Topoi
https://doi.org/10.1007/s11245-025-10355-5

A Tale of Two Anthropocentric Abstractions: On the Concept of the "Human" in AI and Environmental Ethics

Fiorella Battaglia[1,2]

Received: 8 October 2025 / Accepted: 1 January 2026
© The Author(s) 2026

Abstract

This paper examines how anthropocentric assumptions shape the normative frameworks of both AI ethics and environmental ethics, arguing that each domain is dependent on a distinct form of anthropocentric abstraction that is increasingly conceptually unstable. In environmental ethics, anthropocentrism is frequently critiqued as an ontological and moral limitation that obscures nonhuman entities and downplays ecological interdependence. By contrast, the European approach to AI ethics – grounded in principles of trustworthy AI, transparency, risk-based regulation, and the protection of fundamental rights – tends to reaffirm an anthropocentric framework, prioritizing human values, rights, and responsibilities in addressing algorithmic opacity, bias, and environmental impacts, while largely overlooking a critical perspective on the shortcomings and harms perpetrated in the name of a misguided or overly narrow understanding of human-centeredness. Drawing on metaethical resources, the paper disentangles the implicit conceptions of the "human" operative in each domain. It argues that both domains grapple with abstract conceptions of the human, either as the telos of nature – occupying a privileged, central role as if all of creation were oriented toward human flourishing – or as fundamentally defined by moral agency, marked by the capacity of being morally responsible for actions, omissions, consequences, and emotions. These tensions highlight the need for an approach that critically engages with the idea of human-centeredeness, while resisting reliance on a merely assumed – and often contested – appeal to "the human". By clarifying how different anthropocentric models operate across ethical contexts, the paper contributes to a more coherent and critically human-centered ethical framework for addressing both technological and ecological challenges.

Keywords Ontological anthropocentrism · Normative anthropocentrism · Humanism · Metaethics · Moral naturalism and non-naturalism · Moral epistemology · Emotions · Externalism · Internalism

1 Introduction

When we engage in the practice of evaluating actions as right or wrong, people as good or bad, or outcomes as fair or unfair, we are aiming to reach conclusions that are correct in virtue of something independent of our own opinions or decisions (Nagel 1980).[1] Appealing to the "human" as a value capable of justifying actions is a special case of this evaluative practice. Simply put, my thesis is that this is not true. First, because this kind of reference assumes that it inherently carries motivational force. Second, because what underpins moral goodness in some contexts serves to identify moral failure in others. At an intuitive level, the presence of a conflicting motivational force points to a certain inconsistency. An agent's action may be legitimized by invoking the category of 'the human', yet that same appeal may also be used to expose injustice. Given that a justification cannot operate in contradictory ways, this weakens the argument for basing moral action on human-centered considerations. Since it cannot fulfill both roles, the inability to do one compromises its ability to do the other. Thus, it seems that the same normative stance is both justified and called into question by the same grounds. At this point, it might seem that we need to expose a logical inconsistency. This has caused

[1] I share Shafer-Landau's belief that careful reflection on the character of moral inquiry and practice lends support to a realist understanding of moral claims (2003).

✉ Fiorella Battaglia
fiorella.battaglia@unisalento.it

1 University of Salento, Lecce, Italy

2 Ludwig-Maximilians-University, Munich, Germany

Published online: 08 March 2026

many to question whether morality grounded in the notion of "human" can effectively guide action (Estlund 2011). It is this sort of practical character of the idea of "human" which I will consider in this paper. Upon closer examination, what initially seems like an inconsistency does not, in fact, hold. I contend that, when properly interpreted, the action-guiding character of ethics based on the concept of the "human" neither challenges nor diminishes the case for anthropocentrism; on the contrary, it reinforces it. For now, I'll define anthropocentrism as the view that only human interests matter morally – sometimes giving more weight to some humans than others – and that non-human interests matter only if they affect humans. The remainder of this paper is structured as follows. In Sect. 2, I argue that anthropocentrism is flawed in multiple respects: it has historically been characterized by assumptions of human superiority and grants moral status only to a narrowly defined "moral circle." Moreover, this approach neglects emotions, which are closely tied to the motivational dimension of ordinary moral practices. In Sect. 3, I introduce Fischer and Ravizza's alternative model of responsibility and control, which shifts the focus from privilege to the burdens and responsibilities that come with being a moral agent. Since anthropocentrism adopts a functional view of agency, Sect. 4 contrasts two perspectives we can take towards others. In Sect. 5, I present the central argument of the paper: that both Kantian ethics and European approach to AI ethics are, by their very structure, incapable of accommodating the moral emotions essential to the sentimentalist shift toward a relational understanding of moral responsibility. In Sect. 6, I argue that the European approach to AI ethics cannot adequately accommodate moral emotions through a so-called human-centric approach – an approach that is crucial for extending moral consideration beyond human beings, encompassing non-human animals, the environment, and our technological future. In Sect. 7, I contend that the European approach to AI ethics also fails to support a genuinely human-centric responsibility. Finally, in Sect. 8, I conclude the paper.

