Topoi (2023) 42:867–880
https://doi.org/10.1007/s11245-022-09874-2

Decolonizing AI Ethics: Relational Autonomy as a Means to Counter AI Harms

Sábëlo Mhlambi[1,4,5] · **Simona Tiribelli**[2,3]

Accepted: 18 December 2022 / Published online: 8 February 2023

Abstract
Many popular artificial intelligence (AI) ethics frameworks center the principle of autonomy as necessary in order to mitigate the harms that might result from the use of AI within society. These harms often disproportionately affect the most marginalized within society. In this paper, we argue that the principle of autonomy, as currently formalized in AI ethics, is itself flawed, as it expresses only a mainstream mainly liberal notion of autonomy as rational self-determination, derived from Western traditional philosophy. In particular, we claim that the adherence to such principle, as currently formalized, does not only fail to address many ways in which people's autonomy can be violated, but also to grasp a broader range of AI-empowered harms profoundly tied to the legacy of colonization, and which particularly affect the already marginalized and most vulnerable on a global scale. To counter such a phenomenon, we advocate for the need of a relational turn in AI ethics, starting from a relational rethinking of the AI ethics principle of autonomy that we propose by drawing on theories on relational autonomy developed both in moral philosophy and Ubuntu ethics.

Keywords Artificial intelligence · AI ethics · AI ethics principles · Decolonial AI · Relational ethics · Relational autonomy · Ubuntu ethics

1 Introduction

In the last decades, a growing corpus of literature in the field of ethics of artificial intelligence (AI) has been developed to emphasize the importance of using AI for good and a good society through the design of AI systems according to ethical principles (Floridi et al. 2018), insofar as they can help prevent the harms that AI can cause or exacerbate, especially for the globally marginalized and oppressed. However, if the use of AI is to benefit society, particularly the historically marginalized globally, the discipline and use of AI ought to be decolonized. The *decolonizing AI movement*[1] is a movement that responds to this call partly by categorizing AI and specifically algorithmic decision-making (ADM) systems as an implementation of the principles of modernity, whose documented inequalities and inefficacy stem from colonization. These deep inequalities imposed on the non-Euro/American paternalistic and racialized social hierarchies, ways of living, being, and sensing (Mbembe 2019). Accordingly, if AI ethics frameworks discount the colonial legacy in creating these inequalities, they fail to meaningly address them, and to promote AI as a force for good and more just societies.

The principle of autonomy is one core principle of popular AI ethics frameworks that needs to be decolonized and reimagined in a way that provides value to the marginalized. Indeed, AI's ability to mediate and automate social interactions through machine-learning (ML) algorithms and massive digital platforms is at odds with the traditional notion of autonomy. This traditional principle of autonomy, where

✉ Simona Tiribelli
simona.tiribelli@unimc.it

1 Harvard Law School, Harvard University, Cambridge, MA 02138, USA

2 Department of Political Sciences, Communication, and International Relations, University of Macerata, Macerata 62100, Italy

3 Institute for Technology and Global Health, Cambridge, MA 02139, USA

4 Berkman Klein Center for Internet & Society, Harvard University, Cambridge, MA 02138, USA

5 Digital Civil Society Lab, Stanford University, Stanford, CA 94305, USA

[1] To deepen the ethical mission of the *Decolonial AI movement*, see its *manifesto* at https://manyfesto.ai.

individuals rationally decide between the personal costs and benefits of alternative courses of actions (MacIntyre 1988), is weakened through algorithms and platforms that use patterns inferred from large amounts of data to predetermine or direct a users' access to content and social interactions (consider algorithmic techniques such as profiling and personalization like ML-based collaborative filtering) (Tiribelli 2022). This diminishment of human autonomy also undermines internationally-held human rights such as the right to opinion, information, and privacy. In the case of privacy, for example, the nexus of maximizing advertising revenue by recommended content that responds to users' preferences (i.e., empowering decision-making according such traditional concept of autonomy) is enabled through massive surveillance (Zuboff 2019) and extraction of users' data at a colonial like scale (Couldry and Mejias 2019).

However, strict adherence to this traditional notion of autonomy, through technical or policy implementations, is insufficient in achieving equitable outcomes for those who disproportionately are negatively affected by the use of AI. Indeed, to correctly confront the harms exacerbated by AI and maximize the benefits of AI to all, the principle of autonomy should be renegotiated and decolonized from its enlightenment era design, along with the corresponding relationship AI and platforms have within and towards society.

As post-colonial literature points out (Mbembe 2019; Mhlambi 2020), the principle of autonomy has been originally designed to justify and protect the demands and rights of the European/American individual against the oppressive weight of monarchical, feudalistic, and religious institutions. However, in our globalized societies, this foundational conception of autonomy is ethically, geographically, and socially limited in scope and efficacy and therefore inept to adequately confront the plural social harms worsened by AI. The traditional principle of autonomy derives from the enlightenment-era concept of personhood as rationality, and rationality as an individual endeavor historically native to, and the vanguard of, European men, and can be best exercised through individual autonomy (Ramose 1999; Mbembe 2019). This foundational definition and implementation of rationality is ethically inept and for some also morally wrong, insofar as—as many postcolonial scholars argue (Mbembe 2019)—it begins its use via the exclusion and the dehumanization of the non-Euro/American man.

Despite the enlightenment philosophers' fundamental disagreements on the nature of rationality (MacIntyre 1988), rationality was originally applied as the European way of living, thinking, and feeling, often structuring and justifying Euro-American advancements and conquests (Ricaurte 2019; Mignolo and Escobar 2010). The Euro-American colonization of other regions of the world demonstrated the geographic limitation of autonomy as it disregarded the local notions of autonomy from those colonized, while, through colonization, stripped the autonomy of the colonized. Consistent with its problematic conception and historical application, the principle of autonomy within Western-derived AI ethics is a mismatch between the needs of those likely to be marginalized by AI. The mismatch of autonomy as a principle for creating "responsible AI" is also due to modernity that, as many argue, was built on colonization and irrationality (Smith 2019) sustained through neo-colonialism (Nkrumah 1966) and coloniality (Mignolo and Escobar 2010). As a consequence, AI systems, if designed to preserve autonomy, so understood, would only preserve systems of inequality that affect society and directly inform and cause the harms often shallowly attributed to technical flaws in AI models or datasets. Indeed, the current AI ethics principle of autonomy does not consider a priori injustices and inequalities; it exclusively points to examine the decision-making role of an AI system interacting with an individual, while ignoring social aspects that constitute an individual and an individual's ability to make decisions. If AI systems, such as algorithmic recommender systems, limit the autonomy of an individual, the socio-economic determinants that shape how AI is built and which values it encodes, also limit the autonomy of the individual not represented in the ethics, implementation, and values embedded in that AI system.

If the principle of autonomy had been negotiated in an inclusive, universal, and democratized way, it would still be unclear if the current mainstream view of autonomy cited in AI ethics principles would have prevailed. Even if the liberal version of autonomy had prevailed historically and endowed to the rest of humanity, it would still lead to ethical problems with negative social reverberations (Ramose 1999; Molema 1917). Scholars argue how Western philosophers as Locke, Mills, and others justified liberty and read "autonomy" as necessary even if it were to create inequality in wealth and power all the while, without specifying limits to how much of such inequality can be sustained at the expense of inequality (Daniels 1974). US founding fathers also believed liberty and equality were in an eternal state of conflict and that liberty should prevail if a choice between the two should be made (Pittman 1960). It is today still unclear how much inequality the principle of autonomy, as applied in AI ethics, would dictate as necessary in order to preserve liberty. However, the disproportionate AI-perpetuated harm experienced by historically racialized and marginalized populations (Benjamin 2019) indicates that it may be necessary to have a conception of autonomy that considers social relationships. Indeed, a concept of autonomy deprived of positive obligations to society that actively addresses gaps in equity at best maintains the inequalities in the wealth, economic structures, and social hierarchies stemming from colonization (Kunene 1981, 1982; Ramose 1999). But, as scholars have recently highlighted, the impact of colonization cannot be treated as a negligible variable in the development and

use of AI within society (Mohamed et al. 2020). Equally so, the ethical principles designed to lead to the responsible use of AI cannot afford to ignore the *status quo* shaped by colonization. In this sense, a relational concept of autonomy is a crucial step in decolonizing AI and restoring autonomy to those whom it has been systematically denied.

