

Analytic Relationality and the Relational Ethics of the Global South:

Making the Case for Abeba Birhane's Work

Julia Kwak

Stanford University

and

Dr. Nakeema Stefflbauer

Frauenloop, Berlin; Stanford University

In this paper we consider how the work of Abeba Birhane relates to other theories of relationality from European and North American analytical philosophy, and ask whether, given Birhane's critical perspective on western rationalism, analytic frameworks might be at all compatible. Analytic descriptions of relational ethics draw on Watson, Smith, Scanlon, Darwall, and Bovens' theories of relationality and "being held responsible," when an actor contravenes the norms of a relationship with others. Armed with these frameworks, analytical philosophers hope to critically evaluate and—eventually—regulate the global political economies of data and computing industries. Yet, critics like Birhane argue that such frameworks remain mired in the colonial project of western rationality, which is complicit in the digital colonization of the Global South. Even if these analytical relational frameworks address fora which debate accountability about the "many hands" responsible for algorithms, they refer to individual actors and responsibility within western corporate and institutional structures. Moreover, they presume an equal moral status for all actors, which in reality is often not the case since western technology disproportionately harms communities in the Global South. Meanwhile, relational ethics as

Birhane formulates them offer an alternative view that arises from communities. These two philosophical approaches to relationality, while often at odds with one another, do share some conceptual histories and potential compatibilities. We argue the analytic enterprise can ground policy work with clear definitions of contested terms like “harm,” “understanding,” and “responsibility.” Birhane’s relational ethics also draw on both western rationality and concepts of lived experience in these communities. The synthesis of these two types of relationality may help develop an inclusive, actionable, enforceable AI ethics. Still, it is important to remember that in the Global South the relational focus is communities and their well-being, something analytical frameworks aspire to but have yet to adequately address.

I. Analytical frameworks, responsibility, and relationality

Analytic relationality originates outside of analytical philosophy, which rarely elaborates on external influences, especially not from twentieth century existentialist philosophers (Fullbrook & Fullbrook, 1999). Yet, Simone de Beauvoir’s *Ethics of Ambiguity* (1948) offers a strong precedent for later analytic discussions of relationality and intersubjectivity, which draw on her account for definitions the “second-personal perspective” (Darwall 2006; Rini 2020). For Beauvoir, humans are “ambiguous” beings, who simultaneously imagine themselves as “a sovereign and unique subject amidst a universe of objects” while they are also objects for others. At once a subject to oneself and an object to others, a person nevertheless remains an “individual in the collectivity on which he depends (7).” Unlike Heidegger’s *Fürsorge*, caring-for someone and being attentive to their needs (Heidegger 2010) Beauvoir’s “ambiguity” refers to a self that must acknowledge its doubled status. Yet similar to Heidegger’s sense of self, Beauvoir’s retains

its individuality, and therein lies the difference to relationality in the Global South, where relationality is neither merely ambiguity nor dependency, and individuality is no primary assumption one has to learn to accept as doubled. Rather, in the Global South, relationality is inseparable from the self. From Beauvoir to Darwall, who draws on Thompson's conception of the "bipolar obligation" (Darwall 2012; Thompson 2003), this indelible individuality in the Global North grounds all relations and discussions of relationality differently than the Global South.

With its clear definitions of blameworthiness and responsibility, analytic relationality appears promising for efforts to develop ethical policies in the complex relations between actors in the Software Development Life Cycle (SDLC) and the affected communities. Many efforts to articulate responsibility for algorithmic harm draw on Gary Watson, Angela M. Smith, Darwall, and T.M. Scanlon. Watson offers a two-pronged concept "responsibility as attributability" and "responsibility as accountability," wherein both senses of responsibility refer to relationships among actors: "Holding people responsible is not just a matter of the relation of an individual to her behavior; it also involves a social setting in which we demand (require) certain conduct from one another and respond adversely to another's failures to comply with these demands," (Watson, 1996). Responding to Watson and influenced by T.M. Scanlon, Angela M. Smith proposes a single, unified concept of moral responsibility, "responsibility as answerability," which endeavors to describe what the actor owes others (2015). Scanlon's contractualism involves a "mutual recognition" between actors of equal moral status, who decide rightness and wrongness of acts, as well as responsibility and communal impact. According to contractualist theories in ethics, whether an action is acceptable is determined by whether it could be justified to others on grounds no one could reasonably reject (Scanlon 1998). Reasonable rejectability of

principles refers only to individual personal objections. In deciding whether some action is morally objectionable, contractualists must either decide based on the probability of an outcome in advance, *ex ante*, or on the basis of the outcomes of the principles *ex post*. In the case of AI ethics, most contractualists lean toward *ex ante* or a hybrid of both because it is better to consider the probability of harm in design, while also acknowledging the ongoing impacts.

