

Article

Anthropocentrism and Environmental Wellbeing in AI Ethics Standards: A Scoping Review and Discussion

Eryn Rigley , Adriane Chapman ^{*}, Christine Evers and Will McNeill

School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BF, UK

^{*} Correspondence: e.rigley@soton.ac.uk (E.R.); adriane.chapman@soton.ac.uk (A.C.)

Abstract: As AI deployment has broadened, so too has an awareness for the ethical implications and problems that may ensue from this deployment. In response, groups across multiple domains have issued AI ethics standards that rely on vague, high-level principles to find consensus. One such high-level principle that is common across the AI landscape is ‘human-centredness’, though oftentimes it is applied without due investigation into its merits and limitations and without a clear, common definition. This paper undertakes a scoping review of AI ethics standards to examine the commitment to ‘human-centredness’ and how this commitment interacts with other ethical concerns, namely, concerns for nonhumans animals and environmental wellbeing. We found that human-centred AI ethics standards tend to prioritise humans over nonhumans more so than nonhuman-centred standards. A critical analysis of our findings suggests that a commitment to human-centredness within AI ethics standards accords with the definition of anthropocentrism in moral philosophy: that humans have, at least, more intrinsic moral value than nonhumans. We consider some of the limitations of anthropocentric AI ethics, which include permitting harm to the environment and animals and undermining the stability of ecosystems.

Keywords: AI ethics; human-centred AI; environmental ethics; scoping review; AI ethics standards; anthropocentrism



Citation: Rigley, E.; Chapman, A.; Evers, C.; McNeill, W. Anthropocentrism and Environmental Wellbeing in AI Ethics Standards: A Scoping Review and Discussion. *AI* **2023**, *4*, 844–874. <https://doi.org/10.3390/ai4040043>

Academic Editors: Pablo Rivas and Gissella Bejarano

Received: 28 August 2023

Revised: 26 September 2023

Accepted: 28 September 2023

Published: 8 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As artificial intelligence (AI) deployment has broadened, so too has an awareness for the ethical implications and problems that may ensue from this deployment. In response, groups across multiple domains have issued AI ethics standards, ranging from short lists of principles to extensive documents explicating AI ethics problems and instructions on how to mitigate these. New AI ethics standards are continually being published, so it is difficult to ascertain exactly how many currently exist. As a general guide, the *AI Ethics Guidelines Global Inventory* counted 167 AI ethics guidelines in May 2022.

There are now so many individual AI ethics guidelines that groups have begun to survey standards, summarising and consolidating common themes or assertions about the AI ethics landscape [1]. Surveying the AI ethics landscape is now a well-established methodology for examining the trends, themes, and values within current AI research and development (some examples include [2–7]). Surveys combine a range of sources, such as corporate, government, international coalitions, and non-profits, and include a range of formats, ranging from white-papers to professional codes of ethics and reports from private industry.

There is a lack of methodological and theoretical consensus across such a wide variety of AI ethics standards. Crawford [1] argues that this has led to considerable limitations, including conflicting or vague ideals and definitions, and a lack of overarching or standard protocols. Across the AI ethics space, there is little agreement on what is important and how ethical problems that relate to AI ought to be mitigated [8]. The ethical positions that are used in AI ethics tend to be based on principles as mid-level, applicable, and action-guiding

concepts [8]. Mittelstadt argues that these ‘broadly acceptable’ and ‘vague’ principles are applied in order to find some consensus in an otherwise nebulous space [9] (p. 503). Indeed, despite the vast number of AI ethics standards, and despite the lack of consensus, the range of principles applied in AI ethics standards is limited [8,10]. In perhaps the most notable survey of AI ethics standards, Jobin et al. [5] (p. 395) found the most prominent principles across AI ethics standards included transparency, justice and fairness, non-maleficence, and responsibility.

Such vague principles can oftentimes be prescribed without due investigation into their merit and limitations and without clear, established definitions. One such accepted ethical prescription that has seen a notable increase without due investigation is a commitment to ‘human-centredness’. AI ethics principles are often directly linked to established human rights [6], and human-centredness is increasing as a prescribed ethical grounding for AI development and deployment [11]. However, the application of ‘human-centredness’ comes with a range of different definitions [12] and without investigation or instruction on how this ought to interact with other commitments, such as environmental sustainability.

In what is the most prominent and perhaps only survey of AI ethics frameworks from an environmental lens, Owe and Baum [13] re-examined the surveys of Jobin et al. [5] and Baum [14] in terms of concern for nonhumans and anthropocentrism, as it is defined in moral philosophy. Owe and Baum found that, in theory, moral consideration for nonhumans is compatible and consistent with certain anthropocentric perspectives and with many key principles across existing AI ethics frameworks [13] (pp. 519, 525). These principles include transparency and explainability, diversity, non-discrimination, fairness, privacy, and data governance—some of the major ethical principles noted by Jobin et al. [5]. Nevertheless, they argued, “specific treatments of the principles commonly neglect nonhumans” and that concern for nonhumans is vastly outweighed in AI ethics by concern for humans [13] (pp. 519–520).

Though Owe and Baum [13] uncover some key trends across the AI ethics landscape with regards to environmental wellbeing through their qualitative review, some questions remain. In particular, it is unclear whether broader applications of the term ‘human-centred AI’ within AI ethics standards are consistent with anthropocentrism, as defined in moral philosophy, and how these commitments to human-centred AI interact with the moral considerations for nonhumans.

Objectives and Contributions of This Work

The overall objective of this work was to examine the commitments to human-centredness within AI ethics standards and how these commitments interact with other ethical considerations, namely, environmental wellbeing. To approach this, we undertook a scoping review of AI ethics standards, particularly by examining the commitments to human-centredness and considerations for humans and nonhumans. We also critically analysed our findings within the context of human-centred AI more broadly and within the context of anthropocentrism. Overall, we found that the application of human-centredness in AI ethics conforms with the definition of anthropocentrism as established and applied in moral philosophy, namely, that humans have more moral value than nonhumans. The contributions of this work are:

- A discussion of human-centredness in AI, particularly the historical roots, groundings, definitions, and applications of this commitment;
- A scoping review of human-centredness and the moral considerations of humans and nonhumans across 146 AI ethics standards;
- An examination of how applications of anthropocentrism, as defined in moral philosophy, would play out in the development and deployment of AI systems.

Section 2 discusses the background of human-centredness in AI and the philosophical groundings of anthropocentrism. Here, we outline the supposed gap between human-centredness, which is considered as a vague, high-level concept to unify and find consensus among the AI landscape, and anthropocentrism, which is considered as the well-established

claim for moral hierarchies between humans, nonhumans, and the environment. Section 3 will outline the methodology employed to survey the AI ethics landscape for both human-centredness and environmental concern. The results of this survey will be discussed in Section 4. Section 5 will critically analyse the results within the context of anthropocentrism and human-centred AI more broadly. Here, we argue that human-centredness in AI ethics reflects the moral hierarchies of anthropocentrism. We also discuss how these anthropocentric moral hierarchies play out in practice in the regulation of AI systems in Section 5.1.

2. Background

2.1. Asimov

Early AI ethics standards drew inspiration from Isaac Asimov's laws of robotics from his 1950s science fiction novel, *I, Robot*. These are seen as the first governing laws for autonomous technology, explicating an implicit consensus to use and develop technology for the good of humanity [15]. Asimov's three original laws for robotics are as follows: a robot may not injure a human being or, through inaction, allow a human being to come to harm; a robot must obey orders given by human beings except where such orders would conflict with the first law; and a robot must protect its own existence as long as such protection does not conflict with the first or second laws [16]. A fourth law was later added by Asimov, which can be seen across the AI ethics landscape even today: a robot may not harm humanity, or, by inaction, allow humanity to come to harm [15,17]. Together, these laws have been used as a benchmark for "mainstream" AI ethics [18] (p. 209) or as the "essential requirement" for ethical AI [19] (p. 1409).

Murphy and Woods [20] suggest that Asimov's laws have been so successfully inculcated into the public consciousness through entertainment that they now appear to have shaped society's expectations about how technology and humans ought to interact. Multiple domains, including philosophy, AI, and medicine, have discussed the ethics of robots in society using Asimov's laws as a reference [20]. In fact, a draft of the *European Civil Law Rules of Robotics* prescribed Asimov's three laws to designers, producers, and operators of robots, including robots assigned with built-in autonomy and self-learning (sec. T. [21]). On reflection, a later iteration of this civil law conceded that it was culturally and scientifically "wrong" to mention Asimov's laws as the work of fiction could not be taken as true legal principles [22] (p. 8). While Asimov's laws were described as "unfit to protect humanity", it was nevertheless suggested that a framework could be drawn to ensure that future AI benefits humanity [22] (p. 13). This later work posited new rules that remain hauntingly familiar: "The first principle of roboethics is to protect humans against any harm caused by a robot" [22] (p. 20).

2.2. HCAI

Today, a commitment to protect humanity in AI development and deployment seems ubiquitous across the AI ethics landscape. Reflecting the increasing importance and integration of AI in peoples' lives, there is a move towards human-centred artificial intelligence (HCAI), where the goal is to put the human, rather than technology, at the centre of AI development [12] (pp. 25–26).

HCAI is arguably entrenched in and inspired by Asimov's laws. However, Asimov made these laws inherently and purposefully vague, which adds to the plot of *I, Robot* but detracts from their usefulness as real action-guiding principles [22]. Mittelstadt [9] notes that there is often no clear definition of the vague, and oftentimes contested, concepts in AI ethics, and the author also notes that without due definition, these concepts are not specific enough to be action-guiding. Indeed, there is no single, broadly accepted definition of HCAI, but many different definitions that combine the criteria from a human-centred design with AI-specific factors, such as data usage, bias, and uncertainty of outcomes [12]. Some of the applications of the term HCAI include: augmenting rather than displacing

human abilities; explainability and accountability; preserving human control; fair and trustworthy use of data; and providing efficient solutions to users' problems [12].

Without due analysis of the merits and drawbacks of human-centredness as well as an investigation of alternative approaches to AI ethics, it is unclear as to why this commitment ought to be adopted over any other normative grounding. Moreover, without clear definitions of human-centredness, it is unclear exactly what HCAI ought to refer to. This is particularly the case when human-centredness is applied to AI ethics as anthropocentrism has a unique, established definition in moral philosophy.

2.3. Anthropocentrism

In moral philosophy, anthropocentrism, which literally translates to human-centredness, is the view that only humans have intrinsic moral value or, at the very least, humans have more intrinsic moral value than nonhumans [23]. In philosophy, intrinsic value is commonly defined as value that an object has for its own sake, whereas extrinsic value is value that an object has for the sake of something else or in relation to some other object [24]. This definition of anthropocentrism dates back to antiquity, heralded most notably by Aristotle, who summated, “[nature] has made all animals specifically for the sake of man” [25] (p. 1137). This definition of anthropocentrism has been consistently applied within philosophy and has been continued by, among others: Aquinas, who argued, “But plants exist for the sake of animals; indeed, some animals exist for the sake of others, and all exist for the sake of man” [26] (Bk. 3, Pt. 2, Ch. 127); Kant, who stated, “all animals exist only as a means, and not for their own sakes, in that they have no self-consciousness, whereas man is the end” (27:495) [27]; and, more recently, Passmore, who declared, “I treat human interests as paramount. I do not apologise for that fact” [28] (p. 187).

Anthropocentrism can be divided into strong or moderate. Strong anthropocentrism is the claim that humans are the only bearers of intrinsic moral value: “On a bold version of the view, only humans count, morally speaking; nonhumans don't count at all” [29] (p. 2). On this strong anthropocentric account, nonhumans have the most extrinsic value to humans: for survival, pleasure, monetary gain, and so on. Meanwhile, moderate anthropocentrism claims that nonhumans still have intrinsic value, though less so than humans [29] (p. 2). This means that nonhumans or the environment are not solely valuable for the sake of humans but matter for their own sake, albeit, less than humans.

Simply put, both strong and moderate anthropocentrists hold that there is something morally special about humans that gives them significant moral value over nonhumans. This means that, under anthropocentrism, in cases where environmental or nonhuman wellbeing conflicts with human wellbeing, human wellbeing wins out.

