007/s11023-020-09517-8

Hasselbalch, G. (2021). Data ethics of power: A human approach in the big data and AI era: Edward Elgar Publishing.

Ho, A. (2020). Are we ready for artificial intelligence health monitoring in elder care? *BMC Geriatr.* 20:358. doi: 10.1186/s12877-020-01764-9

Horgan, D., Romao, M., Morré, S. A., and Kalra, D. (2020). Artificial intelligence: Power for civilisation—and for better healthcare. Public Health Genomics, 22:Article 5–6. doi: 10.1159/000504785

Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* 1, 389–399. doi: 10.1038/s42256-019-0088-2

Keller, P., and Drake, A. (2020). Exclusivity and paternalism in the public governance of explainable AI. *Comput. Law Secur. Rev.* 40:105490. doi: 10.1016/j.clsr.2020.105490

Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., and Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 3:100074. doi: 10.1016/j.caeai.2022.100074

Khosravi, M., Zare, Z., Mojtabaeian, S. M., and Izadi, R. (2024). Artificial intelligence and decision-making in healthcare: A thematic analysis of a systematic review of reviews. *Health Services Research and Managerial Epidemiology*, 11:23333928241234863. doi: 10.1177/23333928241234863

Lam, M. S., Pandit, A., Kalicki, C. H., Gupta, R., Sahoo, P., and Metaxa, D. (2023). Sociotechnical audits: Broadening the algorithm auditing lens to investigate targeted advertising. *Proceedings of the ACM on Human-Computer Interaction, 7(CSCW2), Article* 360, 1–37. doi: 10.1145/3610209

Lauer, D. (2020). You cannot have AI ethics without ethics. *AI Ethics* 1, 21–25. doi: 10.1007/s43681-020-00013-4

Leikas, J., Johri, A., Latvanen, M., Wessberg, N., and Hahto, A. (2022). Governing ethical AI transformation: a case study of AuroraAI. *Front. Artif. Intell.* 5:836557. doi: 10.3389/frai.2022.836557

Leslie, D. (2019). Understanding artificial intelligence ethics and safety. *arXiv preprint arXiv:1906.05684.*

Liao, B., Ma, Y., and Lei, R. (2023). Analysis of ethics dumping and proposed solutions in the field of biomedical research in China. *Front. Pharmacol.* 14:1214590. doi: 10.3389/fphar.2023.1214590

Lindgren, S. (2023). Critical theory of AI: John Wiley & Sons.

Matheny, M., Thadaney Israni, S., Ahmed, M., and Whicher, D. (Eds.) (2019). Artificial intelligence in health care: The hope, the hype, the promise, the peril. National Academy of Medicine: The Learning Health System Series.

McDermid, J. A., Jia, Y., Porter, Z., and Habli, I. (2021). Artificial intelligence explainability: The technical and ethical dimensions. Philosophical Transactions of the Royal Society A, 379:20200363. doi: 10.1098/rsta.2020.0363

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nat. Mach. Intel.* 1, 501–507. doi: 10.1038/s42256-019-0114-4

Mittelstadt, B. D., Allo, P., Taddeo, M., et al. (2016). The ethics of algorithms: mapping the debate. *Big Data Soc.* 3:205395171667967. doi: 10.1177/2053951716679679

Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., and Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, 2. doi: 10.1016/j.patter.2021.100347

ODSC-Open Data Science (2018) Meet Aurora—Finland's AI assistant answer to Siri and Alexa. Medium. Available at: https://odsc.medium.com/meet-aurora-finlands-ai-assistant-answer-to-siri-and-alexa-f82b3f14b553

ÓhÉigeartaigh, S. S., Whittlestone, J., Liu, Y., et al. (2020). Overcoming barriers to cross-cultural cooperation in AI ethics and governance. *Philos. Technol.* 33, 571–593. doi: 10.1007/s13347-020-00402-x

Pacia, D. M., Ravitsky, V., Hansen, J. N., Lundberg, E., Schulz, W., and Bélisle-Pipon, J. C. (2024). Early AI lifecycle co-reasoning: Ethics through integrated and diverse team science. The American Journal of Bioethics, 24, 86–88. doi: 10.1080/15265161.2024.2377106

Powles, J., and Nissenbaum, H. (2018) The seductive diversion of 'Solving' Bias in artificial intelligence. OneZero. Available at: https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53

Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. *Ethics Inf. Technol.* 20, 5–14. doi: 10.1007/s10676-017-9430-8

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., and Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 33–44). doi: 10.48550/arXiv.2001.00973