Because Watson, Smith, Darwall, and Scanlon refer to individual perspectives even when they address relational norms, it becomes important to find a framework that promotes public and institutional conversation about blame, whether in the *ex ante* case— you knew this algorithm might harm— or in the *ex post*, you see it has harmed. Thus, analytic philosophy frameworks also adopt political philosopher Mark Bovens' accountability theory to consider relationality as a series of interdependent relations between actors and the social, legal, and political structures that legitimize actors' decisions. For Bovens, accountability weighs the decisions of the actor, the response of a forum or many fora, like governments, public discussions, or auditing groups, which hold the actor accountable. It also considers the relationship between actors and these publics, the content of the account, and the consequences of the account (Bovens, 2014). Actors who design algorithms may be blameworthy for their harmful social impact, and those who've been affected are entitled to respond as the forum. This moral philosophy is relational in the sense that the actor owes the forum an accounting for algorithmic harm and the forum deserves restitution.

Algorithmic accountability theory that builds on Bovens hopes to describe relationality in terms of institutions and potential policy-making. For example, Maranke Wieringa's systematic literature review "What to account for when accounting for algorithms" (2020) examines relationality at every stage of the SDLC and the relevant actors, whose involvement may forecast

impacts *ex ante*, that is, consider design and implementation, or *in medias res* as the account progresses, or *ex post* from a retrospective position. In all stages of algorithmic development, Wieringa describes algorithmic accountability as relationality: it is a relationship with multiple decision makers, developers, users, each of whom “have the obligation to explain and justify their use, design, and/or decisions of/concerning the system and the subsequent effects of that conduct,” (10). This empirical, procedural application of relationality aims at design justice *ex ante* or blameworthiness *post ex*. Other efforts that draw on Bovens include Busuioc (2021), Meijer and Grimmelikhuijsen (2020), and Cooper et al. (2022), which revisits Nissenbaum (1996) and combines moral philosophy with Boven’s accountability theory.

Even when analytical relationality tends toward procedural recommendation, like Wieringa, these frameworks remain less interested in developing “end-to-end” auditing protocols (Raji et al. 2020) than in conceiving of a relationality where norms extend beyond the individual and refer to a network of people responsible for algorithmic development. Whether these frameworks actually become implementable and address those communities depends on their enforceability, and that remains currently still more aspiration than an implementation.

Largely stalled at this enforceability challenge, analytic relational ethics teeters between *ex ante* and *ex post* discussions as it endeavors to eventually affect a real-world impact. Even with policy-minded relational discussions in analytic philosophy, ethics remain better prepared to challenge definitions of algorithmic metrics like “risk” than actual structural and social relations. This relationality seeks to answer questions of social impact in an incremental “principled manner, rather than a mandatory checklist” (Green, 2022). No variety of ethics, in the Global North or South, would affirm a corporate-issued checklist designed to placate critics, but the “incremental” and “principled” thinking of western philosophy often fails to directly address the

urgencies of communal needs. It's no surprise that some Black AI ethicists like Brandeis Marshall assert "ethics of any kind is of limited utility" (Marshall, 2022, p. 27). For critics like Marshall who claim that moral philosophy fails to "show up" for marginalized groups, analytic philosophy's move to relationality hardly ameliorates western philosophy's traditional white elitist ethos.

Considering analytic philosophy's presumption of egalitarian conditions for all actors we ask: how constructive and competent are theories which lie in settings that at best evade, or at worst, diminish the differently positioned members in human society? A fundamental lack of these theories is not only inapplicability, but their incomprehensive, therefore, incomplete view of the human moral capacity. Are people of moral equal status? If so intrinsically, do they receive treatment as such? Who constructs these "agreed upon" standards of morality? What epistemic resources are utilized in the construction of these standards? What are the mechanisms of power within relations of moral actors? How is one enabled or inhibited in their actualization of moral relations? These questions are critical for a theory that strives to understand human moral relations.