Crucially, this definition of anthropocentrism differs from some uses of the term ‘human-centred’ in AI or in HCAI, particularly in the uses found in design or safety, which aim to ensure the safety and wellbeing of human users of AI; however, this prioritisation of humans does appear in other instances of HCAI. Consider, for instance, “empower both individuals and society” or “augment rather than displace human abilities, as HCAI seeks to enhance human performance and human-AI collaboration by integrating artificial and human intelligence” [12] (p. 2). Here, the wellbeing of specifically human individuals is seen as of the utmost importance. Consider also the high-level expert group that states that reasoning between conflicting principles should never violate “fundamental rights and correlated principles” such as “human dignity”, as such principles are “absolute and cannot be subject to a balancing exercise” [30] (p. 13). These examples of prioritising human beings may intend to place human wellbeing over and above technology, monetary gain, or machine efficiency as opposed to over and above nonhuman wellbeing or environmental sustainability. However, this is not made explicit. Herein lies the problem of relying on vague, high-level principles: though they are likely to find consensus, in doing so, we are left without clear definitions of terms and without guidance on how to use and balance these principles with other commitments and concerns.

2.4. Environmental Ethics and the Value of Nonhumans

There is a long-standing debate within environmental ethics over anthropocentrism and the value of nonhumans: are nonhumans valuable in and of themselves, or are they valuable for the sake of something else, such as humans; does harming nonhumans in turn harm humans?

There are many reasons why humans would want to extend care to the nonhuman world. We rely on the nonhuman world for shelter, food, and medicine and, in many cases, environmental destruction motivated by monetary gain or pleasure has threatened not only the wellbeing of nonhumans, but our own survival as well.

Anthropocentrism can be consistent with extending moral consideration to the non-human world. Moderate anthropocentrists maintain that nonhumans have some intrinsic value, and so the wanton destruction of nonhumans would be impermissible in many cases. Moreover, even if humans are the sole object of intrinsic value, as strong anthropocentrists claim, we can still have indirect duties to nonhumans for the sake of humans.

For instance, Kant argued that our duties to nonhuman objects “allude, indirectly, to our duties towards men” (27:460) [27]. Causing needless destruction to the natural world indirectly affects other humans, as others may have use for it (27:460) [27]. Moreover, the wanton destruction of flora and harm caused to nonhuman animals will impede on our own aspirations towards moral perfection, and as this moral perfection is the overarching goal of humans, we have indirect duties to not cause needless harm to the nonhuman world [27,31]. More recently, Passmore states that we ought not to impede on the freedom of any human by “destroying the natural world which makes that freedom possible” [28] (p. 195).

Meanwhile, non-anthropocentric environmental ethicists argue that nonhumans do have intrinsic value, with varying degrees of inclusivity. Sentientists argue that all sentient animals, that is, those with the capacity to suffer, deserve equal moral consideration with humans [32]. Biocentrists argue, more inclusively, that all living beings have intrinsic moral value in so far as they have morally significant interests, including survival, wellbeing, and flourishing, which matter beyond what they provide for humans [33,34]. Even more inclusive is a relational ecocentric ethic, which considers the interdependent relations between ecological objects, including humans, nonhuman animals, and plants, such that these individuals are considered part of one interacting whole [35,36]. Humans are considered “plain” members of this community, just like any other organism, which grounds the extension of moral consideration from humans alone to the wider environment [36] (p. 194). Deep ecology, one branch of ecocentrism, extends the “self” to identify with the broader environment, including animals, plants, and ecosystems [37]. Deep ecology describes all parts of this broader “self” as having intrinsic value [37].

The value of nonhumans remains the object of debate within environmental ethics, and different branches have different reasons to consider nonhumans as worthy of moral concern. While strong anthropocentrists may care for the environment indirectly, for the sake of humans, moderate anthropocentrists and nonanthropocentrists will see parts and even all of the nonhuman world as deserving of moral concern for its own sake.

2.5. Background Conclusion

Human-centredness remains an under-examined concept in AI. It is commonly discussed, accepted, and prescribed without due investigation into its merits and limitations, nor with a clear, shared understanding of its meaning and how it ought to be embedded into the wider framework and commitments of AI ethics. This paper examines the conformance of AI ethics standards to human-centredness by undertaking a scoping review of 146 AI ethics standards. As anthropocentrism in moral philosophy discusses the moral considerations of humans in relation to that of nonhumans, this paper will also examine ethical considerations for nonhumans and the environment throughout these standards. Upon a critical analysis of our findings, we find that human-centredness in AI reflects the

definition of anthropocentrism in moral philosophy, having higher levels of moral concern for humans than for nonhumans.

3. Materials and Methods

This review of AI ethics standards follows the five-step methodology for a scoping review as set out by Arksey and O'Malley [38] (pp. 22–23). Paterson et al. [39] note that although there is no universally accepted definition or purpose for a scoping review [40–43], they are particularly useful in providing an overview of a broad topic [42,44] and can be flexible when the literature is vast and complex [38]. Because the AI ethics landscape is indeed vast, diverse, and complex, a broader scoping review was chosen over a systemic review, which can be more narrow and limited in its focus and inclusion criteria [39].

Scoping reviews differ substantially from systematic reviews in the methodology, expectations, tools, and outputs [45]. For instance, because scoping reviews are designed to provide a broad and inclusive overview of the existing evidence base, formal assessments of the bias or quality of the included studies are generally not performed [45,46]. Although there is no exacting set of procedures for scoping reviews, the following five-step methodology for a scoping review as set out by [38] is commonly applied [39,42]:

1. Identifying the research question;
2. Identifying relevant standards;
3. Standard selection;
4. Charting the data;
5. Collating, summarising, and reporting the results;

Section 3.1 outlines the research question. Section 3.2 identifies the relevant standards. Section 3.3 discusses the sources, including the inclusion and exclusion criteria, for the standard selection. Section 4 charts the data and collates and reports the results.

3.1. Identifying the Research Question

The research question this scoping review aims to answer is: how prevalent is human-centredness in AI ethics standards, along with moral considerations of humans and nonhumans?

3.2. Identifying Relevant Standards

Regulation, codes of conduct, standardisation, certification, and accountability and governance frameworks are some examples of the non-technical tools for practitioners to implement and conform to AI ethics principles [30] (pp. 22–23). This work surveys 146 examples of these non-technical tools, henceforth referred to as 'standards'.

3.3. Standard Selection

The sources used to collect AI ethics standards were the study by Jobin et al., (2019) [5] and *The AI Ethics Guidelines Global Inventory* [47] (accessed May 2022). Jobin et al. [5] was chosen as a primary source for guidelines as this work is so far the largest and most comprehensive survey of AI ethics standards. *The AI Ethics Guidelines Global Inventory* was chosen as a supporting source, as this inventory includes standards not included in and/or published after Jobin et al. [5]. The citations identified from the sources were compiled with the identifying review criteria: *Issuer*, *Year*, and *Domain*. The compiled citations were then manually screened in accordance with predetermined exclusion criteria: foreign languages, unofficial blog posts that do not constitute non-technical AI ethics compliance tools [30], duplicates, and standards that have since been deleted. In total, we identified and surveyed 146 ethical AI standards across 8 domains and published over 10 years. Figure 1 abstracts this process of identifying relevant standards.

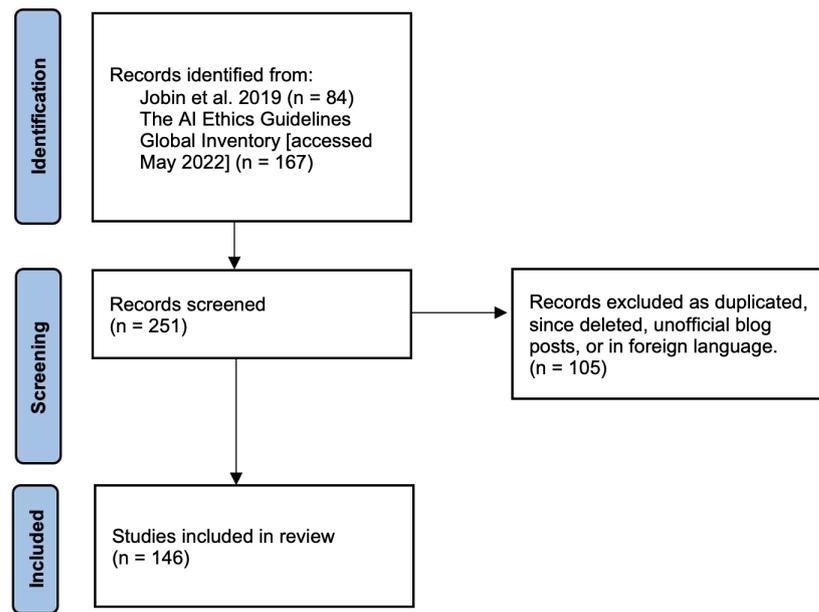


Figure 1. PRISMA flow diagram for source identification using sources Jobin et al. [5] and The AI Ethics Guidelines Global Inventory [47].

3.4. Review Criteria

A data extraction sheet was developed through iterative pilot testing on a small subset of included articles, which refined both the design and review criteria to be objective with the aim of reducing author bias and subjectivity. Table 1 explicates the final review criteria by which each framework was surveyed. Appendix A.1 lists all of the standards that we surveyed. We also reviewed the domains under which each framework was published, which included *government, non-profit, professional association, research group, political party, private, religious, and research council*. The domain under which each framework was published was identified by Jobin et al. [5] and *The AI Ethics Guidelines Global Inventory* [47].

Table 1. Review criteria.

Feature	Criteria for Inclusion.
Human-centric	Explicitly states ‘human-centredness’ or equivalent phrasing
Inclusion of humans	Extends concern to humans anywhere within the standard
Inclusion of nonhumans	Extends concern to nonhumans anywhere within the standard
Humans in core principles	Includes a single principle/foundations/priority/aim of human respect/care/wellbeing/values/or equivalent phrasing
Nonhumans in core principles	Extends concern to humans within one or more core principles

3.4.1. Inclusion of Humans and Nonhumans

We examined whether the standards included humans and/or nonhumans within moral concern. *Inclusion of humans* and *nonhumans* was defined using broad criteria, namely, did the framework include humans and/or nonhumans anywhere within the text or principles? This broad criteria were used to offer a generous perspective over the concern for humans and nonhumans across AI ethics standards.

3.4.2. Humans and Nonhumans in Core Principles

A binary review was made as to whether standards included a set of core principles or values. Recommendations were not included as core principles as many standards had both principles and recommendations and this work is primarily interested in the normative motivations behind the recommendations.

The content of these core principles was examined. We examined whether any of these core principles were related to human wellbeing, respect, dignity, or equivalent phrasing. For example, “AI should positively contribute to the wellbeing of humans”. If so, we stated that this standard *included humans in the core principles*. We also examined whether any of these core principles were related to nonhumans; this might be in addition to humans. For example, “AI should not harm humans or animals”. In this case, this standard was defined as *including humans and nonhumans in the core principles*. Alternatively, the principles might be solely related to nonhumans. For example, “AI development should be environmentally sustainable”. In this case, this standard would be defined as *including nonhumans in the core principles*. We also examined whether nonhumans were included in the core principles for their own sake or for the sake of humans. For example, “AI development should not damage the natural environment on which humans depend for survival” would be defined as *including nonhumans in the core principles for the sake of humans*. Appendix A.2 includes the sources with the core principles categorised by their inclusion of *humans, nonhumans, nonhumans and humans, and nonhumans for humans* within their core principles.

We can see the relationship between these criteria in Figure 2 below. Within the set of standards that include humans and nonhumans anywhere in the text are the standards with core principles. Within the set of standards with core principles are the standards that have core principles that relate to humans and core principles that relate to nonhumans. We have clarified this further using examples from the AI standards included in this survey in Table 2 below.

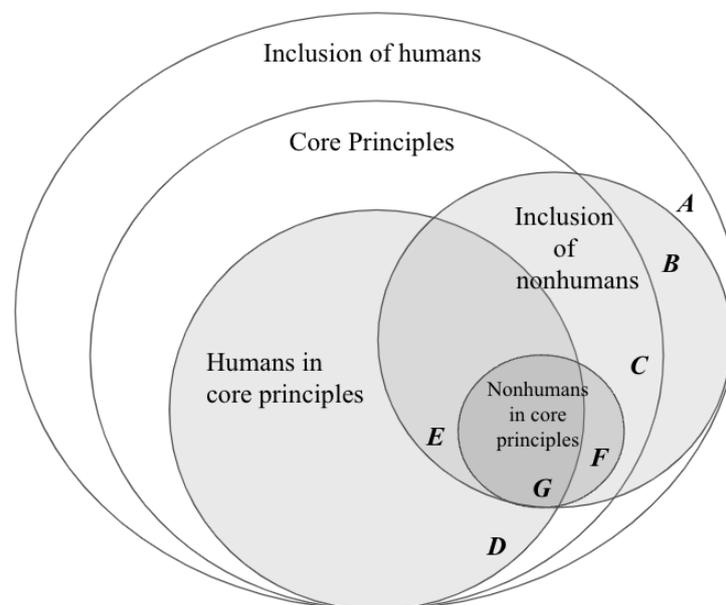


Figure 2. Venn diagram of various possible relationships between review criteria represented by letters A–G.