In this paper we claim that to advance the benefits of AI systems within society while decreasing the harms inherent in the development and use of AI, the AI ethics principle of autonomy should be understood and practiced within the context of relationality. That is considering the role of society in the development of an individual and the individual participation within society as means to achieve a higher degree of autonomy that benefits individuals and society. To this aim, a first section of the paper is devoted to highlight the specific concept of autonomy informing many AI ethics frameworks and the scholarship in AI ethics and to argue its limits and the need for a relational turn in its understanding in order to design AI systems in a truly inclusive way. The second and third sections are devoted to the novel introduction in the debate in AI ethics of the concept of relational autonomy as a mean to decolonize AI (ethics). We first draw attention on how there are meaningful contributions on relational autonomy developed also in Western non-mainstream philosophy, whose consideration is crucial to revise the AI ethics concept of autonomy in a more adequate way. Then, we zoom in on a non-Western, non-ethnocentric concept of relational autonomy, as embraced in African philosophy of Ubuntu ethics and highlight its value to revise the AI ethics concept of autonomy in a more inclusive way.

2 Autonomy in AI Ethics: On Its Limits and the Need for a Relational Turn

Autonomy is one of the most widely acknowledged ethical principles in the field of AI and ethics (Mittelstadt et al. 2016; Tsamados et al. 2022; Jobin et al. 2019), namely, within the scholarship and initiatives focused on how to design responsible AI, by minimizing its risks while harnessing its benefits for individuals and societies. While autonomy has always played a central role both in political and moral philosophy (Raz 1986; Korsgaard 1996; Mackenzie and Stoljar 2000b; Roessler 2021) and in applied ethics, from bioethics to medical ethics (Taylor 2009; Beauchamp and Childress 2013), its specific import in the field of AI and ethics is especially due to a recent and large number of risks raised by the design and uses of AI systems in a wide array of domains, from social media and online marketing to healthcare, education, and justice.

Indeed, on one side, AI systems can open up novel opportunities to support and promote human autonomy, ranging—just to mention a few of them—from empowering very often limited human time resources via the automation of time-consuming repetitive tasks (Floridi et al. 2018), or via sorting information to navigate data and informational overloads (Mittelstadt et al. 2016), up to enabling the discovery of meaningful knowledge and correlations. However, on the other side, a number of autonomy-undermining AI-enabled phenomena has been also widely shown in the debate on AI and ethics. Such phenomena extend from AI-facilitated risks of individuals' deception, manipulation, and coercion (Susser et al. 2019; Jonjepier and Klenk 2022), especially enabled by the deployment of ML profiling techniques to exploit users' emotion and/or vulnerabilities to meet third-party goals above the individual's[2](Applin and Fischer 2015; Newell and Marabelli 2015; Zarsky 2016; Helberger 2016; O'Neil 2016; Milano et al. 2020), to the narrowing of the range of informational choice options (Royakkers et al. 2018; Tiribelli 2020, 2023) and informational diversity (Newell and Marabelli 2015; Pariser 2011; Sunstein 2008)—both pre-conditions for autonomy (Van den Hoven and Rooksby 2018)—via personalization techniques and the construction of paternalistic algorithmic choice-architectures (Tene and Polonetsky 2013; Tiribelli 2023), up to morally unjustified opportunity losses or exclusion, from the access to a job opportunity or a college admission, up to the preclusion of an healthcare facility or of freedom itself (probation), due to gender and racial bias embedded via data in AI systems (Angwin et al. 2016; Dastin 2018; Obermeyer et al. 2019; Simonite 2020).

Such ethical risks have made the protection of autonomy a core AI ethics pillar in the majority of existing AI ethics frameworks, guidelines, and recommendations (Jobin et al. 2019). For example, the European Commission (EC)'s High-Level Expert Group on AI (HLEG-AI) and the World Health Organization (WHO) list the 'protection of autonomy' as the first in their list of key ethical principles for the development of trustworthy and ethical AI (HLEG-AI 2019, p. 8; WHO 2021, p. 25). Other AI ethics initiatives and documents, such as the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, the Montreal Declaration for Responsible Development of Artificial Intelligence, the Association for Computing Machinery's Code of Ethics, and many more (Jobin et al. 2019) stress similarly the pivotal import to respect human autonomy as a priority in the design, implementation, and use of AI.

However, despite the centrality of the protection of autonomy in the debate on AI ethics, the concept of autonomy itself in leading AI ethics frameworks and policy guidelines is still very much opaque, ambiguous, and very often

[2] Consider, for example, the renowned case of Cambridge Analytica (Rosenberg 2018) or the Facebook experiment on emotion contagion (Kramer et al. 2014).

undefined (Prunkl 2022), making also unclear what are both the sub-conditions and measures required to protect autonomy, that is, what autonomy as AI ethics principle demands. This is because the philosophical complexity and the cultural richness of a pivotal ethical concept as that of autonomy are poorly explored in AI ethics. The lack of such an in-depth ethical inquiry on autonomy is problematic as it hampers a proper understanding of all the many and diverse risks that AI can raise to human autonomy, that is, all the ways in which human autonomy can be non-respected and violated by AI and especially the autonomy of *whom* is mostly impacted by that. Such unclarity, in turn, hinders the design of AI systems and policy guidelines that can effectively promote human autonomy in adequate ways, that means also in different social conditions across diverse geographies and according to different meanings that autonomy acquires in plural and heterogeneous cultural contexts, given the transnational nature and application of AI.

This is evident if we consider, for example, an AI ethics benchmarking document as the *Guidelines for Trustworthy AI*, where HLEG-AI exclusively refers to autonomy as the individuals' "capacity to keep full and effective self-determination over themselves, and be able to partake in the democratic processes" (2019, p. 12). As it has been shown (Pariser 2011; Sunstein 2008), when AI is designed to promote individuals' self-determination, for example, by showing informational options according to filtered information based on individuals' preferences, it can paradoxically also undermine autonomy according to such definition, by eroding some key individuals' capacities to adequately participate in democratic processes (e.g., public sphere dialogue and joint pursuit of societal goals), such as the capacity of open mutual understanding, recognition, and cooperation (Giovanola and Sala 2021). Jobin et al. (2019) show such unclarity and vagueness on autonomy in their analysis of more than 80 AI ethics frameworks of principles and guidelines developed at the global level on AI, where autonomy emerges widely and it is mainly understood in terms of "self-determination", "informational self-determination" and/or "privacy-preserving human control and oversight" on AI, or in terms of "freedom to withdrawn consent" or as "freedom from exploitation, manipulation, and surveillance" (p. 11). Following such definitions, unsurprisingly, the most prominent methods to operationalize the respect of autonomy consist of "giving people notice and consent", "refraining from data collecting and spreading data in absent of informed consent", "not reducing options and knowledge of citizens", "increasing people's knowledge on AI", and in "transparency and predictable AI" (Jobin et al. 2019, p.11).

As we will highlight broadly in the next two sections, it is highly questionable that from an ethical, social, and geographical standpoint autonomy exclusively overlaps with (informational) self-determination, as well as that the measures and conditions mainly proposed to protect it—e.g., transparency, AI data literacy, and privacy techniques, nonetheless per se important—are truly sufficient or adequate for its respect. Indeed, beyond such multiplicity of vague definitions and circumscribed measures proposed within different frameworks, it is questionable that the respect of human autonomy can be reduced, as it mainly emerges, to maintain people's full control over themselves and AI systems (Floridi et al. 2018; IEEE 2019; HLEG-AI 2019; OECD 2019; Fjeld et al. 2020), as well as to privacy-preserving techniques based on informed consent (European Parliament 2017; IEEE 2017; Google AI 2018; WHO 2021). The inefficacy of people's digital informed consent has been widely argued in the AI ethics' literature, showing how it can be easily bypassed and therefore cannot guarantee the protection of individuals' autonomy and identity, as well as the prevention of phenomena of unfair discrimination via AI (Wachter 2020). Moreover, as Peña and Varon (2019) argue, digital informed consent—and data protection techniques based on that—has been uncovered to be a colonial tool to strengthen, perpetuate, and silently legitimate historical autonomy-abusive practices for body control, oppression, and exclusion on the basis of racial and gender characteristics.

