

# Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications

Mark Ryan

*Department of Philosophy, KTH Royal Institute of Technology, Stockholm, Sweden, and*

Bernd Carsten Stahl

*School of Computer Science and Informatics, Centre for Computing and Social Responsibility, De Montfort University, Leicester, UK*

Received 11 December 2019  
Revised 18 March 2020  
20 April 2020  
Accepted 13 May 2020

## Abstract

**Purpose** – The purpose of this paper is clearly illustrate this convergence and the prescriptive recommendations that such documents entail. There is a significant amount of research into the ethical consequences of artificial intelligence (AI). This is reflected by many outputs across academia, policy and the media. Many of these outputs aim to provide guidance to particular stakeholder groups. It has recently been shown that there is a large degree of convergence in terms of the principles upon which these guidance documents are based. Despite this convergence, it is not always clear how these principles are to be translated into practice.

**Design/methodology/approach** – In this paper, the authors move beyond the high-level ethical principles that are common across the AI ethics guidance literature and provide a description of the normative content that is covered by these principles. The outcome is a comprehensive compilation of normative requirements arising from existing guidance documents. This is not only required for a deeper theoretical understanding of AI ethics discussions but also for the creation of practical and implementable guidance for developers and users of AI.

**Findings** – In this paper, the authors therefore provide a detailed explanation of the normative implications of existing AI ethics guidelines but directed towards developers and organisational users of AI. The authors believe that the paper provides the most comprehensive account of ethical requirements in AI currently available, which is of interest not only to the research and policy communities engaged in the topic but also to the user communities that require guidance when developing or deploying AI systems.

**Originality/value** – The authors believe that they have managed to compile the most comprehensive document collecting existing guidance which can guide practical action but will hopefully also support the consolidation of the guidelines landscape. The authors' findings should also be of academic interest and inspire philosophical research on the consistency and justification of the various normative statements that can be found in the literature.

**Keywords** Artificial intelligence, AI ethics guidelines, Recommendations for emerging technologies, Policymaking for disruptive technologies

**Paper type** Research paper



© Mark Ryan and Bernd Carsten Stahl. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

This project (SHERPA) has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 786641.

## 1. Introduction

Ethical consequences of artificial intelligence (AI) is a hot topic of debate across academia, policy and general media. It has been shown that there is a large degree of convergence in terms of the principles that guidance documents are based on (Jobin *et al.*, 2019). At the same time, the principle-based approach adopted by much of the discourse has been criticised as insufficient in dealing with the practical issues raised by AI (Mittelstadt, 2019). The quickly growing set of tools that are being developed and provided to address AI ethics are often difficult to map with regards to the categories or principles they could help to address (Morley *et al.*, 2019).

In this paper, we move beyond the high-level ethical principles that are common across the AI ethics guidance literature and provide a description of the *content* that is covered by these principles. We build on Jobin *et al.*'s (2019, p. 395) robust categorisation of ethical principles. While their work provides a comprehensive overview of currently available AI ethics guidelines, their contribution is merely descriptive about these guidelines, rather than discussing the normative content of them. Our paper builds upon these foundations and uses their cohesive approach to develop a presentation of the normative content of these ethics guidelines for organisations developing and using AI.

While there is an abundance of AI ethics guidelines, these guidelines remain separate and distinct from one another. As a consequence, it is difficult for individuals involved in the development or use of AI to determine which ethical issues they should be aware of, how these can present themselves and how they may be addressed. The reference to particular ethical principles, such as fairness, transparency or sustainability may be a good starting point, but further detail is required that allows AI organisations to think through the implications of these principles for their work.

A further issue of AI ethics guidelines is that they are aimed at a range of stakeholders: not only policymakers, users and developers but also educators, civil society organisations, industry associations, professional bodies and more. As a consequence, the guidelines that are currently available are often difficult to understand and are written for technical users who constitute one key user group.

In this paper, we therefore provide a detailed explanation of the normative implications of existing AI ethics guidelines but directed towards developers and organisational users of AI. [1] We believe that the paper provides the most comprehensive account of ethical requirements in AI guidelines currently available, which is of interest not only to the research and policy community engaged in the topic but also to the user communities that require guidance when developing or deploying AI systems. It must be made clear here that we are not providing prescriptive recommendations, but rather, are mapping the prescriptive recommendations found in these guidelines.