Table 2. Example text of relationship between review criteria.

Venn Diagram Area	Example Text	Reference
(A) Includes humans	“Robots are rapidly transitioning into human social environments (HSEs), interacting proximately with people in increasingly intrusive ways”	A Code of Ethics for the Human-Robot Interaction Profession by Riek and Howard [48], p. 1
(B) Includes humans and nonhumans	“We cannot see the global environmental and labor implications of these tools of everyday convenience, nor can we meaningfully advocate for fairness, accountability, and transparency in AI systems, without an understanding of this full stack supply chain”	AI Now Report 2018 by AI Now [49], p. 34
(C) Has core principles	The 12 Principles of Automated Driving: safe operation; operational design domain; vehicle operator-initiated handover; security; user responsibility; vehicle-initiated handover; interdependency between vehicle operator and ADS; safety assessment; data recording; passive safety; behaviour in traffic; and safe layer	Safety first for automated driving by Wood et al. [50] pp. 7–10.
(D) Has core principles relating to humans	“AI should be at the service of society and generate tangible benefits for people”	AI Principles by Telefonica [51], principle 3
(E) Includes humans and nonhumans and has core principles relating to humans	“we must guarantee an outlook in which AI is developed with a focus not on technology, but rather for the good of humanity and of the environment”; “every human being has equal dignity”	Rome Call for AI Ethics by Rome Call [52], p. 4; principle 2.
(F) Has core principles relating to nonhumans	“Digitization should serve to conserve natural resources”	10 ethical guidelines for the digitalisation of companies by Hochschule der Medien [53], principle 10
(G) Has core principles relating to nonhumans and humans	“Promoting well-being, preserving dignity, and sustaining the planet”	Ethical Framework for a Good AI Society by Floridi et al. [54], core principle 1

3.4.3. Human-Centred

This work defines ‘human-centred’ standards as those that used the term ‘human-centred’, ‘human-centric’, or equivalent phrasing. This definition of ‘human-centred’ therefore reduces to an explicit commitment to ‘human-centredness’. This removes subjectivity and author bias and reduces doubt as to whether a framework actually supports human-centredness. This definition of ‘human-centred’ also differs from anthropocentrism as defined in Section 2.3 and may encompass some of the examples of HCAI as outlined in Section 2.2.

3.5. Limitations

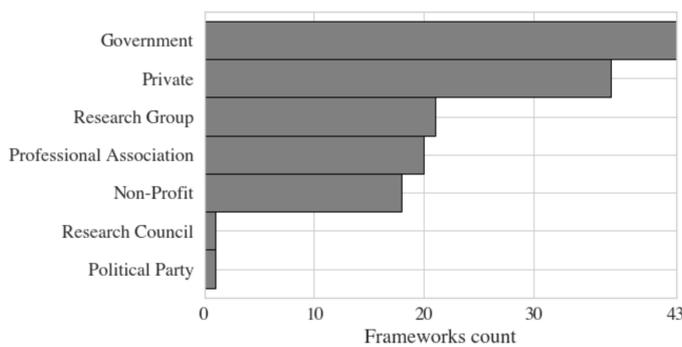
There are several limitations to our survey. A language bias may have skewed our corpus towards English or English-translated results. Our survey took place in 2022 and thus excludes more recent standards.

Our review faces the typical limitations of quantitative analyses of reviewed studies, including that the themes that are numerically identified may not necessarily indicate actual weight or significance. A higher count of “inclusions” in this manner does not necessarily indicate significance but does indicate a form of prioritisation. This is further exacerbated by the fact that philosophy, particularly ethics, is not an exact science but inherently difficult to quantify and quantifiably compare. These limitations have been mitigated by developing review criteria that was explicit, objective, and data-driven, with the aim of reducing subjective author biases. We also address these limitations with an analysis of our results using qualitative data, drawing out quotes from the reports to substantiate our

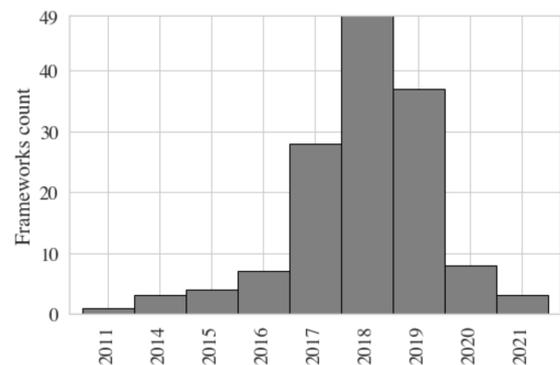
findings, and by critically analysing our findings within the wider context of HCAI and moral philosophy.

4. Findings

The range of AI ethics standards included in this survey reflect the broad and varying AI ethics landscape. Given increasing concern for the moral problems of AI systems, it is encouraging to see such a wide variety of AI ethics guidelines being offered from a range of disciplines across such a large time frame. The largest domain that published AI ethics guidelines is government bodies, which issued 30% of the standards included in this survey. This is followed by private industry, which issued 27%, and research groups, which published 16%. Professional associations and non-profits each contributed to 14% of the standards surveyed. Other domains were religious groups and research councils, which each issued one framework, and political parties, which issued two. There were also standards that came from a collective of various domains. For example, Fairness, Accountability, and Transparency in Machine Learning (FATML) [55] constitutes both a non-profit and professional association, and the *Rome Call for AI Ethics* was published by a collective of religious, private, and governmental groups [52]. Figure 3a charts the domain results as a horizontal bar graph.



(a) Number of standards across domains



(b) Number of standards across years

Figure 3. Overviews.

There is a clear diversity in the parties involved in AI who wish to publish ethical guidelines over its development and usage. The AI ethics standards also heavily diverged with respect to the issuers and their focus or specific area of expertise, including business ethics, future employment trends, and human rights. The standards also mentioned a variety of AI systems, including machine learning and social embodied systems. The publishing year of the surveyed standards was also broad, spanning 2011 to 2021, although most standards surveyed were published between 2017 and 2019. Figure 3b charts the year distribution.

4.1. Concern for Humans and Nonhumans

Figure 4 charts the different results for the inclusion of humans and nonhumans both within and outside of the core principles across standards. Despite such a diversity of standards, the AI ethics landscape we surveyed as a whole extends concern for humans more so than concern for nonhumans. All standards included concern for humans and 26% of those surveyed included concern for nonhumans.

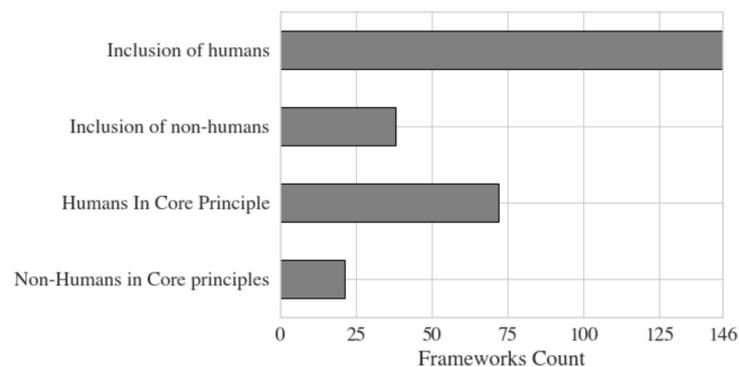


Figure 4. Inclusion of humans and nonhumans.

Concern for nonhumans ranges from moral consideration of the nonhuman world to practical consideration of how the nonhuman world may undermine the success of AI deployment. Cigref’s *A Guide for Professionals of the Digital Age* extended consideration towards the environmental footprint of digitisation [56] (p. 12), and the *AI Now 2019 Report* includes concern for the environmental and labour costs of AI systems [1] (p. 6). Other standards considered the application of AI in pursuing environmental wellbeing. For instance, Korea’s *Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society* includes environmental protection and energy as possible future applications for AI systems [57] (p. 42). Others considered the environment only in terms of how this may affect the deployment of AI systems. For example, *Safety First for Automated Driving* includes natural landscapes and environmental conditions as factors worth considering in AI development to ensure robustness across a range of environments [50].

The majority (71%) of standards included core principles. Within this, 69% included a single principle for the benefit or respect of humans, whereas only 20% included concern for nonhumans within those core principles. This difference is charted in Figure 4.

We further examined whether these core principles included nonhumans for their own sake or for the sake of humans. Examples of standards that included principles relating to nonhumans for their own sake include Hochschule der Medien’s *10 Ethical Guidelines for the Digitalisation of Companies*. This standard includes the principle that “digitization should serve to conserve natural resources” [53]. The Future of Life Institute also includes the conservation of natural resources in its longer-term principles: “Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources” [58]. The Chinese National Governance Committee for the New Generation Artificial Intelligence, [59], IA-Latam [60], and the UNI Global Union [61] among others developed principles relating solely to the nonhuman world.

Others developed a principle relating to humans and nonhumans together. AI4Peoples’ *Ethical Framework for a Good AI Society* includes in the principle of beneficence, “promoting well being, preserving dignity, and sustaining the planet” [54] (p. 696). Similarly, the Alan Turing Institute promotes the protection of human individuals, “future generations”, and the “biosphere as a whole” [62] (p. 11). Meanwhile, the *Beijing AI Principles* “promote the sustainable development of nature and society, to benefit all mankind and the environment, and to enhance the well-being of society and ecology” [63]. The Machine Intelligence Garage Ethics Committee [64], Itchlaw [65], Tieto [66], and the European Parliament [67] were among others to state that AI should be developed for the benefit of both humans and the environment.

Some extended extrinsic concern to the environment only in so far as this would benefit humans. For instance, the IEEE states that their principle of wellbeing “encompasses the full spectrum of personal, social, and environmental factors that enhance human life and on which human life depend” [68] (p. 70). Meanwhile, the European Commission states, “AI technology must be in line with the human responsibility to ensure the preconditions for life

on our planet, continued prospering of mankind, and preservation of a good environment for future generations” [69] (p. 19).

Finally, some standards extended concern to humans alone. For instance, Vodafone states, “We will ensure that we respect international human rights standards” [70]. Similarly, the National Research Council of Canada states that it will “preserve human and legal rights” [71]. Meanwhile, the UK Department of Health and Social Care have two core principles relating to human wellbeing and respect: “respect for persons” and “respect for human rights” [72]. A full list of all standards with their respective core principle categories is available in Appendix A.2.

Overall, as can be seen in Figure 4, though nonhumans are considered in some AI ethics standards for a variety of reasons, humans are included in the concern far more often than nonhumans. Moreover, the standards that adopt core principles include a core principle relating to humans more often than a core principles relating to nonhumans. Appendix A.2 shows that, where nonhumans are considered in core principles, this is most often in relation to or as an extension to humans; few standards have core principles solely relating to nonhumans.

4.2. Human-Centredness

Figure 5 charts the data on human-centredness in AI ethics standards. We found that 27% of the ethics standards we surveyed explicitly supported human-centred approaches. ‘Human-centric’ and ‘human-centred’ were the most common examples of such language. Other examples included ‘benefiting first and foremost humans’; ‘bringing humans’ and ‘human rights to the centre’; and having a ‘primacy fiduciary duty to humanity’.

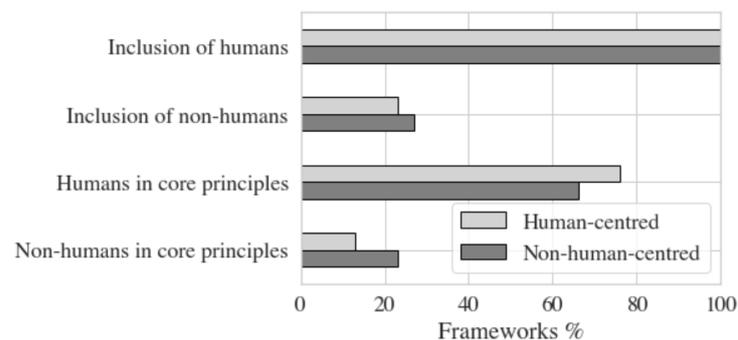


Figure 5. Human-centredness. This figure shows the percentage of human-centred and nonhuman-centred standards that include humans anywhere, include nonhumans anywhere, include humans in core principles, and include nonhumans in core principles. This graph expands on the data visualised in Figure 4, breaking this data further down by whether the standards are human-centred. So, for example, just under 80% of human-centred standards with core principles include humans within those core principles.