2 The Moral Flaws of Anthropocentrism

This paper is organized around a central problem with two interrelated aspects. The first concerns normative anthropocentrism: How can it simultaneously grant human beings a responsible position within nature, while also supporting a sense of belonging and mutual connection with environmental outcomes, non-human animals, and techno-futuristic values? The second aspect emerges when we recognize that solving the first could still leave the relationship between normative and motivational anthropological reasons ultimately unsatisfying. Specifically, it might fall short of the internalist view, which maintains that a moral judgment granting humans a privileged place in the universe should, by its very nature, entail a corresponding motivation to care for the broader natural world. Instead, we are faced with the fact that a belief that one ought to do something morally does not entail the corresponding emotions that would motivate one to act accordingly (Midgley 1994). Anthropocentrism introduces an internal tension: while it structures our responses to nature, it simultaneously undermines the very moral attitudes – such as love, respect, and awe – that could ground a more meaningful relationship with the non-human world. It provides a framework but erodes the moral depth needed for genuine ethical relation to the nonhuman. The practicality of moral judgements based on this idea thus seems to have a metaethical implication of its own (Smith 1994). The framework that positions humans at the moral center must resort to externalism to account for motivation toward actions and practices that express care for both the world we currently inhabit and the one we will inhabit in the future. Externalism holds that moral judgments are separate from motivation, meaning that recognizing something as morally right or wrong doesn't automatically move us to act (Brink 1986). The irony lies in this: because we don't find moral motivation or affirmation in the world itself – we don't see ourselves reflected in it, we must take on the responsibility of self-care, actively cultivating the motivation to live ethically in a world that neither compels us nor in which we feel at home – a world from which we remain detached.

Kant offers an argument of this sort in his reflections on human duties regarding animals. While Kant does not attribute autonomous moral status to animals, he nevertheless formulates a normative stance consistent with his ethics of duty, which maintains that inflicting suffering on animals is ethically wrong (Wood 1998). What is particularly interesting are the reasons offered to support this claim (Kant 2000): His focus is not on animal rights per se, but rather on the moral consequences of human conduct toward animals. For Kant, cruelty toward animals poses a risk not because it wrongs the animals themselves, but because it threatens to degrade the moral character of human beings. As McShane (2007) points out, however, such mistreatment may have no effect on a person's actual attitudes toward non-human beings. From the standpoint of internalism – according to which moral reasons must be capable of motivating action – this raises a broader concern: the appeal to "the human" as a privileged moral category may lack intrinsic motivational force, thereby placing the entire natural world, and other forms of otherness, on equal moral footing with humanity. Within this framework, respect for animals is not owed to them directly, but rather arises as an indirect duty that humans owe to themselves. In this view, acts of cruelty

toward animals do not violate the animals’ rights but instead compromise the agent’s own moral integrity, thus undermining their moral standing (see Recki 2001; Korsgaard 2004).

Untangling tensions of this kind falls to the moral philosopher. While an appeal to “the human” may resonate as marked by discrimination and lack sufficient motivational force to act on behalf of the others, to relinquish “the human” – particularly when that appeal frames the human as responsible for the well-being of the others – leaves ethics without a foundation. The philosopher’s role in metaethics is to provide an understanding of normative moral practice. Metaethical questions explore the nature of moral judgments themselves – for example, whether moral claims can be true or false, what kind of reasoning (if any) justifies them, and whether moral values are objective or subjective. In contrast, substantive moral questions address what we ought to do, what is right or wrong, good or bad – they aim to provide concrete normative guidance. Since the same normative practices may be justified through diverse means, such clarification is essential for preventing ambiguity. This is exemplified by anthropocentrism, where identical ethical practices – such as valuing human *and* other’s interests – can be supported by very different philosophical arguments. Recognizing this helps clarify why discussions about anthropocentrism can become confused or contested, highlighting the importance of distinguishing between the various justificatory frameworks at play. I refer here both to Kantian ethics and, for instance, to the European Commission’s guidelines, the metaethical rationale of which – when explicitly articulated – may leave a somewhat bitter aftertaste. I will return to this in Sect. 6.

3 Human Agency and Control

Starting from Italian humanism of the 14th century, a conception of human superiority over all other entities has become established, whereby non-human beings are frequently regarded as mere instruments for human use. This idea carries a strong normative orientation toward a hierarchical, pyramid-like model. It asserts that no human ought to remain confined to the form shaped by nature or their particular social circumstances. Instead, everyone ought to develop beyond the limitations naturally and socially imprinted upon them (Pico della Mirandola G 1965).

As already mentioned, the anthropocentric connotation does not necessarily have to be predatory. Indeed, it can challenge Hayward’s definition, according to which anthropocentrism connotes “attitudes, values, or practices which give exclusive or preferential concern to human interests at the expense of the interests or well-being of other species or the environment” (Hayward 2014). This particular mode of anthropocentrism is rooted in a framework of thought from which I aim to dissociate myself. At the very least, it is possible to conceive of an action-theoretical and human-centered paradigm, shaped by data derived from human activity, without thereby endorsing normative assumptions of human superiority. This is exemplified, for instance, by Fischer and Ravizza, who refer to the characteristically human capacity for control over one’s actions.

Ordinarily, we simply assume that we and other human beings are persons and are at least sometimes morally responsible agents. Thus, we assume that we (most of us) at least sometimes have the kind of control that grounds moral responsibility and personhood. Typically, this assumption is deemed so obvious as not to command any attention or elicit even the slightest bit of controversy (Fischer and Ravizza 1998, 14).

Fischer and Ravizza (1998) have developed a conception of personhood that shifts the focus from privilege to the burdens and responsibilities inherent in being a moral agent. Conceptually, this shift marks a move away from the anthropocentrism associated with human supremacy toward an emphasis on the distinctively human capacity to exercise control over one’s actions in the interest of oneself and of others – whether human or nonhuman.