Beyond the necessity of an in-depth understanding of what autonomy as ethical concept and principle consists of and/or demands, an ethical question we should raise concerns specifically the autonomy of whom such definitions or formalizations of autonomy adopted in AI ethics aim to and can effectively preserve. Indeed, the lowest common denominator on autonomy that emerges from the majority of AI ethics frameworks and initiatives (Jobin et al. 2019) and from the consistent corpus of literature focused on risks raised by AI on autonomy in AI ethics (Mittelstadt et al. 2016; Floridi et al. 2018; Milano et al. 2020; Calvo et al. 2020; Floridi and Cowls 2019; Tsamados et al. 2022) is its main understanding via the lens of the traditional (also known as mainstream) Western philosophy, where autonomy is mainly grounded on individual's capacity of rational deliberation amongst alternative options, the capacity of self-governance and control via the exercise of reflective endorsement on one's own motives, reasons, preferences, and beliefs (emotions included) motivating an action. In this sense, the concept of autonomy in AI ethics is limited as it is mainly confined to a traditional conception of autonomy rooted in rational deliberation and decision-making, lacking of an in-depth recognition of the precious contributions on autonomy offered within Western philosophical scholarship by feminist and relational scholars.

As we will see in the next section, however, from within the debate in moral philosophy, such mainstream conception of autonomy, mainly advocated by the Kantian-inspired liberal tradition, has been widely criticized to be individualistic and procedural, to do not account for the relational

contexts and features that inform, shape, and characterize people's autonomy and identity, and also to do not account for those who live in conditions of oppression of various kind. Moreover, scholars in AI ethics critique such understanding of people as rational decision-makers, on which the mainstream liberal conception of human autonomy rests, and therefore the design of AI systems according to such rational model (see Dignum 2022), insofar it does not reflect realistic conditions of bounded rationality (Simon 1991) in which the many live, and because it hides via its veil of rational objectivity those unfair asymmetrical and hierarchical power dynamics of colonial heritage that nurture and silently perpetuate systemic injustice and structural inequalities (Mhlambi 2020; Birhane 2021; Dignum 2022).

In this regard, indeed, scholars in AI ethics are pointing to the need for a relational turn in AI ethics, shifting from a rational to a relational approach, to the extent the latter would allow to center in the design of AI those people that are most at the margins, marginalized, and vulnerable, and thus far the most impacted by AI (Birhane 2021; Mohamed et al. 2020; Mhlambi 2020), along with the need to decolonize AI and AI ethics and policy via non-Western approaches (Mhlambi 2020; Mohamed et al. 2020; Roche et al. 2022). Indeed, as the global inquiry on AI ethics carried by Jobin et al. show (2019), non-Western voices and geographical areas such as Africa, South and Central America, and Central Asia are under-represented in the AI ethics debate,[3] thus reinforcing asymmetries in power, autonomy, and agency of certain more economically developed countries rather than others, and eroding the possibility of cultural pluralism towards global justice. As Mohamed et al. (2020) point out, "risks are likely to arise if we neglect to explore the current variation of ethical standards based on identity and geography" (p. 669). In other terms, if we do not examine AI phenomena by enriching and revising current benchmarking epistemological and axiological ethical paradigms with alternative views, especially those of the more marginalized, situated mainly in the outskirt of Western Euro-American tradition and especially in non-Western perspectives, by centering them in AI research and design practices, as called for by relational and decolonial approaches, we do not only be unable to mitigate unfair historically rooted power asymmetries, but we end to reinforce them via AI, as unable to acknowledge them as risks and for whom (Mohamed et al. 2020, p. 664; McDowell and Chinchilla 2016, Birhane 2021).

[3] This data can be paradoxical if we consider how much such non-Western countries are more populous than Western ones and how they are the most exploited and affected by AI. Mohamed et al. 2020 provide many examples of such exploitation ranging from ghost workers' labor to beta-testing techniques (p. 668).

Adopting a decolonial approach in AI means putting the most marginalized and vulnerable people "who continue to bear the brunt of negative impacts of innovation and scientific progress" at the center of the design of such technology (Mohamed et al. 2020, p. 664). By arguing for a relational revision of AI and ethics, Birhane stresses (2021, p. 7) that for an ethical development of such technology we should center the needs and well-being of those that are disproportionately impacted by AI, instead of on solutions that benefit the majority (objective cost–benefit analysis). Power asymmetries and structural inequalities indeed happen in social and relational practices and contexts, therefore, adopting a relational approach to AI and AI ethics principles sounds like the only means to detect and decolonize them, therefore, to design AI systems that can discover and compensate morally wrong and unfair historical power asymmetries by inquiring the relational contexts and practices in which they arose, develop and perpetuate. For example, a relational approach to AI ethics entails overcoming the idea of solving unfair AI just by fixing bias in AI via statistical neutral or parity models; it asks, instead, to use AI to discover and investigate why such bias emerges in certain contexts, what historically asymmetrical relations of power they reflect, and what and who continue to nurture them (Birhane 2021; Giovanola and Tiribelli 2022a, 2022b). Therefore, we are called to investigate such relational contexts and practices to understand how to empower the most marginalized and vulnerable via AI.

To sum up, to mitigate structural power asymmetries and inequalities, and therefore empower the most marginalized and vulnerable via the design and use of AI, we must understand how to empower their autonomy considering their oppressed agency conditions. It clearly follows that, as we claim and show in the following sections, to do so, we need a relational concept of autonomy.

To respond to the above-mentioned increasingly numerous calls for non-mainstream, relational, and non-Western approaches to AI ethics principles, criteria, and guidelines for the development of truly responsible and inclusive AI, in the next two sections we accomplish a double-level conceptual operation, with the ultimate goal to propose a revision of the current mainly ethnocentric and mainstream concept of autonomy informing the relative AI ethics principle, as mainly overlapping with human control, rational self-governance, and informational self-determination.

More specifically, first, we pursue an ethical inquiry into the concept of autonomy, by drawing on theories on autonomy developed also in non-mainstream Western philosophy to clarify that autonomy and rational self-determination do not overlap, and then provide a more complex account of autonomy that considers also the precious contributions offered by relational scholars (e.g., feminists and communitarians) on autonomy, which are currently at the

outskirt of the debate in the field of AI ethics. Our goal is to show that the current main formalizations of autonomy as rational self-determination, self-governance, and control are partial—and therefore, as such, inadequate—insofar as they exclusively rest on a purely mainstream or traditional (known as "standard", and by critics, as individualistic) liberal account of autonomy, rooted in the Kantian-inspired conception of people as rational decision-makers capable to self-determine themselves on their own, if provided with the right sub-conditions (competency and authenticity) to do so. We will introduce the limits of such accounts according to relational perspectives in order to shed light on the necessity to consider also relational features of autonomy. In particular, we will emphasize the relational dimension of autonomy and its sub-conditions, as introduced by non-mainstream Western accounts on autonomy, insofar as they consider realistic conditions of limit and social oppression in which individuals especially the most marginalized and vulnerable mostly live—that is crucial if we want to design AI systems that can truly respect and promote the autonomy of people in a more effective and, most importantly, inclusive way.

Second, after having unpacked the complexity of the ethical concept of human autonomy, going beyond the exclusively traditional one via non-mainstream relational accounts of autonomy, we zoom in on the concept of relational autonomy developed within the non-Western perspective of Ubuntu philosophy. We argue for its value to revise and specifically decolonize the concept of autonomy underpinning the AI ethics principle of autonomy by acknowledging historically rooted power asymmetries in the design of responsible AI capable to counteract morally unfair forms of injustice and oppression.

By doing so, we aim at paving the way for the debate in AI ethics to accomplish the revision of the concept of autonomy underpinning the AI ethics principle, by considering both a non-mainstream Western account of relational autonomy, which is born as a reaction against traditional mainly liberal accounts of autonomy rooted mainly in the conception of rationality; as well as by considering a non-Western account of relational autonomy, born in its own right as a way of life and widely practiced in many African countries, as in the case of Ubuntu (Menkiti 1984; Mhlambi 2020; Birhane 2021).