To provide this normative account, we start with a brief overview of the current academic and policy-oriented discourse on ethics and AI. We then describe the methodology of our work and how we compiled the relevant insights. The largest section of the paper describes 11 normative principles (transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, sustainability, dignity and solidarity) and the various subcategories of these principles. In conclusion, we highlight the contribution of this work and suggest next steps.

## 2. Research and policy in ethics and artificial intelligence

Ethics guidelines constitute one aspect of the larger academic and policy discourse around ethics and AI. It is probably not contentious to state that an interest in ethics and AI is now a global phenomenon. The amount of attention currently paid to the topic is impressive and

---

the literature has mushroomed to the point where it is difficult to keep on top of it. In this paper, we focus on ethics guidelines, but these need to be seen as one aspect of a broader literature on ethics and AI.

Ethics of AI is not a new topic. What falls under this heading depends on the definition of the term AI. A typical definition is “we define AI as a system’s ability to interpret external data correctly, to learn from such data and to use those learnings to achieve specific goals and tasks through flexible adaptation” (Kaplan and Haenlein, 2019, p. 17). Aspects that are typically described as defining features of AI that can give rise to ethical concerns are the ability to learn and to act more or less autonomously on the basis of external input and adaptation.

If these characteristics are at the core of the ethical discussion of AI, then they can be traced back to the very beginning of discussions of ethics and digital technology in the 1940s and 1950s (Wiener, 1954) and they have been driving at least parts of the debate on ethics and technology, computing and information ever since (Bynum, 2010; Bynum and Rogerson, 2003; Capurro, 2008; Moor, 1985). However, even though the debate can be followed back several decades, it has become invigorated in recent years. The generally accepted explanation for this upsurge in AI ethics is based on recent successes and achievements of some AI techniques, and their widespread application in domains such as smart cities (Ryan and Gregory, 2019; Ryan 2019b), agriculture (Ryan 2019a) and transportation (Ryan 2019c).

In particular, machine learning and deep neural networks have been hugely successful in recent years. While not a fundamentally novel technology, recent successes of machine learning have been made possible by the availability of large data sets for training and testing purposes and the affordability and availability of large amounts of computing power. It is important to note that the field of AI, which has been a long-standing part of computer science, goes beyond machine learning, big data and neural networks, but these are at the heart of the current debate. A typical description of the expectation of AI’s future role is as follows:

[. . .] AI will become as much a part of everyday life as the Internet or social media did in the past. In doing so, AI will not only impact our personal lives but also fundamentally transform how firms take decisions and interact with their external stakeholders (e.g. employees, customers) (Haenlein and Kaplan, 2019, p. 9).

This widely shared and accepted narrative that AI will have a large impact on many aspects of life explains the high level of public interest. There have been numerous high-level policy reports that describe the current and expected effects of AI on society and economy (Executive Office of the President, 2016a, 2016b; HoL, 2018; House of Commons Science and Technology Committee, 2016; OECD, 2019). Many industrialised countries now have AI strategies and government departments (Stix, 2019). This policy-oriented discussion reflects the academic research discourse around AI ethics (Berendt, 2019; Clark, 2019; Floridi, 2019; Johnson *et al.*, 2019; Morley *et al.*, 2019) but looks at it from a policy perspective. Their proposals range from national or international regulation and legislation and the corresponding creation of regulatory bodies to corporate governance mechanisms, the creation of standards and codes of ethics to a range of sector-specific measures (e.g. in health, automation and military) and technical means.

Many of the outputs of research-oriented, private and political organisations on AI ethics take the form of guidelines. Prominent examples include the EU’s high level expert group on AI’s guidelines (High-Level Expert Group on AI, 2019) or the Asilomar AI principles (Asilomar Conference, 2017). These guidelines aim to provide guidance for particular

---

stakeholder groups on how to deal with ethical issues they face. They often contain a set of ethical principles which are then used to deduce more specific guidance. Such guidelines need to be read in the context of the legal structure in which they apply. While ethics guidelines often aspire to be incorporated within policy frameworks, they are in themselves meant as guiding frameworks, rather than indicating or enforcing legal parameters for action. Thus, the guidelines are intended as indications towards ethical behaviour, but their target audiences should abide by current legislation in the area and not negate their legal obligations.