Rakova, B., Yang, J., Cramer, H., and Chowdhury, R. (2021). Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. *Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), Article* 7, 1–23. doi: 10.1145/3449081

Riezebos, S., Saxena, R., Gelissen, T., Oliveira, M., Sibal, P., Dreier, V., et al. (2022). Multistakeholder AI development: 10 building blocks for inclusive policy design. *UNESCO & Innovation for Policy Foundation.* https://unesdoc.unesco.org/ark:/48223/pf0000382570

Russell, S., Dewey, D., and Tegmark, M. (2015). Research priorities for robust and beneficial artificial intelligence. *AI Mag.* 36, 105–114. doi: 10.1609/aimag.v36i4.2577

Ryan, M., and Stahl, B. C. (2020). Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *J. Inf. Commun. Ethics Soc*. doi: 10.1108/JICES-12-2019-0138

Samuel, G., and Derrick, G. (2020). Defining ethical standards for the application of digital tools to population health research. *Bull. World Health Organ.* 98, 239–244.

Schmitt, L. (2022). Mapping global AI governance: A nascent regime in a fragmented landscape. *AI Ethics*, 2, 303–314. doi: 10.1007/s43681-021-00083-y

Schroeder, D., Chatfield, K., Singh, M., et al. (2019a). "Ethics dumping and the need for a global code of conduct" in Equitable Research Partnerships (Springer), 1–4. doi: 10.1007/978-3-030-15745-6_1

Schroeder, D., Chatfield, K., Singh, M., et al. (2019b). "Exploitation risks in collaborative international research" in Equitable Research Partnerships (Springer), 37–50. doi: 10.1007/978-3-030-15745-6_5

Schroeder, D., Cook Lucas, J., Fenet, S., and Hirsch, F., (eds.). (2016). "Ethics dumping" – paradigmatic case studies, a report for TRUST. Available at: http://trust-project.eu/ethics-dumping-trusts-report-on-paradigmatic-case-studies/

Shilton, K. (2018). Values and ethics in human-computer interaction. *Found Trends Hum. Comput. Interact.* 12, 107–171. doi: 10.1561/1100000073

Slota, S. C., Fleischmann, K. R., Greenberg, S., Verma, N., Cummings, B., Li, L., and Shenefiel, C. (2021). Many hands make many fingers to point: Challenges in creating accountable AI. *AI & SOCIETY*. doi: 10.1007/s00146-021-01302-0

Stafie, C. S., Sufaru, I. G., Ghiciuc, C. M., Stafie, I. I., Sufaru, E. C., Solomon, S. M., and Hancianu, M. (2023). Exploring the intersection of artificial intelligence and clinical healthcare: A multidisciplinary review. *Diagnostics (Basel)*, 13:1995. doi: 10.3390/diagnostics13121995

Taddeo, M., and Floridi, L. (2018). How AI can be a force for good. *Science* 361, 751–752. doi: 10.1126/science.aat5991

Tallberg, J., Erman, E., Furendal, M., Geith, J., Klamberg, M., and Lundgren, M. (2023). The global governance of artificial intelligence: Next steps for empirical and normative research. *International Studies Review*, 25. doi: 10.1093/isr/viad040

The White House (2022) Blueprint for an AI bill of rights. Office of Science and Technology Policy. Available at: https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Tinnirello, M. (ed.) (2022). "The Global Politics of Artificial Intelligence". Boca Raton: Chapman and Hall. doi: 10.1201/9780429446726

Umbrello, S. (2024). Technology ethics: Responsible innovation and design strategies: John Wiley & Sons.

Vakkuri, V., Kemell, K. K., Kultanen, J., Siponen, M., and Abrahamsson, P. (2019). Ethically aligned design of autonomous systems: Industry viewpoint and an empirical study. *arXiv preprint* arXiv:*1906.07946*.

Whittlestone, J., Nyrup, R., Alexandrova, A., and Cave, S. (2019). The role and limits of principles in AI ethics: towards a focus on tensions. *Proc AAAI/ACM Conf AI Ethics Soc.* doi: 10.1145/3306618.3314289

Wirtz, B. W., Weyerer, J. C., and Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. *Government Information Quarterly*, 39:101685. doi: 10.1016/j.giq.2022.101685

Zhang, D., Mishra, S., Brynjolfsson, E., et al. (2021) The AI index 2021 annual report. AI index steering committee, human-centered AI institute, Stanford University. Available at: https://aiindex.stanford.edu/ai-index-report-2021/,