3 Beyond Autonomy as Rational Self-Determination: Relational Autonomy from Within Moral Philosophy

In the previous section, we have highlighted how the concept of autonomy informing many AI ethics frameworks and scholarship emerges as mainly opaque and vague, leading to a confusing patchwork of fuzzy definitions and conditions or measures to protect autonomy, which have been argued to be often ineffective or incompatible (Jobin et al. 2019). We have also pointed out that such opacity is linked to an absence in AI ethics of a systematic ethical inquiry into the ethical concept of autonomy underpinning the relative core AI ethics principle capable of clarifying what autonomy consists of and demands. In this section, we draw insights from theories on autonomy developed in moral philosophy to fill this gap.

We have also previously outlined that the concept of autonomy underpinning both the corpus of literature analyzing risks from AI systems to human autonomy and the current few vague formalizations in AI ethics frameworks rarely points to relational features, while it tends to overlap with individual self-determination, which in turn is grounded on a model of the rational person, which has been shown to be inadequate (Dignum 2022) and leading to obscure, perpetuate, and strengthen instead of fixing structural inequalities (Birhane 2021; Mhlambi 2020). In this section, thanks to an ethical inquiry into the concept of autonomy, we provide a more complex account of autonomy as an ethical concept, showing that autonomy and self-determination do not overlap. To do so, we unpack the traditional ethical concept of autonomy and its main sub-requirements according to the liberal tradition and then introduce non-mainstream voices coming from feminist ethics and communitarianism who advocate for a relational account of autonomy within Western ethics. We conclude by pinpointing key aspects and sub-conditions of relational autonomy and stressing the need of considering them to integrate and revise the standard liberal (individualistic) account of autonomy in the light of a more complex and adequate concept of autonomy in AI ethics.

Since Kant (1785), the connotation of autonomy as rational self-legislation or, in its contemporary conceptualization, as rational self-determination underpins the Western traditional moral thought, and more specifically, it constitutes one of the key ethical cornerstones of the liberal tradition (Christman and Anderson 2005),[4] expressed through the formula that recites "act with reason", adopted by many scholars also to underline the Kantian rationalism that rests at its core (Colburn 2010; Christman and Anderson 2005). Indeed, albeit with some internal variations, the prominent notion of autonomy emerging from the liberal tradition, also called as the traditional view, considering its long-standing roots—and that then has become mainstream due to its large influence –, is

[4] Since a detailed survey of the perspectives that will be mentioned is beyond of the scope of our analysis, i.e., to shed light on the richness of Western moral philosophy on autonomy and on diverse accounts developed within Western ethics beyond those offered by the liberal tradition, we focus on representative samples of these accounts. For their in-depth analysis, see Christman and Anderson (2005).

described as rational self-government, control, and independence, and it is rooted in the individuals' capacity to choose, act, and behave according to interests, reasons, beliefs, and values they can reflectively endorse. Such *reflective endorsement* that implies a rational appraisal or deliberation is in fact a "constraint" to which most liberal accounts on autonomy are committed (Kymlicka 1989; Dworkin 2000): autonomy is described as respected if and only if people can embrace and endorse, that is, rationally appraise, critically reflect, identify with, or rejects and modify, what guides their choices and actions, and in this way, self-determine themselves. In this sense, liberal autonomy tends to overlap with self-determination which, in turn, is grounded in the individuals' rational and deliberative capacity to be in control of their choices, and therefore self-govern themselves, by steering their behavior according to factors or action-guiding norms that are somehow their own (Frankfurt 1989; Dworkin 1988; Ekstrom 1993; Michael 2005; Korsgaard 2014; Thomas 2017). While there is enough agreement about the overall idea of autonomy as self-determination within the liberal tradition, less accordance emerges instead on what sub-conditions autonomy demands. Killmister (2017) helps to zoom in this debate and clarifies that (at least) two main families of conditions have been proffered within the liberal tradition: *competency* conditions and *authenticity* conditions. The former mainly refers to a set of competencies that individuals should own to be defined as autonomous, that is, to self-rule and self-govern, which range from rational thought and self-control to freedom from debilitating pathologies and systematic self-deception. Authenticity conditions refer instead to the capacity to critically reflect upon and endorse one's own desires, preferences, values, and so on. We encounter such conditions in the debate in AI ethics when, by observing core ethical frameworks, autonomy is mainly described as prescribing to keep control over AI systems or as freedom from manipulation and/or undesired interferences on human capacity of rational decision-making (Jobin et al. 2019). For instance, the majority of debate on AI and ethics focuses on how AI systems can violate autonomy by enabling systematic self-deception (Susser et al. 2019; Natale 2021) or by exploiting individuals' non-rational elements such as emotions to steer individuals' behavior according third-party goals (see Prunkl 2022; Tiribelli 2020, 2023). Such a state of "freedom from interferences" translates then into measures to safeguard autonomy that focus on protecting an individuals' space of independence or, in other terms, their informational privacy (conceived as a space where to "let us alone", see Floridi 2011).[5]

Competency conditions are widely criticized within Western moral philosophy (and in bioethics and medical ethics) as expressing only a formal or procedural account of autonomy that pretends to be neutral or objective. To make it clearer: such conditions focus exclusively on the procedure through which a person comes to *rationally* endorse available options (i.e., reasons, values, and motives) and thus implicate that exercising such a set of skills is sufficient to determine autonomy. In this sense, these conditions center the element of rational self-reflection but in a way described as *procedural independence*: such conditions exclude the import of substance of the process, that is, if the options on which a person chooses embed the values and projects she deems *truly* meaningful. The critique to liberal procedural accounts of autonomy is severe especially from feminist ethics' scholars (MacKenzie and Stoljar 2000a, b; Oshana 2006; Benson 1990, 2005) who claim that, insofar as such accounts focus on a rational, neutral, and objective procedure, with no stipulations on the contents of desires or values on which one can choose as options, they can easily legitimize historically rooted values of oppression and perpetuate autonomy-undermining social conditions of gender oppression, objectivation, and curtail of women's options (think about arranged marriages). Therefore, such an exclusive focus on competencies results problematic for people who live in oppressive or constraining socio-economic and health conditions such as, for instance, those affected by systemic and epistemic injustice (Fricker 2007) or by physical decline or debilitating health pathologies (Jaworska 2009).[6]

There is much more agreement instead on the second family of conditions above-mentioned, that is, authenticity conditions. Such conditions express the possibility of the individuals' to be autonomous to the extent they can identify with commitments, beliefs, and values by reflectively endorsing them as motives of their choices, actions, and behaviors (MacKenzie 2014). In particular, authenticity conditions stress that autonomy is respected only when the options on which a person can exercise her reflective endorsement substantially incorporate the values and projects she deems as being meaningful. In this sense, such conditions center self-reflection but in a more substantial way (i.e., they reflect a *substantial independence*). However, authenticity conditions are controversially debated too, especially in relation to the concept of self-identification via reflective endorsement (to expand see Veltman

[5] A widespread idea in AI ethics is that individuals' identity can be expressed in informational terms ("we are our information") and therefore its protection requires informational privacy, as safeguarding human control over personal data and rational decision-making (Floridi 2011).

[6] In addition, an exclusive focus on competencies is also at odds with recent discoveries in cognitive sciences (see e.g., Thaler and Sunstein 2009; Kahneman 2011; Simon 1991) according to which individuals rarely choose in optimal conditions, and therefore as rational decision-makers, but rather, in conditions of limited cognitive and time resources that lead them to be very often rationally bounded and biased decision-makers.

and Piper 2014). Specifically, substantial independence is criticized to result as still incapable to account for people living in oppressive conditions; for example: if a person is oppressed, the reflective judgments she makes are tainted by that oppression, therefore, her reflective voices can be just an act of rationalization or self-deceit rather than an authentic expression of her character.