To understand the limitations of analytical relationality, it is important to consider how this unequal moral status affects people. The feminist epistemological framework of the "situated knower" can aid in our aims to ground the positionality of a moral actor. Situated knowledge relies on the concept of "social location." Social location refers to a person's identity markers and social roles, along with their accompanied behaviors, beliefs, and emotions. Each individual in society has a social location that comes with specific epistemic norms that they have been conditioned to embody. What one believes, how one emotes, how one acts, how one constructs their self-hood are all influenced by their situated knowledge within their social location

(Anderson, 2011). This idea of situatedness extends to one's exercise of morality. It is not the case that all people are treated as if they have equal stakes in moral standards. It is also not the case that all people are enabled to act according to their ideal wishes of morality. And certainly, it is not the case that all people have been offered the epistemic resources to cultivate this particular form of discernment of morality. The moral capacity of the woman was heavily undermined. The moral capacity of African Americans under slavery was maliciously distorted. The moral capacity of the child is paternalized. The western feminist tradition unveils the different positions that types of humans, in particular, women, are subject to within relations. They elaborate on the dynamics of power that govern relations, and how they put women, their minds, their bodies, and their experience under standards of morality in distinctly different standpoints in contrast to white, heterosexual, and cisgender men.

Feminist philosopher Miranda Fricker's (2007) theory of hermeneutical injustice provides insight into how one's situatedness can dictate moral exercise and better help us understand which kinds of relationality might best promote inclusion. Hermeneutical injustice is founded on the premise that there are epistemic resources that aid in understanding social experiences. Fricker coins these *hermeneutical resources*. Some examples include concepts, expectations, philosophies, theories, procedures, standards of evaluation, organizational schemata, and language. Essentially, hermeneutical resources are tools that facilitate our understanding of our social world. Fricker further establishes that a society has a common set of hermeneutical resources that most, if not all members make use of. The resources that society as a collective accepts and ubiquitously employs (we will call these *collective* hermeneutical resources), are made only by a subset of the population; namely those in socially powerful locations. Collective hermeneutical resources are constructed in the professional settings of

politics, academia, law, and journalism—spheres that those in socially powerless locations are systematically excluded from. *Hermeneutical marginalization* is this very structural exclusion of the socially less powerful in the construction of collective hermeneutical resources. Even without mal intent (which most certainly exists among the socially powerful), it is natural that one creates hermeneutical resources that aid in understanding social experience in ways that are in their own interests. Reflecting these disparities, collective resources created by the socially powerful *insufficiently* accommodate for interests of the socially powerless, given that their power exists only in relativity to those without power.

When we think of one's ethical discernment as being cultivated by hermeneutical resources, we can see how it subjects vast groups of less privileged peoples under hermeneutical marginalization. Unlike the assumption of humans' ability to operate as if they have equal moral status, hermeneutical injustice displays how certain groups of moral actors directly do not get to participate in the moral construction of their community, and furthermore, is deprived of a moral framework that accommodates their particular interests. Moreover, the relations between constructors of hermeneutical resources and their hermeneutically marginalized presents a stark power asymmetry that prevents the full embodiment of the latter group's ideals of morality.

While Fricker describes how some people are not able to effectively participate in the moral responsibilities outlined by the predominantly white, male analytic tradition, an ethics of care describes, positively, another form of morality that humans can exercise. An ethics of care shows a relationship that is not bound by responsibility or duty, but by love and natural inclination to care. The relation is built of the carer and the cared-for, where the former is "engrossed," in the interests of the latter. They experience a "displacement [of their own] motivation," where their pursuits are directed away from their personal projects and towards the

cared-for. Instead of a contractual reciprocity that is borne out of a sense of obligation or responsibility, care ethics does not demand identical or “morally equal” behaviors from the two parties in a relationship. Rather, it promotes a “response” from the cared-for where they acknowledge the care they receive from the carer.

A very different western relational ethics claims to do better. Patricia Hill Collins’ *Black Feminist Thought* (1986; 2002) and the ethics of care, Gilligan (1982; 1993), Noddings (1984; 2013), Harding (1987), Ruddick (1989), Putnam (1993), Smith (1993) present relationality as intersubjective bonds rather than questions of predicted harm or blameworthiness. Feminist accounts from Collins to Harding understand relational ethics as inquiry into community care and sovereignty over women’s minds, bodies, and increasingly, their data. They point their attention to the positionality of women within relations, and most of all, the privilege of those who shaped relations. Smith (1993) asserts: “If the moral life is to be understood relationally, as many feminists have argued, then feminists perpetually and critically have to ask how the power relations of our moral claims order differences.”