Much of this human-centredness is motivated by an interest in ensuring that AI benefits humans. Telefonica includes “human-centric AI” as a core principle to ensure AI benefits humanity [51] (p. 3). The Institute for Information and Communications Policy (IICP) include the pursuit of a “human-centred society” where all humans benefit from and live in harmony with AI [73] (p. 4). Meanwhile, Data Ethics’ principles look for “sustainable solutions benefitting first and foremost humans” [74] (p. 7).

Others look to human-centredness as a means of aligning AI with the values of humans. The IEEE states that AI should remain “human-centric” to serve “humanity’s values and ethical principles” [68] (p. 2). IBM also states that AI should be “human-centric” to align with humanity’s values [75] (p. 8). Meanwhile, the Chinese AI Alliance’s *Joint Pledge on Artificial Intelligence Industry Self-Discipline* states that AI should be human-oriented to uphold humanity’s values and prevent the replacement of humans [59].

Some standards have taken a ‘human rights lens’ or argued for a ‘human rights perspective’ of AI ethics. Examples of this included *White Paper: How to Prevent Discriminatory Outcomes in Machine Learning* [76] and *Privacy and Freedom of Expression In the Age of Artificial Intelligence* [77].

Human-centredness is also seen as essential for building trustworthy AI (TAI). The European Commission High-Level Expert group’s four ethical principles of TAI are: respect for human autonomy; prevention of harm; fairness; and explicability [30] (pp. 11–12). ‘Human-centredness’ is cited as the unifying feature of these ethical principles, necessary for AI systems to be developed for human benefit and having the goal of improving the welfare and freedom of humans [30] (pp. 9–19). This sentiment is echoed across a range of AI ethics standards. For instance, the *Personal Data Protection Commission* of Singapore states that “human-centricity, as clear baseline requirements can build consumer trust in AI deployments” [78] (p. 3). The Organisation for Economic Co-operation (OECD) also states, “there is a need for a stable policy environment that promotes a human-centric approach to trustworthy AI” [79]. Meanwhile, the *Telia Guided Principles on Trusted AI* includes “human centric” as a key principle [80] (p. 3). Fraunhofer IAIS also places humans “at the centre” of their *Trustworthy Use of Artificial Intelligence* framework [81] (pp. 4–5, 12–13).

Figure 5 draws an overview of the impact that this commitment to human-centredness has on concern for humans and nonhumans. Accounting for the fact that there are more human-centred than nonhuman-centred standards, the percentage of human-centred standards which *include nonhumans*, *include nonhumans in core principles*, and *included humans in core principles* was calculated. This was compared with the percentage of nonhuman-centred standards that made the same considerations. Figure 5 shows that a higher percentage of human-centred standards include humans and exclude nonhumans in their core principles compared with nonhuman-centred standards. Moreover, nonhuman-centred standards more often include nonhumans compared with human-centred ones.

To examine this further, Table 3 shows the human-centred standards that include core principles relating to humans and nonhumans. We broke this down using the criteria set out in Section 3.4.2. In this table, human-centred standards are checked as having a core principle relating to *nonhumans*, to *nonhumans and humans* together, to *nonhumans for the sake of humans*, or for *humans* alone. A single framework may have multiple individual core principles relating to humans and nonhumans. For example, the Alan Turing Institute published a framework with two individual core principles relating to humans and nonhumans.

Based on the results in Figures 3 and 5, we can see that human-centred standards tend to more often include a single core principle relating to humans only, excluding nonhumans. This would accord with the strong anthropocentric view (as defined in Section 2.3) that only humans matter morally. The IEEE’s *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems* (versions 1 and 2) includes core principles relating to humans alone and to nonhumans only for the sake of humans. This would comply with the strong anthropocentric view that nonhumans have at most extrinsic value in how they may benefit humans.

Other human-centric standards that include core principles relating to nonhumans and humans together are compatible with moderate anthropocentrism (as defined in Section 2.3). These include the Alan Turing Institute’s *Understanding AI Ethics and Safety* and OECD’s *Principles for Responsible Stewardship of Trustworthy AI*. Note that no human-centred standard includes core principles that solely relate to nonhumans.

Overall, 27% of the ethics standards that we explicitly surveyed supported human-centred AI ethics approaches. Human-centred standards tend to include humans and exclude nonhumans in their core principles more than nonhuman-centred standards. In contrast, nonhuman-centred standards include nonhumans more often than human-centred standards do. Most human-centred AI ethics standards comply with the definition of strong anthropocentrism, including nonhumans only insofar as they benefit humans. Moreover, no human-centred standard included core principles solely relating to nonhumans.

Table 3. Core principles of human-centric standards. * is per core principle. Therefore, ** in a column means two individual core principles relating to that column.

Framework	Nonhumans	Nonhumans and Humans	Nonhumans for Humans	Humans
A Framework for the Ethical use of Advanced Data Science Methods in the Humanitarian Sector				*
AI Principles of Telefónica				*
Alan Turing Institute Understanding AI Ethics and Safety		**		
Charlevoix Common Vision for the Future of Artificial Intelligence				*
Data Ethics Principles				*
Discussion Paper on Artificial Intelligence (AI) and Personal Data—Fostering Responsible Development and Adoption of AI				*
Draft AI R&D Guidelines for International Discussions				*
Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems, First Edition			*	*
Ethically Aligned Design. A Vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems, version 2			*	*
Ethics Guidelines for Trustworthy Artificial Intelligence (AI)		*		*
Everyday Ethics for Artificial Intelligence. A Practical Guide for Designers & Developers				*
For a Meaningful Artificial Intelligence. Towards a French and European Strategy				*
Human Rights in the Robot Age Report				*
Joint Pledge on Artificial Intelligence Industry Self-discipline				*
OP Financial Group’s Ethical Guidelines for Artificial Intelligence				*
OpenAI Charter				*
Position on Robotics and Artificial Intelligence				*
Principles for Responsible Stewardship of Trustworthy AI		*		*
SAP’s Guiding Principles for Artificial Intelligence				*
Telia Guided Principles on Trusted AI				*
Toward a G20 Framework for Artificial Intelligence in the Workplace				*
Vodafone AI Framework				*

4.3. Tensions between Human-Centredness and Environmental Wellbeing

In order to further ascertain whether human-centredness in AI ethics actually accords with the exclusion of nonhumans, trends in the standards were further analysed over time and by domain.

Figure 3b shows the standards included in this survey that followed a left-skewed distribution in the time that they were published, peaking in 2018. In accounting for the fact that there was not an equal count of standards per year, the various features of the standards (human-centredness, inclusion of humans and nonhumans, both within and outside of core principles) were divided by the number of occurrences per year. This is captured in Figure 6. Because the mention of humans is included in all standards and is thereby static, this was excluded. Figure 6 shows that from 2016 to 2020, which is when most of the

ethics standards are drawn from, all three features (inclusion of nonhumans and inclusion of both humans and nonhumans within core principles) increase and decrease with each other, implying that they accord with each other. Between 2011 and 2016, and from 2020 onwards, there are fewer standards, and so it is less reliable to perform a concrete analysis of trends. Overall, the first finding was that the mentioning of nonhumans and inclusion of nonhumans in core principles closely follow each other in trends over time. The inclusion of humans in core principles also followed this same trend over time, albeit less so.

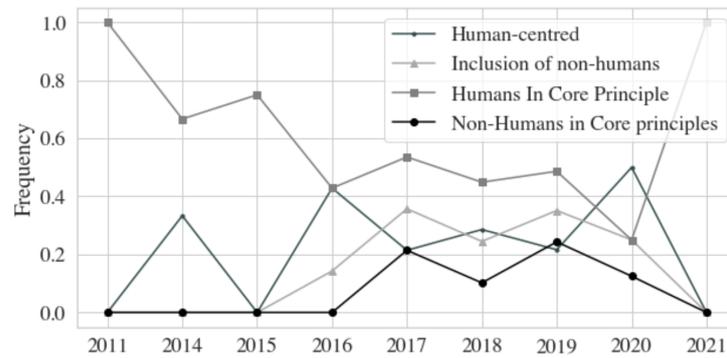


Figure 6. Frequency of trends in standards over time.

The second finding from this analysis was that human-centredness diverges away from concern for nonhumans. Figure 6 shows that, from the same time period of 2016 to 2020, human-centredness as shown with small circle line markings increases where the other lines decrease and vice versa. Over time, mention of nonhumans and the inclusion of both nonhumans and humans in core principles follow the same pattern and seem to accord with each other, whereas human-centredness diverges from this pattern.

To further determine the effects of human-centredness upon extensions of concern, we also analysed the standards by the domains within which these are published. Again, in order to account for the fact that the number of standards published across domains was not equal, the various features of the standards were divided by the number of occurrences per domain. These data are visualised in Figure 7. Because the mention of humans is included in all standards and is therefore both static and equivalent to the number of instances per domain, this feature was again excluded. To improve clarity of the visualisation, the domains having only one or two standards each, *religious group*, *research council*, and *Political party*, were excluded.

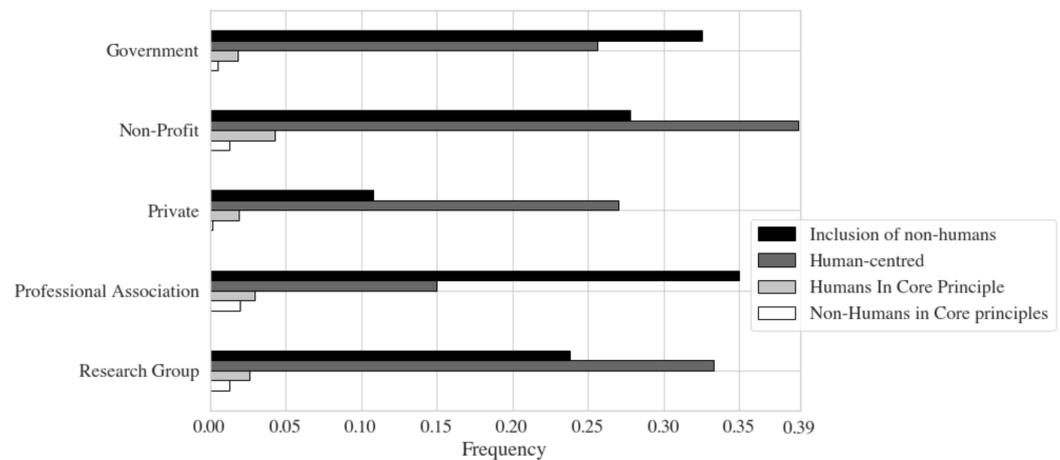


Figure 7. Frequency of standards across domains according to trends.

From Figure 7, we can see that across domains, increases in human-centredness accords with decreases in the inclusion of nonhumans in core principles. Private companies, non-profits, and research groups support human-centredness the most. These domains also mention nonhumans the least. Private companies, in particular, include nonhumans the least and include nonhumans in their core principles the least of all domains. Professional associations and governments support human-centredness the least. Both of these domains also mention nonhumans the most. Professional associations, in particular, support human-centredness the least, mention nonhumans the most, and include nonhumans in their core principles the most out of all domains. Overall, across domains, support for human-centredness tends to relate to lack of concern for nonhumans, whereas lack of support for human-centredness tends to relate to concern for nonhumans.

4.4. Findings Summary

The following key findings of our scoping review of the AI ethics landscape were noted as follows:

- The entire AI ethics landscape includes humans in concern more often than nonhumans and includes more core principles related to humans than nonhumans;
- Wherever nonhumans are included within core principles, this is most often as an extension or in relation to humans; few standards have a core principle relating solely to nonhumans;
- A total of 27% of standards support human-centredness, most of which comply with strong anthropocentrism as defined in Section 2.3;
- The standards that support human-centredness tend to include humans and exclude nonhumans more than nonhuman-centred standards.

Overall, our findings indicate that human-centredness in AI ethics standards tends to correlate with increased concern for human wellbeing and decreased concern for nonhuman wellbeing.

5. Discussion

In Section 2, we discussed the support for human-centredness across the AI landscape and how this differs in the definition and application from security measures, limits of bias, or inequalities to being portrayed as ‘ethical’ or ‘beneficial’ for humans. We discussed this HCAI as inspired by, or perhaps accepting of, Asimov’s laws of robotics, particularly the fourth law: a robot may not harm humanity or, by inaction, allow humanity to come to harm. Overall, we argued that applications and concepts of ‘human-centredness’ across AI were nebulous, without a clear, consistent definition, and lacking investigation into their normative groundings, merits, and limitations. Through our review, we aimed to examine the commitment of AI ethics standards to human centredness and how this commitment interacts with other ethical considerations for nonhumans and the environment. We asked: how prevalent is a commitment to human-centredness in AI ethics standards, and how does this relate to commitments to nonhuman animals and the environment?