Overall, liberal accounts of autonomy as rational self-determination and independence have been severely criticized within Western moral philosophy by scholars coming both from feminist ethics and communitarianism (see Oshana 1998; MacKenzie and Stoljar 2000a, 2000b; Westlund 2009; Gutman 1985; Taylor 1991; Sandel 1982; MacIntyre 1988) to focus exclusively on an individualistic dimension of autonomy, whereby individuals are conceived to be independent and self-isolated. From a communitarian perspective, for example, Sandel criticizes the liberal "disencumbered" conception of the self and of the individuals described as autonomous choosers, by stressing the importance to consider and investigate the social preconditions of human autonomy. Indeed, according to the communitarian view, we cannot neglect that communities are prior to individuals, as well as the associated identity memberships' values and commitments, and therefore avoid to consider how much a person's identity, more than an output of an autonomous choice, is shaped by the belonging to a community and is informed by the related identity. Quoting Taylor (1979, 1991) "we are what we are in virtue of participating in the larger life of our societies" (1979, p. 88) and as follows we should look for conditions of autonomy also in this socio-relational dimension, instead of exclusively in conditions that lead people apart, to isolate, and rationally and independently deliberate. Such atomistic notion of the self and autonomy is also challenged by many feminist scholars who question self-sufficiency and substantive independence as a value that we should really promote (Jaggar 1983). In their view, autonomy as an individualistic concept that does not acknowledge the role of social-relational conditions as required for autonomy is not truly meaningful, as it fails to enable the understanding of social (gender) oppression (Mackenzie 2015) stemming from social-relational conditions and how that can interfere with enjoying autonomy.

As a critical response to this limited (individualistic) account of autonomy, some scholars, especially in feminist ethics, have attempted to rehabilitate autonomy, along with the concept of dependence, as a way to understand and counteract many forms of oppression, starting from the recognition of the key role played by socio-relational dimension in the development and enjoyment of autonomy (Oshana 1998; MacKenzie and Stoljar 2000b; Westlund 2009). Such commitments take the name of 'relational autonomy' (MacKenzie and Stoljar 2000a). Beyond a few conceptual differences, their underpinning argument is that the processes to establish authenticity can only occur in social conditions, to the extent it is through the socio-relational and cultural dimension people develop what have significance for them and truly motivates their agency. Authenticity conditions that—as claimed in many liberal accounts—fail to acknowledge the central role of the relational dimension in the understanding and exercise of human autonomy end to valorize the life of the separated individual and to denigrate the social and interpersonal factors that instead define the possibility for autonomy (Oshana 1998, p. 81). Exponents from relational autonomy stress how, in countless ways, people are constituted by factors that lie beyond their reflective control, but which nonetheless shape their values, thoughts, beliefs, and motivations (Taylor 1991); thus, "we cannot say that we are autonomous only when we can step back from all such connections and critically appraise and endorse them" (Bell 1993, pp. 24–54).

Within Western moral philosophy, some authors by investigating the relational dimension of autonomy have found the conditions to promote relational autonomy in social support or social recognition, insofar as considered critical to preserve and foster individuals' capacity for self-trust, self-esteem, and self-respect, which in turn are argued as necessary for people to properly express their agency (Grovier 1993; Anderson and Honneth 2005; Benson 2005; Westlund 2009). In their view, *authenticity conditions* (via reflective endorsement on meaningful options) and *relatedness conditions* via social support (reciprocity and solidarity) are necessary to respect and foster people autonomy. These contributions advocate for the idea that autonomy requires the ability to endorse one's own values and projects, but that this condition alone is not sufficient to the extent that oppressive conditions of various kind can increasingly threaten autonomy, eroding or removing one's sense of self-confidence required for effective agency. Therefore, it is crucial to consider and work also on relatedness conditions for social support, especially insofar as they can enable (maintain and foster) individuals' self-trusting status, defined as crucial for empowering the autonomy of the most vulnerable and/or oppressed within society, those who are marginalized or live in socio-relational oppressive and/or constraining conditions (Anderson and Honneth 2005; Grovier 1993; Benson 2005; McLeod and Sherwin 2000; Westlund 2009).

To sum up, our ethical inquiry into the concept of autonomy, as developed in mainstream and non-mainstream Western moral philosophy, shows that the current AI ethics principle of autonomy is inadequate, as it is partial also with regard to the Western philosophical tradition of those who have mainly formulated it. From a Western moral philosophy's standpoint, an adequate ethical concept of autonomy would require to consider also its socio-relational dimension and the key conditions identified to promote it. The consideration of such dimension and conditions indeed can

allow to broaden the concept of autonomy that informs the related AI ethics principle in a way that can make it truly effective, that is, in a way that can help to really account for the conditions of limit and oppression in which the majority lives. In the next section, we zoom in on the concept of relational autonomy as spontaneously developed within the African philosophy of Ubuntu and draw attention on key relational elements and relatedness conditions that emerge as similar to those above-mentioned, with the ultimate goal to sketch a conceptual path to decolonize the AI ethics principle autonomy and hence revise it in a more inclusive way.

4 Relational Autonomy in Ubuntu

Ubuntu is a philosophy that we can define as essentially (and ontologically) relational. Indeed, Ubuntu's view is that the universe is organically inextricably interconnected (Lee and Hord 2016; Eze 1997): all that consists of the universe, from the most minute atom to the largest galaxy—and all in between—are all interconnected (Ramose 1999; Kunene 1981, 1982; Ngubane 1979). More importantly, humans are inextricably interconnected to each other through an unbreakable and symbiotic connection maintained throughout time and space (Ramose 1999). Individuals are indeed connected to individuals from the past, present, future and are connected to their environment. Such interconnectedness presents a oneness and a timeless self-similarity (Asante and Abarry 1996): an individual is the community as much as the community is the individual and both the individual and the community co-exist with each other and only through each other (Nyathi 2001; Ngubane 1979). In this sense, the deeply-held African ideal of interconnectedness of humans at the core of Ubuntu implicates a dependency that is understood both as self-preserving and as communal preserving (Kunene 1981; Tutu 1999): a society cannot survive, thrive, or reach ethical maturity without concern for and active participation in the life of the individual (Kunene 1981); similarly, an individual cannot survive, thrive, or reach ethical maturity without concern and active participation in social life (Menkiti 1984). As follows, in Ubuntu, beneficial actions are ethical actions, and ethical actions are actions that both enhance the wellbeing of the individual and that of society (Kunene 1982). In this sense, there is a bidirectional shared reciprocity that takes the form of a duty between the individual and the society.

Indeed, Ubuntu defines personhood as being relational (Ramose 1999): to be an autonomous person, one must first be a person, but a person becomes a person through others ("umuntu ngumuntu ngabantu"—"a person is a person through other persons"). Therefore, to be an autonomous person one must be a relational person one required to have positive duties beneficial to self and in the interest of others (Lee and Hord 2016). However, this does not mean that, in Ubuntu, individual rights are subordinated, but that individual and common (or collective) goods are conceived and pursued inseparably. Indeed, one's humanity is a function of how one enriches or fails to enrich the humanity of others. In virtue of becoming more or less relational, one may become more or less of a person. Personhood is indeed beyond a biological condition: it is a function of one's enrichment of the personhood of others. In turn, positive participation in the personhood of others makes one ethical and a society's participation in enriching the personhood of the individual makes a society ethical (Mhlambi 2020).

In Ubuntu, relational autonomy contains both negative and positive duties that are bidirectional between individuals and society. This means that what one does to others directly affects oneself, and how society treats the individual directly affects the character of society. Autonomy is a function of this shared reciprocity. In this sense, in Ubuntu, autonomy exists only when reciprocity exists between the individual and society. A society that lacks consideration for the individual becomes inhumane and an individual that refuses social duties to others becomes less human (Molema 1917). An individual, dehumanized in an inhumane society, cannot be said to have autonomy and the autonomy or liberty the society uses in dehumanizing an individual can be characterized as oppressive. In this view, an individual at odds with the wellbeing of society may be acting inhumanely. To sum up: relatedness and reciprocity are sub-condition required for humaneness and autonomy. Reciprocity, in turn, requires closeness, oneness, active mutually beneficial involvement as key sub-conditions to promote relational autonomy.