The question that motivated this paper was which practical guidance is available to people who develop or use AI that will help them address ethical concerns they face. Our starting assumption was that the answer to this question should be found in AI ethics guidelines. However, the wealth of existing guidelines raises two related problems that this paper aims to address. First problem is that many of the guidelines are very broad in terms of coverage, i.e. they provide guidance for many different stakeholder groups, including policymakers, companies, users, civil society representatives etc. Second, there is now such a wealth of guidelines that it is very difficult to navigate and understand which pieces of guidance exist and what the specific guidance is.

This paper is aimed particularly at people who develop or use AI systems, and it tries to clarify which ethical principles can guide their work. Most importantly, the paper drills down more deeply into the details of the body of knowledge and specifies which ethical aspects are covered by the range of principles and what users and developers should do to carry out their moral responsibilities. We have compiled the most comprehensive document collecting existing guidance which can guide practical action but will hopefully also support the consolidation of the guidelines landscape. Our findings should also be of academic interest and inspire philosophical research on the consistency and justification of the various normative statements that can be found in the literature.

Before we come to the actual guidance, we give a quick overview of the methodology used in our research.

### 3. Methodology

The most important requirement for our research was to have a comprehensive data set of AI ethics guidelines. To achieve this, we started with a structured search of available databases (Scopus, Web of Science and Google Scholar), using the search terms including “AI ethics”, “AI guidelines”. We compared our findings with existing collection of relevant documents, notably Stix’s European AI Ecosystem ([www.charlottestix.com/european-union-ai-ecosystem](http://www.charlottestix.com/european-union-ai-ecosystem)), the Algorithmwatch AI Ethics Guidelines Global Inventory (<https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/>) but also broader AI repositories, such as the Open AI website (<https://openai.com/>). We collected all the documents that were publicly available and then broadened our search on the basis of references in the published literature as well as references in the guidelines we had already secured.

We used the most comprehensive and rigorously structured overview of guidelines published so far (Jobin *et al.*, 2019) to validate our data set. The result was that we analysed Jobin’s 82 sets of ethical guidelines and an additional 9 guidelines (see articles in bold in Appendix) that they did not include (the total was 91 guidelines). [2]

We then undertook a thematic analysis of all the guidelines (Aronson, 1995; Braun and Clarke, 2006). As a starting point we used the ethical principles that are used by the EU’s HLEG (2019) as high-level coding points. We then identified which ethical principles or guidance fell underneath each of these headings. Table 1 (below) provides an overview of the main principles and the ethical issues that constitute these.

**Table 1.**  
Guiding ethical  
principles and  
constituent ethical  
issues

Principle	Constituent ethical issues or guidance			
Transparency	transparency interpretability	explainability communication	explicability disclosure	understandability showing
Justice and fairness	justice	fairness	consistency	inclusion
Non-maleficence	equality diversity remedy	equity plurality redress	non-bias accessibility challenge	non-discrimination reversibility access and distribution
	non-maleficence protection non-subversion	security precaution	safety prevention	harm integrity
Responsibility Privacy	responsibility privacy	accountability personal or Private information	liability	acting with integrity
Beneficence	benefits social good	beneficence common good	well-being	peace
Freedom and autonomy	freedom	autonomy	consent	choice
Trust Sustainability	self-determination trustworthiness	liberty	empowerment	
	sustainability	environment (nature)	energy	resources (energy)
Dignity Solidarity	dignity solidarity	social security	cohesion	

The identification of ethical principles was done on the basis of a close reading of the guidelines in our data set and following [Jobin et al.'s \(2019\)](#) sub-categories. We tried to stay as close to the data as possible and therefore erred on the side of caution and inclusion. As a consequence, we included a number of concepts that are semantically very similar which might have been possible to merge, but which are discussed separately in different documents.