Our findings showed that 27% of AI ethics standards explicated such a commitment to human-centredness. In contrast, 69% of AI ethics standards employ core principles relating to human wellbeing, dignity, respect, or similar phrasing that closely resemble Asimov’s laws, particularly the fourth law. “We will ensure that we respect international human rights standards” [70], “preserve human and legal rights” [71], “respect for persons”, and “respect for human rights” [72] are some examples. This indicates a strong adherence to Asimovian human-centredness and shows a clear gap between implicitly prioritising human wellbeing and explicating a commitment to human-centredness.

In Section 2.3, we also discussed anthropocentrism, defined in moral philosophy as the claim that humans have, at the very least, more intrinsic moral value than nonhumans, or, more strongly, that nonhumans have no intrinsic moral value. In our scoping review, we compared the commitment to humans and nonhumans across AI ethics standards. We found that overall, AI ethics standards include humans in concern more than nonhumans.

Moreover, where standards included core principles, these were more often related to human wellbeing than to nonhuman wellbeing. We found that very few standards included core principles that solely related to nonhumans.

Human-centred frameworks were defined somewhat broadly; that is, they explicated a commitment to ‘human-centredness’. This means that the frameworks we defined as human-centred would not necessarily commit themselves to the definition of anthropocentrism as used in moral philosophy. As our background discussion showed, the use of the term ‘human-centred’ could relate to safety, bias, inclusion, or a range of other commitments and not necessarily to the claim that humans are more morally valuable than nonhumans. However, our findings show that AI ethics standards that explicate a commitment to human-centredness in whatever regard also tend to exclude nonhumans more often and include humans within concern.

Thus, even when human-centredness may not be explicitly used in the way that anthropocentrism is defined in moral philosophy, the implications of a human-centred AI are that humans are prioritised over nonhumans as in philosophical anthropocentrism. The use of human-centredness on the surface may seem to include a range of possible meanings and applications, not necessarily indicating a moral preference or prioritisation of humans over nonhumans but rather referencing a design preference or broader social justice. However, when interrogated and examined, as we have performed in this review, there comes to light an undercurrent that promotes, or at the very least, accords more strongly with, the moral hierarchies of humans, nonhumans, and the environment, which has been established in antiquity and carried through centuries of moral philosophical debate.

Overall, a commitment to human-centredness in AI may seem a nebulous, vague, or high-level principle that can be used to unite or find consensus among different bodies, stakeholders, ethical norms, and interests. Anthropocentrism is a well established concept used throughout moral philosophy with a specific technical definition and implications. However, when investigated, the seeming gap between human-centredness and anthropocentrism narrows, and the application of this high-level unifying concept is actually exposed to come with real moral implications in terms of who or what stands as objects of moral concern in the development and deployment of AI. Upon analysing our findings within the context of HCAI, Asimov’s laws, and anthropocentrism in moral philosophy, we suggest that human-centredness in AI ethics standards accords with anthropocentrism in the moral prioritisation of humans over nonhumans.

5.1. Anthropocentrism and Environmental Wellbeing in AI Development and Deployment

The reader might reject the notion that AI ethics ought to extend intrinsic moral concern to nonhumans. This is because, one might argue, AI is created and used by humans for human specific purposes; nonhumans are not significant stakeholders that are worthy of the moral consideration of AI makers and users.

However, there is a range of AI systems for which nonhumans are significant stakeholders. There are AI systems that interact with the environment by traversing landscapes, monitoring wildlife, collecting data, and interacting with ecosystems and individuals for commercial, research, or conservation purposes [82]. Other systems interact more directly with ecosystems through pest control, managing predator populations (i.e., killing certain overpowering species members), and chemically stabilising abiotic environments [82]. Research shows that deploying embodied AI systems into established ecosystems can cause distress to certain species [82–84]. In cases of malfunctioning, AI systems can crash, sometimes in delicate ecosystems [82,85,86]. When wildlife are disturbed in this way and either flee or, in extreme cases, are killed, changes in species populations alter the entire structure of the ecosystem, oftentimes undermining its stability for generations to come [85,86].

Some AI systems interact with the environment on a more permanent basis as they are designed and built into an existing ecosystem. Examples include artificially intelligent biosystems [87], intelligent greenhouses [88], or, more rudimentarily, autonomous power stations [89]. Again, ethical questions arise, such as whether it is right to introduce inorganic

engineered elements into a natural environment [82]. If such systems are designed to maintain an ecosystem that would otherwise destabilise, there may also be a sense in which it would be wrong to remove the system from the environment [82].

There is also growing concern over AI's environmental impact as estimates of the substantial energy required to run, as well as the emissions caused by, machine learning models have been released. At the sharp end of machine learning, models can contain billions or trillions of parameters and take months to train, carrying a significant environmental cost. For example, Chat GPT-3, with 175 billion parameters, was estimated to train on 1024 GPUs for 34 days at a cost of USD 4.6 million and an expected energy consumption of 936 MWh [90,91]. For comparison, the average UK household uses between 8.5 kWh and 10 kWh of energy per day [92]. Chat GPT-4 (released March 2023) has even more parameters and is expected to have had an even greater environmental impact [90]. As well as requiring significant amounts of energy for training, machine learning models can lead to the production of large amounts of emissions [93]. For example, the BERTlarge model was estimated to produce as much greenhouse gas emissions as a commercial flight between San Francisco and New York [93,94]. Others have highlighted the environmental impact of an entire AI system's lifecycle, from extracting the necessary materials for hardware to waste production and recycling [95].

One might argue that anthropocentrism does not necessarily stand in opposition to moral concern for the environment. Moderate anthropocentrists maintain that nonhumans have intrinsic value, albeit less than humans. Moreover, caring for the environment and avoiding environmental harm need not rely on nonhumans having intrinsic value. As discussed, some strong anthropocentrists extend extrinsic moral concern to nonhumans for the sake of human wellbeing, survival, or flourishing.

However, under strong anthropocentrism, the harm to nonhuman animals, plants, and ecosystems that come with certain AI systems are only morally significant insofar as humans may be in turn negatively affected. Where humans are not expected to be negatively affected, or where the benefit to humans is significant, these environmental harms are not of moral concern. Harm to the environment may also be permissible under moderate anthropocentrism in cases where a significant benefit to humans outweighs the harm caused to nonhumans. Humans stand to greatly benefit from the development and deployment of AI systems, such as those wherein the environment stands as a significant stakeholder: unmanned delivery vehicles can offer safe delivery of essential materials in areas with low to no infrastructure; data collection systems can collect data from areas that are dangerous or inaccessible for humans; machine learning models can provide decision support for medical professionals, AI-generated code and fixes for programmers, and content for creators; not to mention the profits to shareholders from increased deployment.

In the case of AI systems that interact with the environment, such as unmanned vehicles and robots, anthropocentrism may permit the disturbance of wildlife, altering the population of species within an environment and undermining the stability of ecosystems over generations. As for AI systems that are built into an environment, landscapes may be irreversibly altered, with the future of ecosystems thereafter being dependent on the input of inorganic optimisation systems. An anthropocentric AI ethic may also permit devastating energy consumption and emissions for the advancement of machine learning models.

Strong anthropocentrists who either consider the environment as only extrinsically valuable to humans ought to be concerned about this, as humans will in turn be affected by damage to the environment. Meanwhile, moderate anthropocentrists and non-anthropocentrists ought to also be concerned for the sake of the environment. Therefore, whether the reader supports or opposes anthropocentrism, an anthropocentric AI ethic poses serious risks that are worth consideration.

5.2. Alternative Approaches

Environmental ethics supplies a range of well-established schools of non-anthropocentric ethics that could ground future work on AI ethics standards. Within philosophy, "the idea

of drawing on environmental ethics to address the moral problem of artificial intelligence is not new" [96] (p. 330). AI ethics approaches that are grounded in non-anthropocentric environmental ethics would not permit such harm to nonhumans and would avoid the risks identified in this work.

Recall, for instance, sentiocentrism, defined in Section 2.4 as viewing anything that suffers to possess intrinsic value [32]. A sentiocentric AI ethic may extend moral concern to a wider range of nonhuman animals or study our relations to AI systems in terms of our relations with humans or other sentient animals [97,98]. In terms of the examples discussed above, sentiocentric AI ethics standards may ban the use of rovers or unmanned vehicles that harm sentient wildlife.

Recall also a biocentric or 'life-centred' environmental ethic, which extends moral concern to the living environment and which has also been discussed as an alternative environmental foundation for AI ethics [99,100]. Under biocentrism, our moral duties in the development and deployment of AI systems would extend to all living individuals and systems [99]. In policy, this would force designers of environmentally interactive embodied systems to not only consider the individual harm caused to humans or sentient animals but also to the wider living community of plants and non-sentient animals that may in turn be affected. One example discussed is when AI systems are built into existing ecosystems to optimise environmental conditions or 'parameters'. In such a case, biocentric AI ethics would demand that the system be designed in such a way as to become a stable, contributive ecosystem member. A biocentric AI ethic may also ban the designing of AI systems into ecosystems to avoid the future dependencies of the living environment on artificial optimisation systems.

Finally, recall ecocentrism, which considers all interconnected parts of the environment, including the ecosystems themselves, to possess intrinsic value. In the examples discussed, an ecocentric AI ethic may ban AI systems that undermine ecosystem stability, such as unmanned vehicles that scare away animals, or force designers to consider renewable and sustainable materials for embodied system hardware.

Above are just some examples of how non-anthropocentric environmental ethics can avoid the limitations of anthropocentric AI ethics in considering nonhumans and how such ethics can inform approaches to the moral questions posed by AI systems both in AI ethics regulation and future research.

6. Conclusions

This work set out to examine human-centredness in AI ethics standards. We first introduced the topic of human-centredness in AI more broadly, outlining the range of possible uses, the vagueness of this term, and the lack of due investigation into its meaning and implications. We undertook a scoping review of 146 AI ethics standards, reviewing the commitment to human and/or nonhuman wellbeing as well as human-centredness. We then critically analysed our results within the wider context of HCAI, Asimovian laws, and anthropocentrism in moral philosophy. Overall, we found that the application of what may seem to be a high-level unifying concept actually comes with real moral implications in terms of who or what stands as objects of moral concern in the development and deployment of AI. In particular, human-centredness in AI ethics promotes, or at the very least accords more strongly with, a prioritisation of humans over nonhumans and the environment. This accords with anthropocentrism, as defined in moral philosophy. In fact, we found that most human-centred AI ethics standards complied with strong anthropocentrism: the claim that nonhumans have at most extrinsic moral value to humans.

We have also briefly discussed some of the ways in which anthropocentric approaches to AI ethics could permit harm to nonhuman animals and the environment, including the disturbance of animals, undermining the stability of ecosystems, alteration of landscapes, and devastating energy consumption and emissions. We recommend that the future use and application of 'human-centredness' in AI development and deployment be technically explicated, which should include sufficient discussion of the normative groundings and

moral implications of theory and practice. Without this, the development and application of HCAI will continue without proper investigation of the potential environmental impact; namely, failing to consider in sufficient depth the impacts of AI on nonhumans and their moral standing.

Some alternative approaches to anthropocentric AI ethics grounded in environmental philosophy were also considered; namely, sentiocentrism, biocentrism, and ecocentrism. We have shown the ways in which alternative groundings to AI ethics may avoid some of the limitations of anthropocentrism by extending moral concern to the wider nonhuman world in cases of AI systems that interact, are built in, and impact the environment. The motivation to design future AI systems in accordance with non-anthropocentric environmental ethics provides opportunities for future research across a range of disciplines, including philosophy, political science, social robotics, and computer science. Future environmental AI ethics policy development would benefit from these disciplines working together to widen the scope of who and what counts as deserving of moral consideration and protection.