Departing from the hierarchical structure à la Giovanni Pico della Mirandola, in which human beings occupy a privileged position within the cosmic order, Fischer and Ravizza take a relational approach to responsibility: roughly, a key distinction between persons and other creatures is that only persons can be held morally responsible for their actions. Fischer and Ravizza’s account of moral responsibility is grounded in the notion of control, emphasizing that moral responsibility requires a specific kind of guidance control over one’s actions. Control involves two essential aspects: first, that the agent genuinely owns or identifies with the mechanism generating the behavior; and second, that this mechanism is capable of responding appropriately to reasons. In Fischer and Ravizza’s analysis, “most of us” can meet these conditions of moral responsibility. What is significant in this account is that, in order to acknowledge ontological pluralism, it is not necessary to construct a hierarchical order that begins with being, proceeds through the principle of life, becomes further specified in perception, and culminates in thought. Fischer and Ravizza manage to articulate a compelling paradigm of moral responsibility. They establish a non-anthropocentric paradigm by grounding responsibility in the agent’s capacity for guidance control, rather than in inherently human traits. This framework offers promising implications for environmental ethics and animal ethics, as it allows for the ascription of moral responsibility beyond the human sphere. Moreover, by decoupling moral agency from traditional anthropocentric constraints,

their approach opens new avenues for understanding techno-moral futures, where emerging forms of agency – such as artificial intelligence and bioengineered entities – may be considered within the scope of moral responsibility. The notion of guidance implies a form of control whereby an agent's actions flow from a reasons-responsive mechanism that the agent identifies as their own, ensuring that moral responsibility is grounded not simply in the capacity to act, but in the capacity to act for reasons within a relational and normative framework (Santoni de Sio and Hoven 2018).

In light of this perspective, the traditional notion of human supremacy has been replaced by an emphasis on subjectivity and relationality. This shift foregrounds the importance of the first-person perspective – the agent's own experience and self-understanding – and the second-person perspective, which highlights the intersubjective relations and mutual recognition between agents. Together, these perspectives move the focus away from hierarchical dominance toward a more nuanced framework that values the relational dynamics underpinning moral and ontological considerations.

The second-person perspective as developed by Stephen Darwall primarily centers on human-to-human moral interaction – that is, the direct interpersonal relationship where one agent holds another accountable through claims, demands, or recognition. It emphasizes the "second-person standpoint" as the arena where moral responsibility and authority arise through mutual recognition among rational agents (2006). Traditionally, Darwall's framework is rooted in rational agents which are capable of reciprocal interpersonal relations, which generally means human agents. The second-person standpoint relies on communicative practices, normative demands, and mutual accountability, which are often seen as uniquely human capacities.[2]

However, some contemporary philosophers have begun to explore the possibility of extending or adapting the second-person perspective to include non-human animals and artificial agents. This move is particularly significant in light of ongoing debates in animal ethics and emerging questions about our moral relations with intelligent technologies (Crary 2016; Coeckelbergh 2020; Gunkel 2018). By rethinking the second-person stance – traditionally grounded in human-to-human interactions – as a broader framework for moral address and recognition, these thinkers seek to expand the boundaries of the moral community and challenge assumptions about who (or what) can be a legitimate participant in moral dialogue. Yet, these extensions often require rethinking the conditions for accountability, recognition, and reason-responsiveness. A possible, though less explored, contribution to the themes of relationality, the second-person perspective, and the extension of moral responsibility beyond humans comes from P.F. Strawson's focus on reactive attitudes. Strawson's account centers on the interpersonal emotional responses – such as resentment, gratitude, and forgiveness – that arise within personal relationships and ground moral responsibility. By emphasizing these reactive attitudes, Strawson highlights the inherently relational and second-personal nature of moral practices. While traditionally applied within human contexts, this framework opens conceptual space for considering how such attitudes might be extended or adapted to include non-human animals or artificial agents, thereby expanding the scope of moral responsibility in a relational manner. Drawing on Peter Strawson's seminal essay *Freedom and Resentment* (1962), Fischer and Ravizza demonstrate the fundamental role that reactive attitudes and emotional responses play in our moral lives – underscoring how profoundly we are invested in them. Such attitudes are not merely expressions of feeling but are constitutive of our interpersonal relationships and reveal something essential about personhood. When I invoke Strawson's reactive attitudes in this context, I am not asserting whether Large Language Models or AI systems possess consciousness (Chalmers 2022, 2023). Instead, the focus is on whether they can be seen, from a relational standpoint, as appropriate recipients or participants in reactive attitudes – crucially, as objects of blame, praise, or moral accountability. Therefore, I would like to reformulate the question, which becomes: Can other entities – or potentially even our future selves – be placed within the participant stance, as entities toward whom we might legitimately direct reactive attitudes, in principle, even if that stance would need to be reconstructed or adapted? From a more theoretical perspective: Can the participant stance – and the reactive attitudes it entails – be reconceptualized in such a way that it accommodates a plurality of participants thereby challenging the anthropocentric and synchronic assumptions traditionally embedded in moral responsibility frameworks?

If we aim to enlarge the moral circle, then alongside engaging with the participant stance elaborated by Strawson, we must also engage with the notions of temporality and selfhood, and ultimately with the very structure of moral practices. Only by doing so can we conceive of anthropocentrism not solely in a negative light. The question would be whether moral practices must be grounded in shared human characteristics or needs to take some other points of departure. In principle, other entities could be placed within the participant stance, if the stance is reconstructed or expanded beyond its original anthropocentric

[2] See especially Kant's views about communication, according to which it is the relation to the object that renders the representation valid for each individual and therefore communicable (Foucault 2008 and Gerhardt 2013).

framing. But doing so raises deep philosophical and ethical questions.

This shift beyond the original anthropocentric framing opens space for a form of human-centeredness rooted in relational responsibility – one that is fundamentally distinct from human-centeredness understood as a claim to supremacy. The latter relies on a form of non-naturalism and externalism that often results in a patronizing stance toward those excluded from the dominant moral community. At its worst, this attitude becomes exploitative, especially when it justifies a vision of human flourishing conceived in isolation from other beings and the broader ecological context.