As our main interest was in determining which guidance exists for developers and users of AI, we distilled the guidance that was provided in the guidelines. As a result, our findings are strongly normative, i.e. they give guidance and instructions and are phrased accordingly (e.g. “AI organisations should. . .”), rather than simply recounting what each guideline says on the matter. This is a result of our research approach and our interest in extracting guidance. The formulations we use in the next section does not imply that we are endorsing all of these guidelines or that we are suggesting that individuals always have to follow them. The meaning is that within the corpus of AI ethics guidelines there are suggestions that the indicated activities are morally appropriate. It falls outside of the scope of this paper to do a proper ethical analysis of the guidance, including their detailed ethical justification and check for consistency.

#### 4. Guidelines for the development and use of artificial intelligence

Following this methodology, we analysed the set of guidelines and compiled the detailed guidance that is available to developers and users. We established that while there was a strong degree of overlap about the main issues and themes within the guidelines, they often differed in a number of areas: emphasis on the topic (a greater emphasis on algorithms, privacy and security, or safety), the tone (varying between dogmatic “must do” principles to more open “if possible” recommendations), length (ranging from 1 page to over 266 pages),

level of technicality (very technical to layman terminology) and audience (end-users, developers, companies, policymakers, or society as a whole).

The following subsections highlight the nature of the ethical issues and guidance that has been suggested for developers and users to follow. We reference the relevant guidelines where required but should state that we only provide minimal references per section because our aim is to give an overview of the ethical aspects and the normative content within all of these guidelines, rather than providing a systematic and robust mapping of guidelines to issues, as this has already been done quite well in [Jobin et al. \(2019\)](#).

#### *4.1 Transparency*

Transparency has quickly become one of the most widely discussed principles within the AI ethics debate, with [Floridi \(2019\)](#) and the [High-Level Expert Group on AI \(2019\)](#) viewing it as a defining characteristic within the debate. Transparency can typically be understood in two ways: the transparency of the AI technology itself and the transparency of the AI organisations developing and using it. Throughout our analysis, transparency was regularly discussed directly, or in relation to processes required to ensure it, such as explainability, understandability and communication.

*4.1.1 Transparency.* AI developers need to ensure transparency because it protects many other requirements – such as the fundamental human rights, privacy, dignity, autonomy and well-being ([UNI Global Union, 2017](#)). Organisations using AI should be transparent about their aim for using AI, benefits and harms and potential outcomes that may occur ([IBM, 2017](#)). AI developers should ensure transparency because it allows consumers to make informed choices about sharing their data and using AI ([ADMA, 2013](#)).

*4.1.2 Explainability.* AI must be subject to active monitoring to ensure that they are producing accurate results ([Algo.Rules, 2019](#)). AI organisations should document how their AI makes certain decisions and be able to reproduce them for audits ([SIIA, 2017](#)). AI should be explainable to external algorithmic auditing bodies to ensure the technical and ethical functionality of their AI. If there is a tension between performance and explainability, this should be clearly identified ([Cerna Collectif, 2018](#)).

*4.1.3 Explicability.* AI organisations (i.e. organisations using or developing AI) should be able to intelligibly explain the data that goes in, the data coming out, what their algorithms do, and their objective for doing so ([Demiaux and Abdallah, 2017](#), p. 51). AI organisations should ensure traceability and explicability to guarantee safety ([OECD 2019](#)). AI needs to have a strong degree of traceability to ensure that if harms arise, they can be traced back to the cause ([IEEE, 2017](#)). Data should be traceable back to where, how and when it was captured, retrieved, cleaned and analysed ([Cerna Collectif, 2018](#)). Decisions made by AI should be reproducible by external auditors ([AMA, 2018](#)).

*4.1.4 Understandability.* AI organisations need to implement appropriate methods to monitor the data, algorithms and the decisions that will be arrived at by those processes, and for actions taken by AI to be comprehensible by human beings ([European Parliament, 2017](#)). AI organisations should understand how their AI works and explain the technical functioning and decisions reached by those technologies, whenever possible ([Floridi et al., 2018](#)).