Author Contributions: Conceptualisation, E.R., A.C., C.E. and W.M.; methodology, E.R. and A.C.; formal analysis and investigation, E.R.; writing—review and editing, E.R., A.C., C.E. and W.M.; supervision, A.C., C.E. and W.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded in part by UKRI Trusted Autonomous Systems Hub (EP/V00784X/1) and NIHR Southampton Biomedical Research centre (IS-BRC-1215-20004). The authors would also like to acknowledge support from the Defence and Security Programme at the Alan Turing Institute, funded by the UK Government.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data used in this paper is available upon request from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Full Lists

Appendix A.1. Full List of Ethical Standards Surveyed

Title	Issuer	Year
10 Ethical Guidelines for the Digitalisation of Companies	Hochschule der Medien	2017
10 Principles of Responsible AI	Women Leading in AI	2019
A Code of Ethics for the Human Robot Interaction	Riek, Howard	2014
A Framework for Responsible Limits on Facial Recognition Use-case: Flow Management	WEForum	2020
A Framework for the Ethical Use of Advanced Data Science Methods in the Humanitarian Sector	International Organization for Migration (IOM) Data Science Initiative	2020
A Guide for Professionals of the Digital Age	Cigref	2018
A Typological Framework for Data Marginalization	United Nations University Institute	2019
Advisory Statement on Human Ethics in Artificial Intelligence and Big Data Research	National Research Council Canada	2019
AI—Our approach	Microsoft	2017

Title	Issuer	Year
AI & Data Topical Guide Series 1—Introducing the Series: Can AI and Data Support a More Inclusive and Equitable South Africa?	Policy Action Network	2020
AI in the UK: Ready, Willing and Able?	UK House of Lords, Select Committee on Artificial Intelligence	2018
AI Now 2017 Report	AI Now Institute	2017
AI Now 2018 Report	AI Now Institute	2018
AI Now 2019 Report	AI Now Institute	2019
AI Principles	Future of Life Institute	2017
AI Principles & Ethics	Smart Dubai	2019
AI Principles of Telefónica	Telefonica	2018
AI UX: 7 Principles of Designing Good AI Products	UX Studio	2018
AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations	AI4People	2018
Alan Turing Institute Understanding AI Ethics and Safety	Alan Turing Institute	2019
Algo.Rules	iRights.Lab	2019
Artificial Intelligence and Data Protection	European Council	2018
Artificial Intelligence (AI) in Health	Royal College of Physicians	2018
Artificial Intelligence and Machine Learning: Policy Paper	The Internet Society	2017
Artificial intelligence and Privacy	The Norwegian Data Protection Authority	2018
Artificial Intelligence in Healthcare	Academy of Medical Royal Colleges	2019
Artificial Intelligence: Open Questions about Gender Inclusion	W20	2018
Artificial Intelligence: Opportunities, Risks and Recommendations for the Financial Sector	Commission de Surveillance du Secteur Financier	2018
Artificial Intelligence. Australia's Ethics Framework. A discussion Paper	Commonwealth Scientific and Industrial Research Organisation	2019
Artificial Intelligence. The Public Policy Opportunity	Intel Corporation	2017
Automated and Connected Driving: Report	Federal Minister of Transport and Digital Infrastructure	2017
Beijing AI Principles	Beijing Academy of Artificial Intelligence	2019
Big Data, Artificial Intelligence, Machine Learning and Data Protection	Information Commissioner's Office	2017
Business Ethics and Artificial Intelligence	Institute of Business Ethics	2018
Charlevoix Common Vision for the Future of Artificial Intelligence	Leaders of the G7	2018
Charter of Digital Networking (English translation)	Working group Vernetzte Anwendungen und Plattformen für die digitale Gesellschaft	2014
Civil Rights Principles for the Era of Big Data	The Leadership Conference on Civil and Human Rights	2015
Code of Practice for Disinformation	European Commission	2018
Commitment	Verivox	2019
Data Ethics Canvas	The Open Data Institute	2019
Data Ethics principles	DataEthics.eu	2017
Data for the Benefit of the People: Recommendations from the Danish Expert Group on Data Ethics	DATAETIK (Danish Expert Group on Data Ethics)	2018
Declaration on Ethics and Data Protection in Artificial Intelligence	ICDPPC	2018

Title	Issuer	Year
DeepMind Ethics & Society Principles	DeepMind	2017
Deutsche Telekom AI Guidelines	Deutsche Telekom	2018
Digital Decisions	Centre for Democracy & Technology	2015
Digital Technology and Healthcare. Which Ethical Issues for which Regulations?	French National Ethical Consultative Committee for Life Sciences and Health (CCNE)	2014
Directive on Automated Decision-Making	Government of Canada	2019
Discussion Paper on Artificial Intelligence (AI) and Personal Data—Fostering Responsible Development and Adoption of AI	Personal Data Protection Commission Singapore	2018
Draft AI R&D Guidelines for International Discussions	Institute for Information and Communications Policy (IICP)	2017
Dutch Artificial Intelligence Manifesto	Special Interest Group on Artificial Intelligence	2018
Effective Ad Archives	Mozilla Foundation	2019
Ethical Codex for Data-Based Value Creation: For Public Consultation	Swiss Alliance for Data-Intensive Services	2019
Ethical Guidelines of the German Informatics Society	Gesellschaft für Informatik	2018
Ethical, Social, and Political Challenges of Artificial Intelligence in Health	Future Advocacy	2019
Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition (EAD1e)	Institute of Electrical and Electronics Engineers (IEEE)	2019
Ethically Aligned Design. A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2	Institute of Electrical and Electronics Engineers (IEEE)	2017
Ethics Framework—Responsible AI	Machine Intelligence Garage Ethics Committee	2018
Ethics Guidelines for Trustworthy AI	High-Level Expert Group on Artificial Intelligence	2019
Ethics of AI in Radiology: European and North American Multi-society Statement	American College of Radiology et al.	2019
Ethics Policy	Icelandic Institute for Intelligent Machines (IIIM)	2015
European Ethical Charter on the use of Artificial Intelligence in Judicial Systems and their Environment	European Commission	2019
Everyday Ethics for Artificial Intelligence. A practical guide for Designers & Developers	IBM	2018
Facial Recognition Principles	Microsoft	2018
Five Guiding Principles for Responsible use of AI in Healthcare and Healthy Living	Philips	2020
For a Meaningful Artificial Intelligence. Towards a French and European Strategy	Internet Society	2017
Google People & AI Partnership Guidebook	Google	n.d.
Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence	National Governance Committee for the New Generation Artificial Intelligence	2019
Governing Artificial Intelligence. Upholding Human Rights & Dignity	Data & Society	2018
Guidance for Regulation of Artificial Intelligence Applications	The White House	2020
Guidance on AI and Data Protection	ICO	2020
Hippocratic Oath for Data Scientists	DataForGood	2019
How Can Humans Keep the Upper Hand? Report on the Ethical Matters Raised by AI Algorithms	French Data Protection Authority (CNIL)	2017

Title	Issuer	Year
Human Rights in the Robot Age Report	The Rathenau Institute	2017
IA-Latam Ethics Statement for the Design, Development and Use of Artificial Intelligence	IA-Latam	2019
IBM's Principles for Trust and Transparency	IBM	2018
Initial Code of Conduct for Data-driven Health and Care Technology	UK Department of Health & Social Care	2018
Intel's AI Privacy Policy White Paper. Protecting individuals' Privacy and Data in the Artificial Intelligence World	Intel Corporation	2018
Introducing Unity's Guiding Principles for Ethical AI—Unity Blog	Unity Technologies	2018
It's Time to Do Something: Mitigating the Negative Impacts of Computing Through a Change to the Peer Review Process	Hecht et al.	2018
ITI AI Policy Principles	Information Technology Industry Council (ITI)	2017
Joint pledge on artificial intelligence industry self-discipline	Artificial Intelligence Industry Alliance	2019
Kakao Algorithm Ethics	Kakao	n.d.
Machine learning: the Power and Promise of Computers that Learn by Example	The Royal Society	2017
Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society	Government of the Republic of Korea	2017
MIT Schwarzman College of Computing Task Force Working Group on Social Implications and Responsibilities of Computing Final Report	MIT Schwarzman College of Computing Task Force Working Group on Social Implications and Responsibilities of Computing Final Report	2019
Montréal Declaration: Responsible AI	Université de Montréal	2017
OP Financial Group's Ethical Guidelines for Artificial Intelligence	OP Finland	n.d.
OpenAI Charter	Open AI	2018
Our Principles	Google	2018
Oxford Munich Code of Conduct	Oxford Munich	2019
Policy Recommendations on Augmented Intelligence in Health Care H-480.940	American Medical Association	2018
Position on Robotics and Artificial Intelligence	The Greens (Green Working Group Robotics)	2016
Preliminary Study on the Ethics of Artificial Intelligence	Unesco	2019
Preparing for the Future of Artificial Intelligence	Executive Office of the President; National Science and Technology Council; Committee on Technology	2016
Principles for Accountable Algorithms and a Social Impact Statement for Algorithms	Fairness, Accountability, and Transparency in Machine Learning (FATML)	2016
Principles for Responsible Stewardship of Trustworthy AI	OECD (Organisation for Economic Co-operation)	2019
Principles for the Safe and Effective use of Data and Analytics	New Zealand Privacy Commissioner and the Government Chief Data Steward	2018
Principles of robotics	Royal College of Physicians	2011
Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector	Monetary Authority of Singapore	2018
Privacy and Freedom of Expression In the Age of Artificial Intelligence	Privacy International	2018
Report of COMEST on Robotics Ethics	COMEST/UNESCO	2017
Report on Artificial Intelligence and Human Society (Unofficial translation)	Advisory Board on Artificial Intelligence and Human Society	2017

Title	Issuer	Year
Report with Recommendations to the Commission on Civil Law Rules on Robotics	European Parliament	2017
Responsible AI #AIFORALL Approach Document For India Part 2—Operationalizing Principles For Responsible AI	National Institute for Transforming India	2021
Responsible AI #AIFORALL Approach Document for India Part 1—Principles for Responsible AI	National Institute for Transforming India	2021
Responsible AI and Robotics. An ethical framework	Accenture UK	2018
Responsible AI Practice	Google	
Responsible AI: Global Policy Framework	iTechLaw	2019
Responsible bots: 10 Guidelines for Developers of Conversational AI	Microsoft	2018
Responsible use of Artificial Intelligence (AI)	Government of Canada	2019
Rome Call for AI Ethics	Rome Call	2020
Safety First for Automated Driving	Aptive, Audit, BMW, FCA, Continental, Daimler, VW, Intel, Infineon, Baidu, Here	2019
SAP's Guiding Principles for Artificial Intelligence	SAP	2018
Seeking Ground Rules for AI	New York Times	2019
Sony Group AI Ethics Guidelines	Sony group	2018
Statement on Algorithmic Transparency and Accountability	Association for Computing Machinery (ACM)	2017
Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems	European Commission	2018
Telia Guided Principles on Trusted AI	Telia	n.d.
Tenets	Partnership on AI	2016
The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term	AI Now Institute	2016
The Critical Engineering Manifesto	The Critical Engineering Working Group	2019
The Ethics of Code: Developing AI for Business with Five Core Principles	Sage	2017
The Japanese Society for Artificial Intelligence Ethical Guidelines	The Japanese Society for Artificial Intelligence	2017
The Future Computed—Artificial Intelligence and Its Role in Society	Microsoft	2018
The Good Technology Standard (GTS:2019-Draft-1)	The Good Technology Collective	2018
The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation	Future of Humanity Institute et al.	2018
The National Artificial Intelligence Research and Development Strategic Plan	National Science and Technology Council; Networking and Information Technology Research and Development Subcommittee	2016
The Responsible AI framework	PriceWaterhouseCoopers UK	n.d.
The Responsible Machine Learning Principles	The Institute for Ethical and Machine Learning	n.d.
The Toronto Declaration: Protecting the Right to Equality and Nondiscrimination in Machine Learning Systems	Access Now; Amnesty International	2018
Tieto's AI Ethics Guidelines	Tieto	2018
Top 10 Principles for Ethical Artificial Intelligence	UNI Global Union	2017
Toward a G20 Framework for Artificial Intelligence in the Workplace	CIGI Centre for International Governance Innovation	2018

Title	Issuer	Year
Trustworthy use of Artificial Intelligence	Fraunhofer IAIS	2020
Unfairness by Algorithm: Distilling the Harms of Automated Decision-Making	Future of Privacy Forum	2017
Unified Ethical Frame for Big Data Analysis. IAF Big Data Ethics Initiative, Part A	The Information Accountability Foundation	2015
Universal Guidelines for Artificial Intelligence	The Public Voice	2018
Universal Principles of Data Ethics	Accenture	2016
Užupis Principles for Trustworthy AI Design	Republic of Užupis	2019
Vienna Manifesto on Digital Humanism	Faculty of Informatics, TU Wien	2019
Vodafone AI Framework	Vodafone	2019
White Paper: How to Prevent Discriminatory Outcomes in Machine Learning	WEF, Global Future Council on Human Rights 2016–2018	2018
Work in the Age of Artificial Intelligence. Four Perspectives on the Economy, Employment, Skills and Ethics	Ministry of Economic Affairs and Employment	2018

*Appendix A.2. Full List of Core Principle Categories. * is per core Principle. Therefore, ** in a Column Means Two Individual Core Principles Relating to That Column*