4 The Reflective and Relational Dimensions of the Participant Stance

Although grounded in different philosophical frameworks, both Fischer and Ravizza's theory of guidance control and Strawson's account based on reactive attitudes ultimately support the view that agents can be held morally responsible. To motivate the view that Strawsonian reactive attitudes are central to moral responsibility, it is helpful to examine more closely the role they play. Scholars such as Bennett (2002), Holroyd (2010), and Fricker (2016) have argued that the core function of blame—and related attitudes like indignation – is to prompt an alignment in moral understanding between the blamer and the blamed. In this view, blame is not merely punitive – as a form of punishment or retribution for wrongdoing – or expressive, as a way of showing disapproval or signaling a norm violation. Rather, blame is understood as dialogical: a moral practice aimed at restoring or establishing shared norms through engagement and recognition. So far, the participant stance seems implicitly constrained to human participants, given its reliance on capacities like reason-responsiveness and moral communication, which are traditionally ascribed only to humans. This raises important questions about its applicability to non-human agents or future, technologically mediated forms of subjectivity.

Strawson introduces the participant stance in the context of contrasting two perspectives we can take toward others. The participant stance is the ordinary interpersonal attitude we take toward others when we see them as moral agents – as people with whom we stand in relationships of mutual expectation, accountability, and emotional responsiveness. It includes reactive attitudes like resentment, gratitude, and forgiveness. The objective stance is a more detached, clinical, or external view, where we see the other not as a full moral agent, but as someone to be managed, treated, or understood causally – as we might view children, animals, or people with certain psychological conditions. By itself, the Strawsonian perspective does not break away from anthropological dualism, as it continues to ground moral responsibility in a distinction between those capable of participating in intersubjective practices – through reactive attitudes – and those excluded due to a lack of rational or relational capacities. It is now clear how we should proceed: the definition of the participant stance hinges on two crucial elements, both of which warrant further examination. By unpacking each component, we can better assess the scope and potential extension of the participant stance beyond its traditional human-centered framework. The first element concerns the fact that we ourselves can be the recipients of reactive attitudes. This means we are capable of directing moral emotions such as resentment toward our own actions – for example, when we recognize that we have acted in ways that failed to respect the boundaries of the environment. In this sense, the participant stance also enables a reflective moral relationship with oneself, grounded in shared norms and expectations that extend beyond the human sphere. I will call this the "reflective dimension" of the participant stance: the capacity to regard oneself as a moral participant, subject to the same reactive attitudes – such as resentment or guilt – that one might direct toward others. This reflexivity enables an internalization of moral expectations, including those that concern our relationship with the non-human world.

The second element concerns the possibility that the participant stance need not be limited to human relationships. Other entities – such as non-human animals, artificial agents, or ecosystems – could in principle be situated within the participant stance, provided that the stance is reconstructed or expanded beyond its original anthropocentric framework. This suggests a more inclusive, non-anthropocentric model of moral recognition, one that redefines eligibility for reactive attitudes in terms of relational responsiveness rather than exclusively human capacities. This second element may be called the "relational-expansive dimension" of the participant stance. It reflects the idea that this stance need not be tailored exclusively to human-to-human interactions. Rather, if reconceptualized beyond its original anthropocentric framing, the participant stance can accommodate other entities – non-human animals, artificial agents, or ecological systems – as potential participants in moral relationships. This opens space for a broader, more inclusive ethics grounded in relationality rather than species membership.

Within these reflective and relational-expansive dimensions of the participant stance, it becomes possible to explore the functional role of moral emotions in our practices of holding both others and us morally responsible. Reactive attitudes such as resentment, guilt, or indignation are not merely expressive but serve as mechanisms through which shared moral expectations are communicated, contested,

and internalized. These emotions function as bridges between individual moral agents and the broader normative frameworks that structure our more inclusive ethical relationships.

An interesting example of an agent that is capable of taking responsibility toward the whole ecosystem in which their life is embedded would be the Indigenous steward, whose cultural practices and ethical frameworks emphasize relationality, reciprocity, and care for the land as a living community (Kimmerer 2020).

5 Reassessing Human-Centeredness Through the Reactive Attitudes Approach

As outlined in the preceding sections, the moral perspective is defined by its relational, communicative, and reactive dimensions. Adopting a non-ideal theoretical lens, I now advance the central argument of this paper: that both Kantian ethics and the European approach to AI ethics are, by their very structure, incapable of integrating the moral emotions essential to the sentimentalist turn toward a relational conception of moral responsibility – one that alone can sustain a human-centric approach without reproducing predatory forms of anthropocentrism. The Kantian approach is intrinsically rooted in reason and duty, rather than in feelings or emotions (Herman 1990). It excludes a more comprehensive, faculty-encompassing view of moral experience. It suffices to consider that the most relational aspect of his moral theory is the universalization of the "Categorical Imperative". The most famous version asks whether the principle (maxim) behind our action could be made a universal law without contradiction (Kant 1998, 421). Even this aspect, while connecting individuals in the abstract realm of reason, remains impersonal and detached, emphasizing formal consistency over concrete human relationships and emotional responsiveness. This framework confines moral status to those participants who can be both authors and subjects of universal moral laws. As such, it implicitly excludes beings who lack full rational autonomy raising questions about the inclusiveness and scope of Kantian ethics. Moreover, the Kantian framework, by restricting moral agency to rational, autonomous individuals, fails to account for the full range of human experience. It leaves us unsatisfied in light of our contemporary understanding of how we perceive the world around us.