It follows that, in Ubuntu, an individual cannot be self-complete and therefore autonomous in isolation (Molema 1917). Another condition for an individual to enjoy relational autonomy is the active participation of the community in order to have autonomy and one cannot receive the community's participation if one distances oneself from the community (Ndlovu-Gatsheni 2019). As a key condition to enjoy autonomy, the individual must also work for the benefit of the community for it is the union of the community and individual that allows for an individual to exercise autonomy. This places an ethical mandate on the purpose of autonomy, for if an individual acting in self-interest, undermines society directly, or indirectly, it removes the preconditions for autonomy to exist. In Ubuntu, without a community the individual cannot exercise autonomy for an individual is connected to and is a person through others. This oneness and shared destiny incentivize an explicit duty to the individual and society to align their interests with each other (Rettovà 2016). Put it differently: relational autonomy begins with an ethical mandate originating from the shared state of interconnectedness.

The mainstream principle of autonomy as a rational endeavor to choose courses of actions maximizing one's benefit does not necessarily impose or imply ethical restrictions or positive duties to others or society. Non-relational autonomy requires the individual to maximize their interest without necessarily needing to consider or avoid social asymmetries in wealth, power, or other inequalities. Since the mainstream western philosophy that justifies and exports rationality-based autonomy allows for an unspecified degree of inequality to exist as a necessary condition of autonomy, one does not have a mandate to not worsen inequality. This type of autonomy may allow for high degrees of wealth and power inequality to exist without causing ethical concerns. Even if the Hippocratic principle of "do no harm" is implied, as in a number of AI ethics frameworks, it may only consider direct harms, harms that may be specified as crimes from direct commission of an offense or neglect of statute or duty. Indirect harms, second-order harms, such as lingering inequality stemming from colonization, a direct harm, that do not fit neatly within this myopic view are no longer considered the result of the principal actors or their beneficiaries.

This distancing of responsibility paired with the principle of autonomy void of an explicit ethical mandate is insufficient to address a wider range of harms that may affect the marginalized and compound inequality. When second order harms are no longer the responsibility of those who initiated them or the responsibility of society to address, as society has no special mandate to an individual other than to not interfere in the individual's autonomy, then it creates the idea that the individuals who are affected by such harms have either misused their autonomy, perhaps through non-performance of certain actions that the market or society may reward, errors in rational judgment between courses of action, or are autonomous beings solely responsible to resolve their current conditions. If any economy has non-equitable outcomes, outcomes that may feed into AI systems, it may be believed that individuals that have less favorable outcomes are solely responsible for their condition as the result of their autonomy. This non-relational principle of autonomy is not merely harmful but presents irreconcilable differences that undermine also the mainstream understanding and use of AI.

The mainstream principle of autonomy may be at odds with what AI is largely meant to be or accomplish, namely, to create and use machines that use previous experiences, encoded as digital datasets, to mathematically uncover patterns and correlations that would allow a machine to rationally and autonomously decide between courses of action. If machines are to respect our autonomy, then the autonomy of machines ought to be relational to our autonomy. This discredits the idea that rational decision making is a universal ethical virtue that leads to good outcomes, that autonomy is an ethically beneficial principle for its necessity in allowing one to make rational decisions and acknowledges that autonomy on its own does not necessarily lead to good outcomes. This would challenge an idea within AI that with enough data and computational resources any solvable problem can be solved without ethical considerations.

An AI system that enforces the principle of autonomy to itself without ethical positive duties to others may be harmful to society. If the principle of autonomy justifies inequality as a necessary cost for liberty, what degree of inequality must an autonomous AI system allow for? How much inequality can an AI system automate, perpetuate, or contribute to without causing direct harm? Does AI further the distance from the responsibility to minimize or not tolerate inequality? For example, if an AI system acting autonomously causes direct harm, who should be responsible for restitution and addressing the harm? If an AI system is to be a morally good actor within society, it must be explicitly designed according to a model that goes beyond that of rational autonomous agent, but is oriented and driven by positive duties towards society. Prescribing or encoding positive duties on an autonomous AI system such as ADM requires the system to behave in a relational manner as if it had relational autonomy. It follows that responsible AI can only occur in relational autonomy, and relational autonomy is a necessary step in decolonizing AI.

4.1 Decolonizing AI through Ubuntu's Relational Autonomy: A Call to Action

The ways to decolonize AI can be as varied as the experiences of colonization are varied. To some, colonization never ended, to others it ended only politically but not economically (Nkrumah 1966). Among the numerous manifestations of colonization some challenges were unique to particular regions. Therefore, to create a single framework that adequately addresses the effects of colonization would be infeasible. However, there are common traits and effects from the implementation across regions and time that can serve as a way of understanding current and future harms worsened by the use of AI (Mbembe and Dubois 2017). Mbembe argues that the "world is becoming black" stating that the colonial and post-colonial experiences of Africa are a site of one such framework for understanding present and future harms, that "blackness" serves as a prototype for future harms (Mbembe 2019).

Ubuntu's principles that were effective in fighting for equity, dignity, and autonomy in Africa's decolonial and post-colonial movements (Tutu 1999; Mandela 1994; Samkange and Tommie 1980) may also be useful in decolonizing AI. Ubuntu was used to highlight the failings of modernity and widespread inequality in wealth and power, exposing the broken social contract between and individual and society. Without the reciprocal duties society owed to

an individual and corresponding individual responsibility to society, autonomy could not exist. Without the shifting of AI and the socio-economic structures that allow for AI to be developed and deployed, from the antiquated rationalistic view to the relational view, the contract between society and the individual will continue to be broken.

With the increase of AI in everyday life, society is in a similar place where the struggle for equity, dignity, and autonomy is raging once more. AI is now sustained by structures and institutions that extract resources, surveil (Browne 2015), and control populations in a colonial like way (Couldry and Mejias 2019). Large tech companies have built massive platforms and internet infrastructure deployed globally especially in formerly colonized regions whose economies have yet to fully recover (Couldry and Mejias 2019). The false promise that rationality, modernity would alleviate human problems is the same false promise popularizing the use of AI (Mhlambi 2020).

Ubuntu's concept of relational autonomy allows us to re-imagine AI and the entire ecosystem that produces it. Relational autonomy allows us to envision recommender systems such as social media networks and e-commerce platforms, that have settings that users could control to change the content being recommended to them. AI designed to be relational, respecting the autonomy of humans while providing benefits to society, would allow individual users and the community to have more say and power over their online experiences. In this case the recommender system works as a collaborator of shaping online experiences.

Relational autonomy could also influence the social aspects of how AI is built. Relational autonomy as a principle of AI ethics frameworks would finally dispel the idea of an autonomous AI system that can use rationality to solve all sorts of problems. This removes a core belief that AI is neutral simply by using mathematics. Relational autonomy would make definitively clear that no action is truly neutral but has consequences across a system (Noble 2018) and could encourage companies and teams to seek more input and feedback during the design choices of AI. Relational autonomy would also help to overcome trade-off situations between ethical principles, and/or solve moral dilemmas (Dignum 2022), as it embeds duties of social justice, communal responsibility, and accountability (and redress). Ubuntu's relational autonomy indeed also allows for restorative justice and for reparations to address direct and indirect harms that are perpetuated using AI. This would overcome current contentions on whom to bestow liability when individuals are harmed by AI. This could ethically justify laws such as The Digital Services Act, which also makes possible the idea of restitution when users are harmed by recommended contents.

5 Conclusions

In this paper we have argued that the mainstream ethical principle of autonomy is insufficient to use as a rail against the harms resulting from the use of AI. This ethical principle without having a relational positive duty to others fails to recognize the impact of legacy harms that result from colonization and the present social hierarchies and inequalities that stem from it. To counter such harms, we posit relational AI and specifically relational AI ethics via the introduction of relational autonomy as necessary and missing step in adequately addressing persistent harms perpetuated by AI especially to the most marginalized and vulnerable. By doing so, we have drawn the conceptual path to revise the ethical concept of autonomy informing many popular AI ethics frameworks in a way that considers both non-mainstream Western moral philosophy, such as accounts of relational autonomy that are currently at the outskirt of the debate in AI ethics, as well as a non-Western and essentially relational philosophy, as the African account of autonomy developed in Ubuntu ethics, and showed how both perspectives are necessary and meaningful to revise the AI ethics principle of autonomy in a more adequate and inclusive way.