However, even with the establishment of these feminist theories, actual inclusion, especially along identity lines apart from gender, remains delayed. Only after undergoing a confrontation with their own privilege, did these late 20th century white feminists read Collins and include race in their inquiries about power relations and the privileges of class and gender. Nel Noddings mentions race in her book, but only once as a category among others like “class” (Noddings, 2013, p.184). Unlike her contemporary white feminists, Sandra Harding was one of the first western, white feminists to write about the commonalities between feminist and African relational ethics and move the discussion away from Eurocentrism (1987). Yet, her western

feminist conceptions of care, community, and harmony reveal important differences to African relationality.

II. Relationality in the Global South is positive rather than about blame, but it is not completely free of western thought, since some versions like Abeba Birhane's draws on Black Feminism.

Whereas analytical AI accountability theory ultimately centers around the blameworthiness of the individual or individuals, and when it claims to address structural injustice appears most successful in challenging definitions of terms, relational ethics in the Global South refers to positive communal care and development that focuses *ex ante* on design justice and securing sovereignty over data for data subjects of marginalized groups, which AI disproportionately harms. This relationality refers to communities who co-create their identity for whom ethics is intersubjective. Nigerian-British philosopher Nkiruka Ahiauzu describes African relationality:

The relationship is not one based on a conception of fair play as we see with the fairness theory but on a sense of concern for others. It is not borne out of an idea of implicit consent or ahistorical actual consent but on the humanity of persons and the implication of that humanity for normativity. (Ahiauzu, 2006)

Arising from a sub-Saharan oral tradition and building on the central ideal of human excellence, or “*Ubtunu*” in southern African culture, these ethics derive from an understanding of the self in relation to its community. People live “a life of mutual concern for the welfare of others, such as in a cooperative creation ... feeling integrated with as well as willing to integrate others into a web of relations” (Masolo 2010, p. 240).

Ubuntu is not an easily defined concept. It is a kind of philosophy but also a social practice. As a philosophy, it captures areas of epistemology, ontology, and ethics; and importantly, it is not easily reducible to these categories of western philosophical frameworks. At its most fundamental form, *Ubuntu* is a relational philosophy that humans are intrinsically interconnected in a web of relations with others. Branches of this relational theory argue that one's very own humaneness and humanness is derived from others. A baby is not a human yet, until it interacts with and consumes others' influence. In this way, a human is ontologically relational, and epistemically derives their sense of self and the world in which they inhabit through these relations. Because of the implication of one's humaneness/humanness, *Ubuntu* is also an ethical examination of the human condition, and how we ought to act in our relations. The interconnectedness of humanity that African relational ethics purports, rejects the notion that a moral actor can be an independent entity that acts upon other people. Moreover, the emphasis on interconnectedness also situates *Ubuntu* as a theory embedded in context, as opposed to one within objective distance (Gouws & van Zyl 2015, Cornell & Marle 2015). Critiques of *Ubuntu* problematize how the philosophy has been adapted in African societies to justify heteronormative, patriarchal, and nationalist practices. It is critical to evaluate *Ubuntu* in the political landscape of South Africa, which complicates the ethicality of its uses. *Ubuntu* gained prominence in the 1980s when critical southern African countries began or were transitioning to majority-Black self-rule. In this transitory stage of thwarting histories of colonization, *Ubuntu* proved to be a political device to harken back to a time before white colonialism, and to restore the solidarity of a heavily divided post-colonial society (Gade, 2011). As a result, pre-colonial practices that hinge on patriarchy and heteronormativity were construed aspirationally, and impacted the perception of an *Ubuntu* conception. It is this context of colonization that critiques

of *Ubuntu* must contextualize. Without enabling problematic practices, one must still acknowledge that these ideas of feminism originate in western, colonizer societies; a resistance to adopting these feminist theories may not be completely permissible, but at the very least, understandable. However, there exists Afro-feminist thinkers today, as well thinkers handling issues of algorithmic justice that combine western feminism with *Ubuntu* to construct an ethical path forward.