*4.1.5 Interpretability.* While there is a degree of opaqueness in some machine-learning technologies, AI organisations should be able to understand how a decision was reached and how human oversight ensures that harms caused by algorithmic black-boxing are addressed and prevented ([IEEE, 2019](#)). High-stake domains (such as health care, criminal justice and welfare) should reconsider using black-box AI altogether ([AI Now Institute, 2017](#)). Algorithmic reviews should be done on a regular basis to determine if they are fit-for-



---

purpose and interpretable (Algo.Rules, 2019). Organisations should be able to clearly interpret and demonstrate how their AI is abiding by current legislation, such as the general data protection regulation (GDPR), and be able to demonstrate what measures are being taken to ensure compliance (UK Government, 2018).

*4.1.6 Communication.* End users should be provided with accurate information to ensure that they are not manipulated, deceived, or coerced by AI (High-Level Expert Group on AI, 2019, p. 16). End users should be informed about the intent and outcomes of the technology (IBM, 2018). AI companies should be explicitly clear and discuss in a jargon-free manner, the potential flaws or harm that may arise from their AI (Algo.Rules, 2019). Communication methods may have to change for different industries, expertise and context of use (Floridi *et al.*, 2018). AI organisations should communicate their progress and likelihood to hit particular milestones to governments, so that they can plan for these outcomes (NSTC, 2016a).

*4.1.7 Disclosure.* AI should be designed and used to retrieve little to no personal data, or if required, that any data retrieved is anonymised, encrypted and securely processed, while being able to demonstrate this to a third-party auditor (High-Level Expert Group on AI, 2019). AI should go through internal and external auditing to ensure they are fit for purpose, but the organisation also needs to be able to explain and justify the use of their AI. Organisations should allow for independent analysis and review of their systems (Amnesty International/Access Now, 2018).

*4.1.8 Showing.* Data should be accurate, up-to-date and fit-for-purpose, and companies should be able to demonstrate this (ICO, 2017). Data quality should be transparent, available for periodic assessment and there should be regular and continued anomaly detection set in place [United Nations Development Group (UNDG), 2017]. Developers of AI should also be able to provide their ethics codes to public authorities, organisational users and where possible, the public (University of Montreal, 2017). This can be achieved through periodic review sessions, appropriate oversight mechanisms and collective responsibility approaches within the organisation (ICDPPC, 2018). It should also be clear to the end user that they are interacting with an AI system, rather than a human (EPSRC, 2011).

## 4.2 Justice and fairness

Discrimination and unfair outcomes stemming from algorithms has become a hot topic within the media and academic circles (O'Neil, 2016). It is not surprising that issues of fairness, equality and equity were repeatedly discussed throughout the ethics guidelines. In addition to simply addressing issues of harm and injustice themselves, many of the guidelines provided recommendations on how to implement steps to minimise these harms. Furthermore, some documents also highlighted how different organisations should implement methods to reverse, remedy and allow fair redress, in instances where harms have occurred.

*4.2.1 Justice.* AI practitioners should identify what levels of justice and fairness can be implemented into the AI system during the design process (NSTC, 2016b). For example, if AI is used within the judicial system in any way, accountability should still lie with the human user, e.g. the judge (Rathenau Institute, 2017, p. 43). In addition, AI will replace many human jobs in the future, so it is important that there are effective and just ways to retrain and retool the human workforce (COMEST/UNESCO, 2017, pp. 52-53).

*4.2.2 Fairness.* While AI developers may have their own values, they should not develop algorithms with historically unfair prejudices (Latonero, 2018). There should be steps in place to ensure that data being used by AI is not unfair, or contains errors and inaccuracies, that will corrupt the response and decisions taken by the AI (ICO, 2017). To ensure the

fairness of AI, their design should be fit for purpose, identify impacts on different aspects of society and should be designed to promote human welfare, rather than endanger it (ICDPPC, 2018). Organisations should consider using fairness-aware data mining algorithms (FATML, 2016).

*4.2.3 Consistency.* To prevent harmful actions in the decision-making process, organisations should ensure that accurate and representative sample data is collected, analysed and used [IPC Ontario (Information and Privacy Commissioner of Ontario), 2017]. Organisations need to establish procedures to ensure the identification, prevention and the minimisation of inaccuracies in their AI. To achieve this, data should be of the highest quality (UNDG, 2017), external algorithmic auditing should be carried out (Intel, 2017), and there should be consistent, repeated and regular discussions with end users and stakeholders that may be affected (PwC, 2019).