Another category of normative ethics that falls short in addressing the challenge is what is often referred to as the European approach to AI ethics which is grounded in principles of trustworthy AI, transparency, risk-based regulation, and the protection of fundamental rights (European Commission 2019). The European approach to AI ethics characterizes a shift in how ethical questions about AI are framed and approached – moving beyond abstract principles and toward more contextual, relational, and socio-political dimensions of AI systems. Within this framework, there is a commitment to inclusivity that nominally embraces all participants involved, though it often falls short in fully recognizing their diverse roles and perspectives. Responsibility is carefully allocated in terms of agency, ensuring that those who design, deploy, and regulate AI systems or manage environmental resources are held accountable for their actions and their consequences. This responsibility is understood as mediated through a human-centric abstraction, which operates as a conceptual framework that both centers human experience and preserves the capacity for ethical deliberation and decision-making. By balancing inclusivity with structured responsibility, this approach attempts to navigate the complexities of contemporary ethical challenges without resorting to reductive or exclusionary frameworks. Despite its emphasis on contextual dimensions and systemic concerns, it still struggles to offer a compelling justification for a human-centric approach that avoids the typical wrongs historically associated with such a stance (Willert and Knudsen 2025). By drawing on both ideal and non-ideal theory, I have sought to clarify not only what we should demand of a human-centric perspective, but also what, by contrast, ought to be bracketed or set aside.

6 Human-Centric Approach in Ethics

In this section, I explore the human-centric approach within the European approach to AI ethics in greater depth, as it currently constitutes the most comprehensive framework available for incorporating moral dimensions and providing guidance for ethical practice. This approach has primarily been advocated in policy documents, where it has typically only been stated without further elaboration (Future of Life 2017; European Commission 2019). Some scholars, both individually and collaboratively, have explored various methods to establish a framework in which AI is developed and deployed in ways that prioritize human well-being – effectively operationalizing the idea of "AI for people" (Floridi 2013; Floridi et al. 2018). I take a different approach. I am introducing a shift in focus to a more foundational concern. I propose an alternative approach that engages with a metaethical question: what gives those moral claims their normative force?

The point is that unless we critically examine the anthropocentric value theory (or axiology), which assigns intrinsic value exclusively to human beings and regards all other entities – including other forms of life – as merely instrumentally valuable (i.e., valuable only insofar as they serve

human interests), we cannot rely on a justification that truly considers the human being within the broader context of their environment. Consequently, the responsibility to care is justified solely by appeal to a principle imposed from outside. However, this does not entail a rejection of the human role; rather, it requires a more demanding theoretical account capable of providing a robust normative justification for the exercise of control. In sum, it does not deny that meaningful control can only be exercised by human agents, who must justify its use in relation to the needs and moral standing of the various participants involved. The Asilomar document similarly supports the principle of autonomy, insofar as “humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives”. According to this view, obligation derives from the command of someone who has legitimate authority over the moral agent and so can make laws for them. One of the main shortcomings of this view is that obligations – intuitively perceived as good – are accepted without question. According to this perspective, obligation arises from the command of an authority figure – such as the European Commission or other supranational bodies – who possesses legitimate power over the moral agent and is thereby empowered to enact laws governing their behavior.

You must do the right thing because a political authority commands it, or because a political institution to which you have given your consent enacts it as law. In this framework, the source of normativity – the force that obliges moral agents to act – is grounded in the legislative will of a recognized authority. This authority, whether a government, a supranational organization like the European Commission, or another political body, holds legitimate power to establish binding rules that govern behavior. The moral weight of these obligations derives not from abstract principles alone, but from the fact that they have been formally enacted by a body vested with the power to create laws. Consent plays a crucial role here, as it legitimizes the authority's commands and establishes a social contract between the governing institution and the individual. Thus, normativity emerges from a collective agreement to recognize and abide by these legislated rules, anchoring moral obligation in the structures of political authority and legal enactment rather than purely in individual reason or intuition.

A preliminary question that underlies this system of norms is: would we have obligations even if we had not consented to laws or the establishment of treaties? Would the metaethical position endorsed by the European approach to AI ethics amount to a form of naturalism? Notoriously, such a stance would equate human actions to any other form of physical motion, rendering them morally indifferent in themselves. Values, according to this view, are not discovered within the natural world; rather, it is humans – as intelligent beings – who must impose moral values onto nature. This is the view endorsed by Hobbes. Hobbes maintained that there is no real obligation until there exists an authoritative power capable of enforcing the laws of nature. In the state of nature, individuals are governed solely by self-interest and the drive for self-preservation. The so-called “laws of nature” – such as the imperative to seek peace – may be rational guidelines, but they lack binding force unless upheld by a sovereign authority (Hobbes 1996).[3]

Coming up with an answer to this question compels us to confront a fundamental issue concerning the normative force of moral claims – namely, whether such claims can bind us independently of any legal, political, or institutional authority. Are moral obligations grounded in something intrinsic to reason or human practice, or do they require external validation through law, institutional power, or consensual agreement? This tension lies at the core of any attempt to articulate a coherent ethical framework, particularly in domains such as AI governance and environmental responsibility, where authority is often diffuse, contested, or still in the process of formation.