Data Availability Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors have no competing interests to declare that are relevant to the content of this article. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no financial or proprietary interests in any material discussed in this article.

References

Anderson J, Honneth A (2005) Autonomy, vulnerability, recognition, and justice. In: Christman J, Anderson A (eds) Autonomy and the challenges to liberalism: new essays. Cambridge University Press, New York, pp 127–149

Angwin J, Larson J, Mattu S, Lauren K (2016) Machine bias. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Applin SA, Fischer MD (2015) New technologies and mixed-use convergence: How humans and algorithms are adapting to each other. In: 2015 IEEE international symposium on technology and society (ISTAS). IEEE: Dublin, Ireland: pp. 1–6. https://doi.org/10.1109/ISTAS.2015.7439436

Asante ML, Abarry AS (1996) African intellectual heritage: a book of sources. Temple University Press, Philadelphia

Beauchamp TL, Childress JF (2013) Principles of biomedical ethics, 8th edn. Oxford University Press, New York

Bell D (1993) Communitarianism and its critics. Clarendon, Oxford

Benjamin R (2019) Race after technology: abolitionist tools for the new Jim code. Polity, Cambridge

Benson P (1990) Feminist second thoughts about free agency. Hypatia 5(3):47–64

Benson P (2005) Feminist intuitions and the normative substance of autonomy. In: Taylor JS (ed) Personal autonomy. Cambridge University Press, New York, pp 124–142

Birhane A (2021) Algorithmic injustice: a relational ethics approach. Patterns. https://doi.org/10.1016/j.patter.2021.100205

Browne S (2015) Dark matters: on the surveillance of blackness. Duke University Press, Durham

Calvo RA, Peters D, Vold K, Ryan RM (2020) Supporting human autonomy in ai systems: a framework for ethical enquiry. In: Burr C, Floridi L (eds) Ethics of digital well-being: philosophical studies series. Springer, Cham, p 140

Christman J, Anderson J (2005) Autonomy and the challenges to liberalism: new essays. Cambridge University Press, New York

Colburn, (2010) Autonomy and liberalism. Routledge, London

Couldry N, Mejias U (2019) The costs of connection: how data colonizes human life and appropriates it for capitalism. Stanford University Press, Stanford

Daniels N (1974) On liberty and inequality in Rawls. Soc Theory Pract 3(2):149–159

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

Dignum V (2022) Relational artificial intelligence. https://doi.org/10.48550/arXiv.2202.07446

Dworkin G (1988) The theory and practice of autonomy. Cambridge University Press, New York

Dworkin R (2000) Sovereign virtue: the theory and practice of equality. Harvard University Press, Cambridge

Ekstrom L (1993) A coherence theory of autonomy. Philos Phenomenol Res 53:599–616. https://doi.org/10.2307/2108082

European Parliament (2017) Report with recommendations to the Commission on Civil Law Rules on Robotics.

Eze EC (1997) The color of reason: the idea of "Race" in Kant's anthropology. In: Eze EC (ed) Postcolonial African philosophy: a critical reader. Blackwell Publishers, Hoboken

Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar, M (2020) Principled Artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication, 2020–1. http://dx.doi.org/https://doi.org/10.2139/ssrn.3518482

Floridi L (2011) The informational nature of personal identity. Minds Machines 21:549–566. https://doi.org/10.1007/s11023-011-9259-6

Floridi L, Cowls J (2019) A unified framework of five principles for AI in society. Harvard Data Sci Rev 1:1. https://doi.org/10.1162/99608f92.8cd550d1

Floridi L et al (2018) AI4People—an ethical framework for a good AI society: opportunities, risk, principles, and recommendations. Minds Machines 28:689–707. https://doi.org/10.1007/s11023-018-9482-5

Frankfurt H (1989) Freedom of the will and the concept of the person. In: Christman J (ed) The inner citadel: essays on individual autonomy. Oxford University Press, New York, pp 63–76

Fricker M (2007) Epistemic injustice: power and the ethics of knowing. Oxford University Press, New York

Giovanola B, Sala R (2021) The reasons of the unreasonable: is political liberalism still an option? Philos Soc Crit. https://doi.org/10.1177/01914537211040568

Giovanola B, Tiribelli S (2022a) Weapons of Moral construction? On the value of fairness in algorithmic decision-making. Ethics Inform Technol. https://doi.org/10.1007/s10676-022-09622-5

Giovanola B, Tiribelli S (2022b) Beyond bias and discrimination: redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. AI Soc. https://doi.org/10.1007/s00146-022-01455-6

Google AI (2018) Our principles. https://ai.google/principles/.

Grovier T (1993) Self-trust, autonomy and self-esteem. Hypatia 8(1):99–119. https://doi.org/10.1111/j.1527-2001.1993.tb00630.x

Gutman A (1985) Communitarian critics of liberalism. Philos Public Aff 14(3):308–322

Helberger, N (2016) Profiling and targeting consumers in the internet of things—a new challenge for consumer law. https://doi.org/10.2139/ssrn.2728717

High Level Expert Group on Artificial Intelligence [HLEGAI] (2019). Ethics guidelines for trustworthy AI. European Commission. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

IEEE (2017) The IEEE global initiative on ethics of autonomous and intelligent systems. IEEE Standards Association. https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

IEEE (2019) Global initiative on ethics of autonomous and intelligent systems. Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent systems, 1st ed. (EAD1e). https://standards.ieee.org/wp-content/uploads/import/documents/other/ead1e.pdf

Jaggar A (1983) Feminist politics and human nature. Rowman and Allanheld, New Jersey

Jaworska A (2009) Caring, minimal autonomy, and the limits of liberalism. In: Lindemann H, Verkerk M, Walker M (eds) Naturalized bioethics: toward responsible knowing and practice. Cambridge University Press, Cambridge

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

Jonjepier F, Klenk M (2022) The philosophy of online manipulation. Routledge, London

Kahneman D (2011) Thinking fast and slow. Straus & Giroux, New York

Killmister S (2017) Taking the measure of autonomy: a four-dimensional theory of self-governance. Routledge, New York

Korsgaard CM (1996) The sources of normativity. Cambridge University Press, New York

Korsgaard CM (2014) The normative constitution of agency. In: Vargas M, Yaffe G (eds) Rational and social agency: the philosophy