Framework	Nonhumans	Nonhumans and Humans	Nonhumans for Humans	Humans
10 ethische Leitlinien für die Digitalisierung von Unternehmen (10 Ethical Guidelines for the Digitalisation of Companies)	*			
A Code of Ethics for the Human Robot Interaction				*
A Framework for the Ethical use of Advanced Data Science Methods in the Humanitarian Sector				*
A Guide to Good Practice for Digital and Data-driven Health Technologies				*
A Typological Framework for Data Marginalization				*
AI in the UK: Ready, Willing and Able?				*
Future of Life Institute AI Principles	*			*
Smart Dubai AI Principles & Ethics				*
AI Principles of Telefónica				*
AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations		*		*
Alan Turing Institute Understanding AI Ethics and Safety		**		
Artificial Intelligence, Australia's Ethics Framework: A Discussion Paper				*
Artificial Intelligence. The Public Policy Opportunity				*
Automated and Connected Driving: Report				*
Beijing AI principles		*		*

Framework	Nonhumans	Nonhumans and Humans	Nonhumans for Humans	Humans
Charlevoix Common Vision for the Future of Artificial Intelligence				*
Charter of Digital Networking (English translation)				*
Civil Rights Principles for the Era of Big Data				*
Data Ethics Canvas				*
Data Ethics principles				*
Data for the Benefit of the People: Recommendations from the Danish Expert Group on Data Ethics				*
Declaration on Ethics and Data Protection in Artificial Intelligence				*
Discussion Paper on Artificial Intelligence (AI) and Personal Data—Fostering Responsible Development and Adoption of AI				*
Draft AI R&D Guidelines for International Discussions				*
Ethical Codex for Data-Based Value Creation: For Public Consultation		**		
Ethical Guidelines of the German Informatics Society				*
Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems, First Edition			*	*
Ethically Aligned Design. A Vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems, Version 2		*		*
Ethics Framework—Responsible AI		**		
Ethics Guidelines for Trustworthy Artificial Intelligence (AI)		*		*
Icelandic Institute for Intelligent Machines (IIIM) Ethics Policy				*
European Ethical Charter on the use of Artificial Intelligence in Judicial Systems and their Environment				*
Everyday Ethics for Artificial Intelligence. A practical Guide for Designers & Developers				*
Five Guiding Principles for Responsible use of AI in Healthcare and Healthy Living				*
For a Meaningful Artificial Intelligence. Towards a French and European Strategy				*
Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence		*		*
Guidance for Regulation of Artificial Intelligence Applications		*		
Human Rights in the Robot Age Report				*
IA-Latam Ethics Statement for the Design, Development and use of Artificial Intelligence	*			*
IBM's Principles for Trust and Transparency				*
Introducing Unity's Guiding Principles for Ethical AI – Unity Blog				*
ITI AI Policy Principles				*
Joint Pledge on Artificial Intelligence Industry Self-discipline				*
Kakao Algorithm Ethics				*
Montréal Declaration: Responsible AI	*			*
OP Financial Group's Ethical Guidelines for Artificial Intelligence				*
OpenAI Charter				*

Framework	Nonhumans	Nonhumans and Humans	Nonhumans for Humans	Humans
Google: Our Principles				*
Position on Robotics and Artificial Intelligence				*
Preliminary Study on the Ethics of Artificial Intelligence	*			*
Principles for Responsible Stewardship of Trustworthy AI		*		*
Principles for the Governance of AI				*
Principles for the Safe and Effective use of Data and Analytics				*
Principles of Robotics				*
Report of COMEST on Robotics Ethics				*
Report with Recommendations to the Commission on Civil Law Rules on Robotics		*		*
Responsible AI #AIFORALL Approach Document for India Part 1—Principles for Responsible AI				*
Responsible AI #AIFORALL Approach Document For India Part 2—Operationalizing Principles For Responsible AI				*
Responsible AI: Global Policy Framework	*			*
SAP's Guiding Principles for Artificial Intelligence				*
Sony Group AI Ethics Guidelines				*
Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems			*	*
Telia Guided Principles on Trusted AI				*
Partnership on AI Tenets				*
The Good Technology Standard (GTS:2019-Draft-1)	*			*
The Japanese Society for Artificial Intelligence Ethical Guidelines				*
The Responsible AI framework				*
Tieto's AI ethics guidelines		*		*
Top 10 Principles for Ethical Artificial Intelligence		*		*
Toward a G20 Framework for Artificial Intelligence in the Workplace				*
Unified Ethical Frame for Big Data Analysis. IAF Big Data Ethics Initiative, Part A				*
Universal Principles of Data Ethics				*
Vodafone AI Framework				*
Work in the Age of Artificial Intelligence. Four Perspectives on the Economy, Employment, Skills and Ethics				*

References

1. Crawford, K.; Dryer, T.; Fried, G.; Green, B.; Kazianus, E.; Kak, A.; Mathur, V.; McElroy, E.; Sanchez, A.N.; Raji, D.; et al. AI Now 2019 Report AI Now Institute. 2019. Available online: <https://ainowinstitute.org/publication/ai-now-2019-report-2> (accessed on 7 May 2022).
2. Zeng, Y.; Lu, E.; Cunqing, H. Linking Artificial Intelligence Principles. *arXiv* **2019**, arXiv:1812.04814.
3. Wong, S. Fluxus Landscape: An Expansive View of AI Ethics and Governance. 2019. Available online: <https://icarus.kumu.io/fluxus-landscape> (accessed on 17 May 2022).
4. Thilo, H. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds Mach.* **2020**, *30*, 99–120. [CrossRef]
5. Jobin, A.; Ienca, M.; Vayena, E. The global landscape of AI ethics guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399. [CrossRef]

6. Fjeld, J.; Achten, N.; Hilligoss, H.; Nagy, A.; Srikumar, M. *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*; Berkman Klein Center Research: Cambridge, MA, USA, 2020. [CrossRef]
7. Ayling, J.; Chapman, A. Putting AI ethics to work: Are the tools fit for purpose? *AI Ethics* **2021**, *2*, 405–429. [CrossRef]
8. Stahl, B.C.; Brooks, L.; Hatzakis, T.; Santiago, N.; Wright, D. Exploring ethics and human rights in artificial intelligence—A Delphi study. *Technol. Forecast. Soc. Chang.* **2023**, *191*, 122502. [CrossRef]
9. Mittelstadt, B. Principles alone cannot guarantee ethical AI. *Nat. Mach. Intell.* **2019**, *1*, 501–507. [CrossRef]
10. Ryan, M.; Stahl, B.C. Artificial intelligence ethics guidelines for developers and users: Clarifying their content and normative implications. *J. Inf. Commun. Ethics Soc.* **2021**, *19*, 61–86. [CrossRef]
11. Williams, O. Towards Human-Centred Explainable AI: A Systematic Literature Review. Master's Thesis, University of Birmingham, Birmingham, UK, 2021. [CrossRef]
12. Hartikainen, M.; Väänänen, K.; Olsson, T. Towards a Human-Centred Artificial Intelligence Maturity Model. In *Proceedings of the Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany, 23–28 April 2023*; Association for Computing Machinery: New York, NY, USA, 2023. [CrossRef]
13. Owe, A.; Baum, S.D. Moral consideration of nonhumans in the ethics of artificial intelligence. *AI Ethics* **2021**, *1*, 517–528. [CrossRef]
14. Baum, S.D.; Owe, A. Artificial Intelligence Needs Environmental Ethics. *Ethics Policy Environ.* **2023**, *26*, 139–143. [CrossRef]
15. Rees, C.; Müller, B. All that glitters is not gold: Trustworthy and ethical AI principles. *AI Ethics* **2022**, *16*, 1–4. [CrossRef]
16. Asimov, I. *I, Robot*; Harper Voyager: New York, NY, USA, 2018.
17. Asimov, I. *Robots and Empire*; Harper Voyager: New York, NY, USA, 2018.
18. Fröding, B.; Peterson, M. Friendly AI. *Ethics Inf. Technol.* **2021**, *23*, 207–214. [CrossRef]
19. He, H.; Gray, J.; Cangelosi, A.; Meng, Q.; McGinnity, T.M.; Mehnen, J. The Challenges and Opportunities of Human-Centered AI for Trustworthy Robots and Autonomous Systems. *IEEE Trans. Cogn. Dev. Syst.* **2022**, *14*, 1398–1412. [CrossRef]
20. Murphy, R.; Woods, D.D. Beyond Asimov: The Three Laws of Responsible Robotics. *IEEE Intell. Syst.* **2009**, *24*, 14–20. [CrossRef]
21. European Parliament. European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics. In *Technical Report 2015/2103(INL)*; European Parliament: Eurométropole de Strasbourg, France, 2017.
22. Nevejans, N. European Civil Law Rules in Robotics. In *Technical Report*; European Parliament: Eurométropole de Strasbourg, France, 2016.
23. Brennan, A.; Lo, N.Y.S. Environmental Ethics. In *The Stanford Encyclopedia of Philosophy*, Winter 2021 ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2021. Available online: <https://plato.stanford.edu/archives/win2021/entries/ethics-environmental> (accessed on 12 January 2021).
24. Zimmerman, M.J.; Bradley, B. Intrinsic vs. Extrinsic Value. In *The Stanford Encyclopedia of Philosophy*, Spring 2019 ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2019. Available online: <https://plato.stanford.edu/entries/value-intrinsic-extrinsic/> (accessed on 12 January 2021).
25. Aristotle. *The Basic Works of Aristotle*; Modern Library: New York, NY, USA, 2001.
26. Aquinas, T. *Summa Contra Gentiles*; Burns & Oates and B. Herder, N.D: London, UK, 1905.
27. Kant, I. *Lectures on Ethics*, 1997 ed.; The Cambridge Edition of the Works of Immanuel Kant; Cambridge University Press: Cambridge, UK, 1930.
28. Passmore, J. *Man's Responsibility to Nature*; Gerald Duckworth & Co. Ltd.: London, UK, 1974.
29. Kagan, S. What's Wrong with Speciesism. *J. Appl. Philos.* **2016**, *33*, 1–21. [CrossRef]
30. High-Level Expert Group on Artificial Intelligence. *The Ethics Guidelines for Trustworthy Artificial Intelligence (AI)*; Report; European Commission: Brussels, Belgium, 2018.
31. Svoboda, T. Duties Regarding Nature: A Kantian Approach to Environmental Ethics. *Kant Yearb.* **2012**, *4*, 143–163. [CrossRef]
32. Singer, P. *Animal Liberation: The Definitive Classic of the Animal Movement*, 2015 ed.; Open Road Media: New York, NY, USA, 1975.
33. Goodpaster, K.E. On being morally considerable. *J. Philos.* **1978**, *75*, 308–325. [CrossRef]
34. Palmer, C. Does nature matter? The place of the nonhuman in the ethics of climate change. In *The Ethics of Global Climate Change*; Arnold, D.G., Ed.; Cambridge University Press: Cambridge, UK, 2011; pp. 272–291.
35. Rolston, H. Is There an Ecological Ethic? *Ethics* **1975**, *85*, 93–109. [CrossRef]
36. Leopold, A. The Land Ethic. In *Environmental Ethics: The Big Questions*; Keller, D.R., Ed.; Wiley-Blackwell: Chichester, UK, 1949; pp. 193–201.
37. Naess, A.; Ecosophy T. Deep Versus Shallow Ecology. In *Environmental Ethics: Readings in Theory and Application*, 7th ed.; Pojiman, L.P., Pojiman, P., McShane, K., Eds.; Cengage: Boston, MA, USA, 2017; pp. 222–231.
38. Arksey, H.; O'Malley, L. Scoping studies: Towards a methodological framework. *Int. J. Soc. Res. Methodol.* **2005**, *8*, 19–32. [CrossRef]
39. Peterson, J.; Pearce, P.F.; Ferguson, L.A.; Langford, C.A. Understanding scoping reviews: Definition, purpose, and process. *J. Am. Assoc. Nurse Pract.* **2017**, *29*, 12–16. [CrossRef]
40. Daudt, H.M.L.; van Mossel, C.; Scott, S.J. Enhancing the scoping study methodology: A large, inter-professional team's experience with Arksey and O'Malley's framework. *BMC Med Res. Methodol.* **2013**, *13*, 48. [CrossRef]
41. Levac, D.; Colquhoun, H.; O'Brien, K.K. Scoping studies: Advancing the methodology. *Implement. Sci.* **2010**, *5*, 69. [CrossRef] [PubMed]