If we turn away from the view that grounds morality in authoritative command and instead take our moral beliefs as the starting point of ethics, we enter a framework where ethics begins with our considered judgments about what is good, which actions are morally impermissible, what kind of distribution is just, and so on. Ethical theory, within this perspective, aims to develop general criteria for what is good, right, and just – criteria that are consistent with certain moral beliefs that appear to be indispensable, while also offering orientation in cases where our moral intuitions are uncertain or even internally conflicting. These ethical criteria are not merely descriptive, they are normative. They do not simply report existing moral beliefs but instead articulate and express moral judgments themselves. According to the realist, our moral or normative thinking isn’t just a projection of our feelings or societal conventions. Instead, we are responding to real features of the world – we understand that certain things truly have normative properties. When we think it right to give to famine relief then, the realist believes that to give to famine relief is actually good, independently of our personal views – and we can *grasp* this truth, much like we grasp that gravity causes objects to fall. (Smith 2005).

The metaethical claim advanced at the outset was that a plausible working hypothesis treats human-centric norms as justified only insofar as they are ultimately reducible to norms of care and responsibility. This perspective emphasizes that moral obligations arise primarily from our relationships and interdependencies with other human beings,

[3] *“Where there is no common power, there is no law; where no law, no injustice.”* Thomas Hobbes, *Leviathan*, Chapter 13.

highlighting the importance of empathy, attentiveness, and accountability. Unlike abstract or purely principle-based ethical systems, a care- and responsibility-centered framework roots morality in the concrete realities of social life, where individuals are mutually vulnerable and dependent. Such norms recognize the lived experiences of moral agents and prioritize responsiveness to the needs and well-being of others. Moreover, this approach aligns well with contemporary challenges, including AI ethics and environmental concerns, by foregrounding the ethical significance of sustaining human dignity and promoting responsibility for wellbeing within interconnected systems. By framing human-centric norms through care and responsibility, metaethics can provide a robust foundation for normative ethics that is both practically relevant and morally sensitive.

The benefits of adopting this approach, I trust, are now evident. For we have seen compelling reasons to adopt a human-centric perspective on normative judgments within the domain of norms concerning both environment and AI. If moral norms ultimately reduce to norms of care and responsibility, then it follows that we have equally strong grounds to be human-centered in our moral judgments as well.

On the contrary, the European approach to AI ethics cannot adequately accommodate ethical sentimentalism through its advocated human-centric approach. Without including the relational and communicative dimension, the inclusivity of the model collapses. Consequently, it fails to extend moral consideration beyond human beings on an equal footing and, as a result, excludes non-human animals, the environment, and our technological future.

7 Unpacking the Normative Divergence of AI and Environmental Ethics

Early approaches to environmental ethics were characterized through a utilitarian framework of cost-benefit calculation (Passmore 1974; Baxter 1974). It took some time before environmental ethics took shape as a metaethical reconsideration of the conditions for the validity of its moral rules (Kramm 2020). Juxtaposing the evolution of these two discourses is especially valuable, as it enables a comprehensive understanding of both the ideal and non-ideal dimensions of foundational ethical debate. I contend that the European approach to AI ethics also fails to ground a genuinely human-centric approach, which does not rely on human rights. Human rights are inherently tied to the concept of *humans* as moral subjects, which can exclude or include non-human animals, ecosystems, or future entities from moral consideration. Onora O'Neill offers a powerful and widely respected critique of human rights discourse – especially its use in contemporary ethics and politics. While she doesn't reject the concept of human rights altogether, she raises critical metaethical and practical concerns about how rights are invoked, justified, and implemented (2005). Moral claims are valid when justified through shared reasoning that is attuned to emotion and respects others. Instead, the kind of anthropocentrism based on human rights limits the scope of ethics and may fail to justify extending moral concern beyond humans. These shortcomings can be explained by drawing on the historical development of the two ethical discourses. For the most part, AI ethics and environmental ethics have progressed in isolation, with little intersection between the two. These diverging developments are particularly significant for rethinking the conceptual boundaries of human-centric ethics. In environmental thought, human-centered approaches are often met with strong criticism, as they are widely regarded as contributing significantly to the environmental crisis. In contrast to environmental ethics, where human-centeredness is often criticized, AI ethics tends to embrace human centrality as a constructive and ethically grounded approach. It is only in recent years that the gap between AI ethics and environmental ethics has begun to receive attention, largely due to rising awareness of the environmental costs and energy consumption associated with the unchecked growth of AI technologies (Bender et al. 2021). Although there is a growing awareness of the issue, the conceptual means to fully engage with it are still lacking. The posthumanism approach, in particular, does not provide a satisfactory response. Although its critical ambition to question the privileged status of homo sapiens is clear, the lack of an action-theoretical framework limits its theoretical effectiveness. (Henry 2020).

The European approach to AI ethics proves to be even more problematic. Indeed, it is unable to disentangle the control that humans have over their actions from the attribution of a privileged moral status. Above all, it fails to offer a foundation that moves beyond human rights, which in this context appear to rest on circular reasoning, amounting to a *petitio principii*. It becomes evident, then, that the stated anthropocentrism must be both scrutinized and more precisely defined. Given the absence of a sufficiently expansive value theory capable of integrating the natural world, future generations, and emerging technological axiologies alongside a reaffirmation of human moral responsibility, the consideration of these entities amounts to a superficial juxtaposition. In this context, juxtaposition is neither employed as a method of contrast nor as a means of clarification. It is not used to deepen understanding by placing two ideas in proximity to reveal what might otherwise remain hidden when considered separately. Instead, juxtaposition here exposes a conceptual void, reducing these considerations to a mere list of discrete items. This suggests that the inclusion

of the environment, non-human animals, and future generations operates merely as an additive "and" rather than being genuinely integrated into a substantive theoretical reframing – which continues to be lacking. From a metaethical standpoint, the shortcomings of the proposed solution become increasingly evident. Moreover, the proposal's normative force is further undermined by its reliance on a voluntaristic model – one that draws on fundamental rights, including those articulated in the Treaties of the European Union and the Charter of Fundamental Rights of the European Union. These frameworks are united by a common foundation rooted in respect for human dignity, which again grants human beings a unique and inalienable moral status. This view, once more affirming that human beings possess a unique and inalienable moral status, underscores their exceptional standing in moral consideration.