Abeba Birhane (2021) criticizes biases of western philosophy contributors, whom she sees as perpetuating rather than challenging algorithmic injustice. Rejecting western individualist rationality as the dominant view as a definition for personhood, Birhane describes an African relational ethics, which links one's personhood to the personhood of others. These relational ethics elaborate why western rationality and its metrics may be inappropriate for assessing harms in non-western communities. Describing *Ubuntu* as having a "relational personhood diametrically opposed to rationality as personhood," she argues that such a philosophy can counteract the Western world's asymmetric relationship with much of the world, computing culture, and tech corps' quixotic quest for general AI understanding and mechanical personhood. Birhane is also careful to be inclusive in her description of African philosophy, because for her, not all relational frameworks are non-western, nor southern African. Birhane's adoption of western frameworks into her AI Ethics, especially Black feminist thought, shows her developing a hybrid theory of western and non-western relationality much like Sandra Harding. Birhane's drawing on these diverse sources thus also elides some of the complexities of African relational ethics, which confront traditional African social and patriarchal hierarchies as well, and which African ethics are currently contesting in their own re-fashioning of relational ethics (Jecker, 2022).

Birhane calls for more contextual understanding of the power relations and historical experiences that define the data settings in which human data is collected. If this is a more labor-intensive approach to identifying bias than mathematical retro-fitting, it is also a practice that centers lived experience of harm rather than hypothetical scenarios. Birhane notes that, unfortunately, there are no easy solutions to the problem of “AI” or machine learning tools being used as “the hammer every messy social challenge is bashed with.” However, there is a growing body of well-documented instances where marginalized individuals’ experiences of algorithmic systems have been harmful. How to prioritize these individuals’ protection from further harm - as well as any individual who is more likely to receive unfair assessment relative to others - is where Birhane’s ethical practice differs from other approaches. Her focus is on *ex-ante* ethical practice as a human-centered design process where biased outcomes and harms are mitigated via contextual analysis. Departing from overly generalized “fairness” definitions, Birhane asserts that people’s relations to others within hierarchical societies demands continual assessment and incorporation into an accurate account of fairness for any real-world group of people.

Luckily, the concept of ensuring that digital barriers are minimized or removed from different communities is not new: accessibility standards have been widely adopted to accommodate a range of groups. The reason is that these groups would either struggle with or face exclusion from digital properties without accessible designs —e.g., have sub-optimal experiences. Is there a less optimal experience of algorithmic systems than to be unfairly denied a loan, denied a job, denied medical treatment, denied due legal process or denied social assistance? If people building digital experiences now have an obligation to proactively implement accessibility design for varied populations, it is also possible to build algorithmic systems to deliver proactive, community-focused bias-mitigation design for those who

consistently experience algorithmic harm. One example is the ally framework, which gives a starting point for acknowledging and incorporating people's real-world circumstances and needs into the built digital experiences promoted globally. Dr. Birhane has defined a relational ethics process that draws on Global South philosophies and marginalized communities' experiences. Operationalizing these practices involves little more than re-framing the people most at risk of algorithmic harm as the audience to be accommodated, rather than excluded from consideration.

In this sense, Birhane's relational ethics have greater potential to be actionable than the analytic ones, but there is a place for both frameworks. African and non-western relationality has a similar capacity to consider the *ex ante* and *ex post* questions of algorithmic justice, but being communally focused may well be better positioned to describe how communities in the Global South can demand that if western technologies and corporations want global data, they must guarantee sovereignty over data for data subjects in the Global South. There is ample evidence that analytic definitions do help with policy-making, and could also be of use beyond the Global North (Cairney, 2016). Thus, both types of relational frameworks can challenge tech to transform large models from their currently nascent shaky foundations built on unconsented data into more reliable infrastructures, but only if large models think smaller. Currently, large models exhibit what Birhane describes as a western hubris claiming to offer generalized knowledge, where one algorithm purports to "understand" all domains of inquiry and aspires to "reflect human values." Both analytical and Birhane's relational frameworks cast a critical eye on such often overstated and vague goals, foregrounding the volatility of large models, institutional privilege of those who build them and their global decision-making power. The best use of relationality is to center the communities most disproportionately affected by algorithms, to gain their input in design and allow these communities ownership over the data and technology.

Relational frameworks must also question the dubious project of instilling human values in technology and instead to question models' claims of universal application and tech solutionism. Relational frameworks can be helpful if they interrogate industry and research claims to render algorithmic technologies “democratically accessible” and better enable data and algorithmic workers in the Global South to maintain sovereignty over their data as well as their labor.

References:

- Anderson, E. S. (1999). What is the Point of Equality? *Ethics*, 109(2), 287-337.
- Ahiau, N. (2006). Ubuntu and the Obligation to Obey the Law. *Cambrian Law Review*, 37, 17.
- De Beauvoir, S. (1962). *The ethics of ambiguity*, tr. Citadel Press.
- Birhane, A. (2021). Algorithmic injustice: a relational ethics approach. *Patterns*, 2(2), 100205.
- Bovens, M. (2007). Analyzing and assessing accountability: A conceptual framework 1. *European law journal*, 13(4), 447-468
- Bovens, M. (2014). Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism. *In Accountability and European governance*, 28-49.
- Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, 81(5), 825-836.
- Cairney, P. (2016). *The politics of evidence-based policy making*. Springer.