42. Pham, M.T.; Rajić, A.; Greig, J.D.; Sargeant, J.M.; Papadopoulos, A.; McEwen, S.A. A scoping review of scoping reviews: Advancing the approach and enhancing the consistency. *Res. Synth. Methods* **2014**, *5*, 371–385. [CrossRef]
43. Rumrill, P.D.; Fitzgerald, S.M.; Merchant, W.R. Using scoping literature reviews as a means of understanding and interpreting existing literature. *Work* **2010**, *35*, 399–404. [CrossRef]
44. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *J. Clin. Epidemiol.* **2009**, *62*, 1006–1012. [CrossRef]
45. Munn, Z.; Peters, M.D.J.; Stern, C.; Tufanaru, C.; McArthur, A.; Aromataris, E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med. Res. Methodol.* **2018**, *18*, 143. [CrossRef] [PubMed]
46. Peters, M.D.J.; Godfrey, C.M.; Khalil, H.; McInerney, P.; Parker, D.; Soares, C.B. Guidance for conducting systematic scoping reviews. *JBI Evid. Implement.* **2015**, *13*, 141–146. [CrossRef]
47. Algorithm Watch. AI Ethics Global Inventory. 2020. Available online: <https://inventory.algorithmwatch.org/> (accessed on 14 December 2021).
48. Riek, L.; Howard, D. A code of ethics for the human-robot interaction profession. In Proceedings of the We Robot 2014, Coral Gables, FL, USA, 4–5 April 2014.
49. Whittaker, M.; Crawford, K.; Dobbe, R.; Fried, G.; Kaziunas, E.; Mathur, V.; West, S.M.; Richardson, R.; Schultz, J.; Schwartz, O. AI Now 2018 Report AI Now Institute. 2018. Available online: <https://ainowinstitute.org/publication/ai-now-2018-report-2> (accessed on 7 May 2022).
50. Wood, M.; Robbel, P.; Maass, M.; Tebbens, R.D.; Meijs, M.; Harb, M.; Reach, J.; Robinson, K.; Wittmann, D.; Srivastava, T.; et al. Safety first for automated driving; Aptive, Audit, BMW, FCA, Continental, Daimler, VW, Intel, Infineon, Baidu, Here. 2019. Available Online: <https://group.mercedes-benz.com/documents/innovation/other/safety-first-for-automated-driving.pdf> (accessed on 2 June 2022).
51. Telefonica. *AI Principles of Telefonica*; Report; Telefonica: Madrid, Spain, 2018.
52. RenAIssance Foundation. Rome Call. 2020. Available online: <https://www.romecall.org/the-call/> (accessed on 2 June 2022).
53. Hochschule der Medien. 10 Ethische Leitlinien für die Digitalisierung von Unternehmen (10 Ethical Guidelines for the Digitalisation of Companies). 2017. Available online: https://www.hdm-stuttgart.de/digitale-ethik/digitalkompetenz/ethische_unternehmensleitlinien (accessed on 2 June 2022).
54. Floridi, L.; Cows, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds Mach.* **2018**, *28*, 689–707. [CrossRef]
55. Fairness, Accountability, and Transparency in Machine Learning (FATML). Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. 2016. Available online: <https://www.fatml.org/resources/principles-for-accountable-algorithms> (accessed on 2 June 2022).
56. Cigref. *A Guide for Professionals of The Digital Age Extended Consideration Towards the Environmental Footprint of Digitisation*; Report; Cigref: Paris, France, 2018.
57. Government of the Republic of Korea. *Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society*; Report; Government of the Republic of Korea: Seoul, Republic of Korea, 2017.
58. Future of Life Institute. Asimolar AI Principles. Available online: <https://futureoflife.org/2017/08/11/ai-principles/> (accessed on 2 June 2022).
59. Laskai, L.; Webster, G. Translation: Chinese Expert Group Offers ‘Governance Principles’ for ‘Responsible AI’; New America. 2019. Available online: <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-expert-group-offers-governance-principles-responsible-ai/> (accessed on 2 June 2022).
60. IA-Latam. IA-Latam Ethics Statement for the Design, Development and Use of Artificial Intelligence. Available online: <https://ia-latam.com/etica-ia-latam/> (accessed on 2 June 2022).
61. UNI Global Union. *Top 10 Principles for Ethical Artificial Intelligence*; Report; UNI Global Union & The Future World of Work: Nyon, Switzerland, 2017.
62. Leslie, D. *Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design And Implementation of AI Systems in The Public Sector*; Report; Alan Turing Institute: London, UK, 2019.
63. Beijing Academy of Artificial Intelligence. Beijing AI Principles. 2019. Available online: <https://www-pre.baai.ac.cn/news/beijing-ai-principles-en.html> (accessed on 2 June 2022).
64. Machine Intelligence Garage Ethics Committee. *Ethics Framework—Responsible AI*; Report; Machine Intelligence Garage: London, UK, 2018.
65. ITechLaw. *Responsible AI: Global Policy Framework*; Report; ITechLaw: Toronto, ON, Canada, 2019.
66. Tieto Corporation. *Tieto’s AI Ethics Guidelines*; Report; Tieto Corporation: Espoo, Finland, 2018.
67. Delvaux, M.; Mayer, G.; Boni, M. *Report with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))*; Report; European Parliament: Eurométropole de Strasbourg, France, 2017.
68. IEEE. *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems*, 1st ed.; Report; IEEE: New York, NY, USA, 2019.

69. European Commission, Directorate-General for Research and Innovation, European Group on Ethics in Science and New Technologies. *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*; Report; European Commission: Brussels, Belgium, 2018.
70. Vodafone Group Plc. *AI Framework*; Report; Vodafone: Berkshire, UK, 2019.
71. National Research Council Canada. *Advisory Statement on Human Ethics in Artificial Intelligence and Big Data Research (2017)*; Report; Government of Canada: Ottawa, ON, Canada, 2017.
72. Department of Health and Social Care. *A Guide to Good Practice for Digital and Data-Driven Health Technologies*; Report; National Health Service (NHS): London, UK, 2021.
73. The Institute for Information and Communications Policy (IICP) of the Ministry of Internal Affairs and Communications (MIC). *Draft AI R&D Guidelines for International Discussions*; Report; The Conference toward AI Network Society; MIC: Tokyo, Japan, 2017.
74. Tranberg, P.; Hasselbalch, G.; Olsen, B.K.; Byrne, C.S. *Data Ethics Principles*; Report; DataEthics.eu. 2017. Available online: <https://dataethics.eu/wp-content/uploads/Dataethics-uk.pdf> (accessed on 2 June 2022).
75. Cutler, A.; Pribić, M.; Humphrey, L. Everyday Ethics for Artificial Intelligence. In *A Practical Guide for Designers & Developers*; Report; IBM: Armonk, NY, USA, 2018.
76. World Economic Forum and Global Future Council on Human Rights 2016–2018. White Paper: How to Prevent Discriminatory Outcomes in Machine Learning. 120318—Case 00040065; White Paper; 2018. Available online: https://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf (accessed on 2 June 2022).
77. Privacy International. Privacy and Freedom of Expression in the Age of Artificial Intelligence. Privacy International. 2018. Available online: <https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf> (accessed on 2 June 2022).
78. Personal Data Protection Commission Singapore. Discussion Paper on Artificial Intelligence (AI) and Personal Data - Fostering Responsible Development and Adoption of AI. 2018. Available online: <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf> (accessed on 2 June 2022).
79. OECD (Organisation for Economic Co-operation). Principles for Responsible Stewardship of Trustworthy AI; 2019. Available online: <https://oecd.ai/en/ai-principles> (accessed on 2 June 2022).
80. Telia Company. *Guiding Principles on Trusted AI*; Report; Telia Company: Stockholm, Sweden, 2019. Available online: https://www.teliacompany.com/assets/u5c1v3pt22v8/2vc3JrcTrq177ww43dChjh/e7277ac89ac0c75926eba76625f37dd7/TC_guiding_principles_on_trusted_AI_Jan11.pdf (accessed on 8 June 2022).
81. Cremers, A.B.; Englander, A.; Gabriel, M.; Hecker, D.; Mock, M.; Poretschkin, M.; Rosenzweig, J.; Rostalski, F.; Sicking, J.; Volmer, J.; et al. *Trustworthy Use of Artificial Intelligence*; Report; Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS: Sankt Augustin, Germany, 2020.
82. van Wynsberghe, A.; Donhauser, J. The Dawning of the Ethics of Environmental Robots. *Sci. Eng. Ethics* **2018**, *24*, 1777–1800. [[CrossRef](#)] [[PubMed](#)]
83. Ditmer, M.A.; Vincent, J.B.; Werden, L.K.; Tanner, J.C.; Laske, T.G.; Iazzo, P.A.; Garshelis, D.L.; Fieberg, J.R. Bears Show a Physiological but Limited Behavioral Response to Unmanned Aerial Vehicles. *Curr. Biol.* **2015**, *25*, 2278–2283. [[CrossRef](#)]
84. Cawthorne, D.; Juhl, P.M. Designing for Calmness: Early Investigations into Drone Noise Pollution Management. In Proceedings of the 2022 International Conference on Unmanned Aircraft Systems (ICUAS), Dubrovnik, Croatia, 21–24 June 2022; pp. 839–848. [[CrossRef](#)]
85. Rosenberg, L. Drones Affect the Environment in Many Different Ways. 2021. Available online: [GreenMatters.com](https://www.greenmatters.com) (accessed on 6 April 2022).
86. Wigglesworth, A. A Generation of Seabirds Was Wiped out by a Drone in O.C. Scientists Fear for Their Future. *LA Times*, 7 June 2021. Available online: <https://www.latimes.com/california/story/2021-06-07/thousands-of-eggs-abandoned-after-drone-crash-at-orange-county-nature-reserve> (accessed on 6 April 2022).
87. Blersch, D.M.; Kangas, P.C. Towards an Autonomous Algal Turf Scrubber: Development of an Ecologically-Engineered Technoecosystem. Ph.D. Thesis, University of Maryland, College Park, MD, USA, 2010. Available online: <https://api.drum.lib.umd.edu/server/api/core/bitstreams/ceb56e49-531b-4f33-b90e-916ed5a6040c/content> (accessed on 23 May 2023).
88. Clark, O.G.; Kok, R. Engineering of highly autonomous biosystems: Review of the relevant literature. *Int. J. Intell. Syst.* **1998**, *13*, 749–783. [[CrossRef](#)]
89. Kawai, K.; Takizawa, Y.; Watanabe, S. Advanced automation for power-generation plants—Past, present and future. *Control Eng. Pract.* **1999**, *7*, 1405–1411. [[CrossRef](#)]
90. Chen, C.; Fu, J.; Lyu, L. A Pathway Towards Responsible AI Generated Content. *arXiv* **2023**, arXiv:2303.01325.
91. Narayanan, D.; Shoeybi, M.; Casper, J.; LeGresley, P.; Patwary, M.; Korthikanti, V.; Vainbrand, D.; Kashinkunti, P.; Bernauer, J.; Catanzaro, B.; et al. Efficient Large-Scale Language Model Training on GPU Clusters Using Megatron-LM. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, St. Louis, MO, USA, 14–19 November 2021*; Association for Computing Machinery: New York, NY, USA, 2021. [[CrossRef](#)]
92. Longley, J. What Is an Average Household's Energy Usage? *Utility Bidder*. Available online: <https://www.utilitybidder.co.uk/compare-business-energy/what-is-an-average-households-energy-usage/> (accessed on 23 May 2023).
93. Tamburrini, G. The AI Carbon Footprint and Responsibilities of AI Scientists. *Philosophies* **2022**, *7*, 4. [[CrossRef](#)]

94. Strubell, E.; Ganesh, A.; McCallum, A. Energy and Policy Considerations for Deep Learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 3645–3650. [CrossRef]
95. Joler, V.; Crawford, K. Anatomy of an AI System. 2018. Available online: <https://anatomyof.ai/> (accessed on 23 August 2022).
96. Laukyte, M. Against Human Exceptionalism: Environmental Ethics and the Machine Question. In *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence*; Springer: Cham, Switzerland, 2019; Volume 134, pp. 325–339. [CrossRef]
97. Darling, K. Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects Robot Law. In *Robot Law*; Calo, R.A., Froomkin, M., Kerr, I., Eds.; Edward Elgar Publishing: Cheltenham, UK, 2016; pp. 213–234.
98. Gunkel, D.J. A vindication of the rights of machines. *Philos. Technol.* **2014**, *27*, 113–132. [CrossRef]
99. Taylor, P.W. The Ethics of Respect for Nature. *Environ. Ethics* **1981**, *3*, 197–218. [CrossRef]
100. Gibert, M.; Martin, D. In search of the moral status of AI: Why sentience is a strong argument. *AI Soc.* **2022**, *37*, 319–330. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.