This is why anthropocentrism remains widely discredited in environmental ethics: for Callicott, it reflects a moral bias that unjustifiably centers human interests at the expense of ecological integrity (1984). Floridi, by contrast, resists this rejection (2013). He reframes anthropocentrism not as a hierarchical or exclusionary stance, but as a relational position grounded in the ethical responsibilities of human agents within the broader infosphere. Callicott discredits anthropocentrism as a fundamentally flawed moral orientation that places human beings at the center of ethical concern to the detriment of the natural world (2013). Drawing from Aldo Leopold's land ethic and deep ecology, he argues that anthropocentric ethics reduce nature to a set of resources valued only in terms of their utility to humans (Leopold 1949, 224–239). For Callicott, this instrumental view fosters ecological degradation by denying intrinsic value to non-human entities – animals, ecosystems, or even entire species. In his view, the very criteria for inclusion within the moral circle must be reconsidered (Callicott 1984). Moral consideration should extend beyond humans to encompass a more holistic, ecocentric perspective, wherein nature is respected as a moral community rather than merely a backdrop to human activity.

Floridi and Sanders, by contrast, do not reject anthropocentrism outright but seek to reinterpret it within a broader ethical framework (2004). Rather than viewing anthropocentrism as inherently exploitative or morally limiting, Floridi understands it as a situated perspective – an inevitable consequence of the fact that ethical reasoning begins from the standpoint of human agents. In his information ethics and philosophy of the infosphere, he argues for a shift from *human exceptionalism* to *human responsibility* (2013). That is, humans are not privileged because they are superior, but because they are capable of reflecting on and shaping the moral and informational environments in which they operate. Floridi thus reclaims anthropocentrism as a relational and accountable position: one that acknowledges our central role not to dominate, but to care for and maintain the integrity of complex, interdependent systems – including artificial agents and digital ecologies.

This divergence reveals two distinct responses to the crisis of human-centeredness: Callicott calls for decentering the human altogether in favor of a more ecologically balanced ethics, while Floridi proposes a redefinition of human centrality as a moral burden rather than a claim to superiority. The tension between these views is particularly relevant in the context of emerging technologies like AI, where questions of agency, responsibility, and value attribution are being reconfigured beyond traditional humanist boundaries. In my view, it is necessary to mobilize the full resources of metaethical reflection to critically deconstruct the mere reliance on frameworks dominated by fundamental rights. Human-centric approaches grounded exclusively in human rights prove inadequate in addressing relativistic, culturalist, and constructivist challenges. To encompass a broader variety of subjects, the theory of moral responsibility should extend beyond accountability for actions to include responsibility for the emotions we experience toward other entities.

8 Conclusion

The central argument of this paper is that a human-centric attitude is essential for addressing AI and environmental challenges. However, both Kantian anthropocentrism and the European approach to AI ethics are constitutionally incapable of supporting a human-centric stance that does not rely on human exceptionalism. By virtue of its capacity to address grounding morality in human emotions, relational and contextual sensitivity, and motivational force, moral sentimentalism is well-suited to inform the reconceptualization of human-centeredness. Consequently, human-centeredness must be reconceptualized as relational moral responsibility grounded within a sentimentalist framework. Interpreted as an essentially theoretical, not applied discipline, the most important philosophical task for environmental ethics is the development of a non-anthropocentric value theory. Conveyed in a more ideal perspective, this represents the development of a value theory expansive enough to integrate the natural world, future generations, and emerging technological axiologies, alongside a reaffirmation of human moral responsibility. Positively construed, this involves the task of rethinking the moral circle so that it is sufficiently broad and radically relational – that is, grounded in an understanding of ethical responsibility as emerging from the interconnectedness and mutual responsiveness between humans, non-human beings, future generations, and technological systems. Such a reconfiguration

challenges traditional boundaries that isolate moral agents and instead embraces a dynamic, networked view of morality, where obligations arise through relationships rather than abstract rights or isolated individuals. Moreover, such a reconfiguration challenges the paradigm of anthropocentrism not merely from the standpoint of applied ethics, but rather as indicative of a deeper paradigm shift within moral theory itself. This shift moves away from traditional frameworks centered on human exceptionalism and individual autonomy toward an ethic that foregrounds relationality, interdependence, vulnerability, and the intrinsic value of nonhuman entities. It resonates with contemporary developments in environmental ethics, care ethics, and posthumanist philosophy, which collectively question the separateness of the human from the natural and technological worlds.

Author Contributions Single author paper.

Funding Open Access funding enabled and organized by Project DigitHuman. This article was written as part of the European Union-Next Generation EU, Missione 4 Componente 1 CUP F53D23010740001 "DigitHuman-Digital Transition and Human Values" and by the project "Ethical Design for AI", part of the Spoke 6 of the project FAIR - Future AI Research (PE00000013), under the NRRP MUR program funded by the NextGenerationEU.

Data Availability Not applicable.

Declarations

Competing Interests There are no competing interests to declare.

Consent for Publication Granted.