- Cech, F. (2021). The agency of the forum: Mechanisms for algorithmic accountability through the lens of agency. *Journal of Responsible Technology*, 7, 100015.
- Collins, P. H. (1986). Learning from the outsider within: The sociological significance of Black feminist thought. *Social Problems*, 33(6), s14-s32.
- Cooper, A. F., Moss, E., Laufer, B., & Nissenbaum, H. (2022, June). Accountability in an algorithmic society: relationality, responsibility, and robustness in machine learning. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 864-876.
- Cornell, D. & Van Marle, K., 2015, Ubuntu feminism: Tentative reflections, *Verbum et Ecclesia* 36(2), Art. #1444, 8 pages. <http://dx.doi.org/10.4102/ve.v36i2.1444>
- Darwall, S. (2012). Bipolar obligation. *Oxford Studies in Metaethics*, 7, 333-67.
- Fricke, Miranda. *Epistemic Injustice*. Oxford University Press, 2007.
<https://doi.org/10.1093/acprof:oso/9780198237907.001.0001>.
- Fullbrook, E., Fullbrook, K., & Murphy, J. S. (1999). The absence of Beauvoir. *Feminist interpretations of Jean-Paul Sartre*, 45-63.
- Gade, C.B. (2011). The historical development of the written discourses on Ubuntu. *South African Journal of Philosophy*, 30(3), 303-329.
- Gilligan, C. (1993). *In a different voice: Psychological theory and women's development*. Harvard University Press.
- Gouws, A. & van Zyl, M. (2015). Towards a feminist ethics of Ubuntu: Bridging rights and Ubuntu. *Care Ethics and Political Theory*, 164-186.
<https://doi.org/10.1093/acprof:oso/9780198716341.003.0010>.
- Green, B. (2022). Escaping the impossibility of fairness: From formal to substantive algorithmic fairness. *Philosophy & Technology*, 35(4), 1-32.

- Harding, S. (1987). The curious coincidence of feminine and African moralities: *Challenges for feminist theory*.
- Heidegger, M. (2010). *Being and Time*. Trans. by Joan Stambaugh. Albany: State University of New York Press.
- Jecker, N. S. (2022). African ethics, respect for persons, and moral dissent. *Theoria* 88: 666-678.
- Masolo, D. A. (2010). *Self and community in a changing world*. Indiana University Press.
- Meijer, A., & Grimmelikhuijsen, S. (2020). Responsible and accountable algorithmization: How to generate citizen trust in governmental usage of algorithms. In *The Algorithmic Society*. Routledge: 53-66.
- Metcalf, J., Moss, E., Watkins, E. A., Singh, R., & Elish, M. C. (2021, March). Algorithmic impact assessments and accountability: The co-construction of impacts. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*: 735-746.
- Mhlambi, S. (2020). From rationality to relationality: Ubuntu as an ethical and human rights framework for artificial intelligence governance. *Carr Center for Human Rights Policy Discussion Paper Series*, 9.
- Nissenbaum, H. (1996). Accountability in a computerized society. *Science and engineering ethics* 2, 1 (1996), 25–42.
- Noddings, N. (2013). *Caring: A relational approach to ethics and moral education*. University of California Press, 184.
- Putman, D. (1991). Relational ethics and virtue theory. *Metaphilosophy*, 22(3), 231-238.
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Barnes, P. (2020, January). 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*, 33-44.
- Rini, R. (2020). Contingency inattention: against causal debunking in ethics. *Philosophical Studies*, 177, 369-389.
- Scanlon, T. (1998;2000). *What we owe to each other*. Belknap Press.
- Smith, A. M. (2015). Responsibility as answerability. *Inquiry*, 58(2), 99-126.
- Smith, R. L. (1993). Relationality and the ordering of differences in feminist ethics. *Journal of Feminist Studies in Religion*, 9(1/2), 199-214.
- Thompson, M. (2003). *What is it to wrong someone?: A puzzle about justice*. Faculty of Law, University of Toronto.
- Watson, G. (1996). Two faces of responsibility. *Philosophical Topics*, 24(2), 227-248.
- Wieringa, M. (2020, January). What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 1-18.