of Michael Bratman. Oxford University Press, New York, pp 190–214
Kramer ADI, Guillory JE, Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci USA 111:8788–8790. https://doi.org/10.1073/pnas.1320040111
Kunene M (1981) Anthem of the decades: a Zulu epic. Heinemann, London
Kunene M (1982) The ancestors & the sacred mountain: poems. Heinemann, London
Kymlicka W (1989) Liberalism, community and culture. Clarendon, Oxford
Lee JS, Hord FL (2016) I am because we are: readings in Africana philosophy. University of Massachusetts Press, Amherst
MacIntyre (1988) Whose justice? Which rationality? University of Notre Dame Press, Notre Dame
Mackenzie C (2014) Three dimensions of autonomy: a relational analysis. In: Veltman A, Piper M (eds) Autonomy, oppression, and gender. Oxford University Press, Oxford, pp 15–41
Mackenzie C (2015) Responding to the agency dilemma: autonomy, adaptive preferences, and internalized oppression. In: Oshana M (ed) Personal autonomy and social oppression. Routledge, New York, pp 48–67
Mackenzie C, Stoljar N (2000a) Relational autonomy: feminist perspectives on autonomy, agency, and the social self. Oxford University Press, New York
Mackenzie C, Stoljar N (2000b) Introduction: autonomy refigured. In: MacKenzie C, Stoljar N (eds) Relational autonomy: feminist perspectives on autonomy, agency, and the social self. Oxford University Press, New York, pp 3–31
Mandela N (1994) Long walk to freedom: the autobiography of Nelson Mandela, 1st edn. Times Warner book, London
Mbembe A (2019) Necropolis. Duke University Press, Durham
Mbembe A, Dubois L (2017) Critique of black reason. Duke University Press, Durham
Mcdowell C, Chinchilla MY (2016) Partnering with communities and institutions. In: Gordon E, Mihailidis M (eds) Civic media: technology, design, practice. MIT Press, pp 461–480
McLeod C, Sherwin S (2000) Relational autonomy, self-trust, and health care for patients who are oppressed. In: MacKenzie C, Stoljar N (eds) Relational autonomy: feminist perspectives on autonomy, agency, and the social self. Oxford University Press, pp 259–279
Menkiti IA (1984) Person and community in African traditional thought. In: Wright R (ed) African philosophy: an introduction. University Press of America, Lanham, pp 171–182
Mhlambi S (2020) From rationality to relationality: ubuntu as an ethical and human rights framework for artificial intelligence governance. Harvard carr center discussion paper series. https://carrcenter.hks.harvard.edu/publications/rationality-relationality-ubuntu-ethical-and-human-rights-framework-artificial
Michael B (2005) Planning agency, autonomous agency. In: Taylor JS (ed) Personal autonomy. Cambridge University Press, New York
Mignolo WD, Escobar A (2010) Globalization and the decolonial option. Routledge, London
Milano S, Taddeo M, Floridi L (2020) Recommender systems and their ethical challenges. AI & Soc 35:957–967. https://doi.org/10.1007/s00146-020-00950-y
Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L (2016) The ethics of algorithms: mapping the debate. Big Data Soc. https://doi.org/10.1177/2053951716679679
Mohamed S, Png MT, Isaac W (2020) Decolonial AI: decolonial theory as sociotechnical foresight in artificial intelligence. Philos Technol 33:659–684. https://doi.org/10.1007/s13347-020-00405-8
Molema L (1917) The Bantu past and present, an ethnographical & historical study of the native races of South Africa (2012[1917]). Forgotten Books, London
Natale A (2021) Deceitful media: artificial intelligence and social life after the turing test. Oxford University Press, Oxford
Ndlovu-Gatsheni SJ (2019) Provisional notes on decolonizing research methodology and undoing its dirty history. J Dev Soc 35(4):481–492. https://doi.org/10.1177/0169796X19880417
Newell S, Marabelli M (2015) Strategic opportunities (and challenges) of algorithmic decision-making: a call for action on the long-term societal effects of 'datification.' J Strateg Inf Syst 24(1):3–14. https://doi.org/10.2139/ssrn.2644093
Ngubane J (1979) Conflict of minds. Books in Focus, New York
Nkrumah (1966) Neo-colonialism: the last stage of imperialism. International Publishers, New York
Noble, (2018) Algorithms of oppression: how search engines reinforce racism. New York University Press, New York
Nyathi (2001) Traditional ceremonies of Amandebele. Mambo Press, Gweru
O'Neil C (2016) Weapons of math destruction: how big data increases inequality and threatens democracy. Crown, New York
Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. Science 366:447–453. https://doi.org/10.1126/science.aax2342
Organization for Economic Co-operation and Development (OECD) (2019). Recommendation of the Council on Artificial Intelligence (Technical Report OECD/LEGAL/0449). https://oecd.ai/en/ai-principles
Oshana M (1998) Personal autonomy and society. J Soc Philos 29(1):81–102. https://doi.org/10.1111/j.1467-9833.1998.tb00098
Oshana M (2006) Personal autonomy in society. Ashgate, Hampshire
Pariser E (2011) The filter bubble. Penguin, London
Peña P, Varon, J (2019) Decolonizing AI: a transfeminist approach to data and social justice" [Global Information Society Watch 2019]. Association for Progressive Communications. https://www.giswatch.org/node/6203
Pittman RC (1960) Equality versus liberty: the eternal conflict. Am Bar Assoc J 46(8):873–880
Prunkl C (2022) Human autonomy in the age of artificial intelligence. Nat Mach Intell 4:99–101. https://doi.org/10.1038/s42256-022-00449-9
Ramose BM (1999) African philosophy through Ubuntu. Mond Books Harare, Zimbabwe
Raz J (1986) The morality of freedom. Clarendon Press, Oxford
Rettovà A (2016) African philosophy as a radical critique. J Afr Cult Stud 28(2):127–131. https://doi.org/10.1080/13696815.2016.1159123
Ricaurte P (2019) Data epistemologies, the coloniality of power, and resistance. Television New Media 20(4):350–365. https://doi.org/10.1177/1527476419831640
Roche C, Wall PJ, Lewis D (2022) Ethics and diversity in artificial intelligence policies, strategies and initiatives. AI Ethics. https://doi.org/10.1007/s43681-022-00218-9
Roessler B (2021) Autonomy: an essay on the life well-lived. John Wiley, New Jersey
Rosenberg M (2018) Bolton was early beneficiary of Cambridge Analytica's Facebook data. New York Times. https://www.nytimes.com/2018/03/23/us/politics/bolton-cambridge-analyticas-facebook-data.html
Royakkers L, Timmer J, Kool L, van Est R (2018) Societal and ethical issues of digitization. Ethics Inf Technol 20(2):127–142. https://doi.org/10.1007/s10676-018-9452-x
Samkange S, Tommie M (1980) Hunhuism or ubuntuism: a Zimbabwe indigenous political philosophy. Graham Pub, Salisbury

Sandel M (1982) Liberalism and the limits of justice. Cambridge University Press, Cambridge
Simon H (1991) Bounded rationality and organizational learning. Organ Sci 2(1):125–134. https://doi.org/10.1287/orsc.2.1.125
Simonite T (2020) Meet the secret algorithm that's keeping students out of college. Wired. https://www.wired.com/story/algorithm-set-students-grades-altered-futures/.
Smith EHJ (2019) A history of the dark side of reason. Princeton University Press, Princeton
Sunstein C (2008) Democracy and the internet. In: van den Hoven J, Weckert J (eds) Information technology and moral philosophy. Cambridge University Press, New York, pp 93–110
Susser D, Roessler B, Nissenbaum H (2019). Technology, autonomy, and manipulation. Internet Policy Review, 8 (2). https://doi.org/10.14763/2019.2.1410
Taylor C (1979) Hegel and the modern society. Cambridge University Press, Cambridge
Taylor C (1991) The ethics of authenticity. Harvard University Press, Cambridge
Taylor JS (2009) Practical autonomy and bioethics. Routledge, London
Tene O, Polonetsky J (2013) Big data for all: privacy and user control in the age of analytics. Nw. J. Tech. Intell. Prop 11:239
Thaler R, Sunstein C (2009) Nudge: improving decisions about health, wealth and happiness. Penguin, London
Thomas P (2017) Self-determination: the ethics of action. Oxford University Press, Oxford
Tiribelli S (2020) Predeterminazione algoritmica e libertà di scelta. In: Alici L, Miano F (eds) Etica nel Futuro. Orthotes, Napoli, pp 431–441
Tiribelli S (2023) Moral freedom in the age of artificial intelligence. Mimesis International, Milan, London
Tsamados A, Aggarwal N, Cowls J, Morley J, Roberts H, Taddeo M, Floridi L (2022) The ethics of algorithms: key problems and solutions. AI Soc. https://doi.org/10.1007/s00146-021-01154-8
Tutu D (1999) No future without forgiveness. Doubleday, New York
van den Hoven J, Rooksby E (2008) Distributive justice and the value of information: a (broadly) Rawlsian approach. In: van den Hoven J, Weckert J (eds) Information technology and moral philosophy. Cambridge University Press, Cambridge, pp 376–396
Veltman A, Piper M (2014) Autonomy, oppression, and gender. Oxford University Press, Oxford
Wachter S (2020) Affinity profiling and discrimination by association in online behavioural advertising. Berkeley Technol Law J. https://doi.org/10.2139/ssrn.3388639
Westlund A (2009) Rethinking relational autonomy. Hypatia 24:26–49. https://doi.org/10.1093/phe/phu025
World Health Organization (2021) Ethics and Governance of Artificial Intelligence for Health. https://www.who.int/publications/i/item/9789240029200
Zarsky T (2016) The trouble with algorithmic decisions an analytic road map to examine efficiency and fairness in automated and opaque decision making. Sci Technol Human Values 41(1):118–132. https://doi.org/10.1177/0162243915605575
Zuboff S (2019) The age of surveillance capitalism: the fight for a human future and the new frontier of power. Public Affairs, New York