Ethical Approval Not applicable.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

Baxter W (1974) People or penguins: the case for optimal pollution. Columbia University Press, New York

Bender EM, Gebru T, McMillan-Major A, Mitchell M (2021) On the dangers of stochastic parrots: can language models be too big? In: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. ACM. pp. 610–623 https://doi.org/10.1145/3442188.3445922

Bennett C (2002) The varieties of retributive experience. Philos Q 52:145–163. https://doi.org/10.1111/1467-9213.00259

Brink DO (1986) Externalist moral realism. South J Philos 24:23–41. https://doi.org/10.1111/j.2041-6962.1986.tb01594.x

Callicott JB (1984) Non-anthropocentric value theory and environmental ethics. Am Philos Q 21(4):299–309

Callicott JB (2013) Thinking like a planet: the land ethic and the earth ethic. Oxford University Press, New York

Chalmers DJ (2022) Reality+: virtual worlds and the problems of philosophy. Allen Lane, London

Chalmers DJ (2023) Could a large language model be conscious? Boston Review, August 9.

Coeckelbergh M (2020) AI ethics. MIT Press, Boston

Crary A (2016) Inside ethics: on the demands of moral thought. Harvard University Press

Darwall S (2006) The second-person standpoint: morality, respect, and accountability. Harvard University Press, Cambridge

Estlund D (2011) Human nature and the limits (if any) of political philosophy. Philos Public Aff 39(3):207–237

European Commission. High-Level Expert Group on Artificial Intelligence (2019) Ethics guidelines for trustworthy AI. https://ec.europa.eu/futurium/en/ai-alliance-consultation

Fischer JM, Ravizza M (1998) Responsibility and control: a theory of moral responsibility. Cambridge University Press, Cambridge

Floridi L (2013) The ethics of information. Oxford University Press, Oxford

Floridi L, Sanders JW (2004) On the morality of artificial agents. Minds Mach 14(3):349–379

Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, Schafer B (2018) AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach 28(4):689–707. https://doi.org/10.1007/s11023-018-9482-5

Foucault M (2008) Introduction to Kant's anthropology, Nigro R, Briggs K (Eds.), Semiotext(E)

Fricker M (2016) What's the point of blame? A paradigm based explanation. Nous 50(1):165–183

Future of Life Institute (2017) Asilomar AI principles. https://futureoflife.org/ai-principles/

Gerhardt V (2013) Bewusstsein als Funktion der Mitteilung. In: Hüning D, Klingner S, Olk C Das Leben der Vernunft: Beiträge zur Philosophie Kants, De Gruyter, Berlin, Boston

Gunkel DJ (2018) Robot rights. MIT Press, Cambridge

Hayward T (2014) Anthropocentrism. In: Barry J (ed) International encyclopedia of environmental politics. Routledge, London

Henry B (2020) What remains of the human in the Anthropocene? Living between 'nature' and 'culture' in the posthuman condition. Itinerari LIX:157–172. https://doi.org/10.7413/2036-9484028

Herman B (1990) Morality as rationality: A study of Kant's ethics. Garland Publishing, New York

Hobbes T (1996) Leviathan. Edited by Tuck R, Cambridge University Press, Cambridge

Holroyd J (2010) Punishment and justice. Soc Theory Pract 36:78–111

Kant I (1998) Groundwork of the metaphysics of morals. Cambridge University Press, Cambridge, Translated by Gregor M

Kant I (2000) Critique of the power of judgment. Cambridge University Press, New York

Kimmerer RW (2020) Braiding sweetgrass: indigenous wisdom, scientific knowledge and the teachings of plants. Penguin, Harmondsworth

Korsgaard C (2004) Fellow creatures: Kantian ethics and our duties to animals. Tanner Lect Hum Values 24:77–110

Kramm M (2020) When a river becomes a person. J Hum Dev Capabil 21:307–319
Leopold A (1949) The land ethic. And sketches here and there. Oxford University Press, Oxford, In a sand county almanac
McShane K (2007) Anthropocentrism vs. nonanthropocentrism: why should we care? Environ Values 16(2):169–185. https://doi.org/10.3197/096327107780474555
Midgley M (1994) The end of anthropocentrism? Royal Institute of Philosophy Supplement 36:103–112
Nagel T (1980) The limits of objectivity. In: McMurrin S (ed), The Tanner lectures on human values I, University of Utah Press, Sa1t Lake City
O’Neill O (2005) The dark side of human rights. Int Aff 81(2):427–439
Passmore J (1974) Man’s responsibility for nature: ecological problems and Western traditions. Duckworth
Pico della Mirandola G (1965) Oration on the dignity of man. Translated by Lindsay AD, Bobbs-Merrill, Indianapolis, originally published 1486
Recki B (2001) Achtung vor der zweckmäßigen Natur Die Erweiterung der kantischen Ethik durch die dritte Kritik. In: Gerhardt V, Horstmann RP, Schumacher R (eds) Kant und Die Berliner Aufklärung: Akten des IX Internationalen Kant-Kongresses. Walter de Gruyter, New York, pp 296–304
Santoni de Sio F, van den Hoven J (2018) Meaningful human control over autonomous systems: a philosophical account. Front Robot AI 5:15. https://doi.org/10.3389/frobt.2018.00015
Shafer-Landau R (2003) Moral realism. A defence. Oxford University Press, Oxford
Smith M (1994) The moral problem. Basil Blackwell, Oxford
Smith M (2005) Meta-ethics. In: Jackson F, Smith M (eds) The Oxford Handbook of Contemporary Philosophy. Oxford University Press, pp 3–30
Strawson PF (1962) Freedom and resentment. Proc Br Acad 48:1–25
Willert KB, Knudsen N (2025) Anthropocentrism in the Anthropocene: toward an ostensive humanism. Environ Hum 17(3):638–656. https://doi.org/10.1215/22011919-11956643
Wood AW (1998) Kant on duties regarding nonrational nature. Proc Aristot Soc Suppl Vol 72(1):189–228

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.