*4.2.4 Inclusion.* AI should not become another tool for exclusion within society (AI for Humanity, 2018). Particular attention should be given to under-represented and vulnerable groups and communities, such as those with disabilities, ethnic minorities, children and those in the developing world (High-Level Expert Group on AI, 2019). Data that is being used should be representative of the target population and should be as inclusive as possible (High-Level Expert Group on AI, 2019). AI organisations should not only reduce exclusion issues but should promote active inclusion of women and minority groups into the development and design of AI (Gilburt, 2019; WEF, 2018).

*4.2.5 Equality.* AI should not harm, and where ever possible, should promote, the equality of individuals in respect to their rights, dignity and freedom to flourish (The Future Society, 2018; Tieto, 2018). One way equality can be enabled is through greater diversity in AI teams and data sets and designs (Sage, 2017). More steps need to be taken to address sexist, misogynistic and gender-biased harms resulting from some AI (World Wide Web Foundation, 2018).

*4.2.6 Equity.* The aims of AI, generally, should be to empower and benefit individuals, provide equal opportunities while distributing the rewards from its use in a fair and equitable manner (EGE, 2018; IEEE, 2019; SIIA, 2017). AI should be developed so that it can be used within society in a fair and equal way (Japanese Society for Artificial Intelligence, 2017).

*4.2.7 Non-bias.* AI organisations should invest in ways to identify, address and mitigate unfair biases (ICDPPC, 2018). Developers should examine unfair biases at every stage of the development process and should eliminate those found (The Public Voice, 2018). There should be close attention paid to the training data used, potential human biases and bias derived from the results of algorithmic processes (Cerna Collectif, 2018). Developers and organisational users of AI should conduct analysis to identify unfair bias, and there should be explicit attempts to avoid individual and societal bias, continual mechanisms in place and dialogue with stakeholders to raise awareness and reverse any biases detected (IBM, 2018). If there is any indication of unfair bias, the AI organisations should demonstrate the elimination of such bias before a competent authority (Council of Europe, 2017).

*4.2.8 Non-discrimination.* AI should be designed for universal usage and not discriminate against people, or groups of people, based on gender, race, culture, religion, age or ethnicity (Cerna Collectif, 2018). There should be mechanisms in place to effectively prevent, remedy and reverse discriminatory outcomes resulting from AI use (Amnesty International/Access Now, 2018). AI use should not lead to discrimination against individuals or groups of individuals in accordance with the Equality Act 2010, and organisations should create “discrimination impact assessments” to identify issues before their AI are used (AI for Humanity, 2018).



*4.2.9 Diversity.* To promote diversity, AI organisations should instil an inclusionary working environment (Cerna Collectif, 2018), hire teams from a range of backgrounds (IBM, 2018) and disciplines (SAP, 2018), conduct regular diversity sessions and incorporate the viewpoints from a wide range of stakeholders (Amnesty International/Access Now, 2018). Organisations implementing and using AI should encourage a diversity of opinions throughout every stage of its use (Smart Dubai, 2019).

*4.2.10 Plurality.* AI developers should consider the range of social and cultural viewpoints within society and should attempt to prevent societal homogenization of behaviour and practices (University of Montreal, 2017). Organisations should not only be focused on “pipeline model” changes in their organisation but should ensure that the plurality of individuals within their organisation have a voice and they create a culture of inclusion, which should be reflected in the AI technology (AI Now Institute, 2018). Create a multi-stakeholder dialogue and incorporate the viewpoints of women, underrepresented groups and marginalised individuals at every stage of AI applications (Leaders of the G7, 2018).

*4.2.11 Accessibility.* Organisations should protect the rights of data subjects, such as the right of information access about them (Datatilsynet, 2018). Individuals have a right to access data that is being stored and used about them, and subsequently, to request that this is rectified or deleted (Datatilsynet, 2018). When decisions are made about individuals, explanations should be available that are easily accessed, free of charge and user-friendly (Smart Dubai, 2019).

*4.2.12 Reversibility.* It is important to clearly articulate if the outcomes of AI decisions are reversible, e.g. if individuals are refused a loan because of an AI algorithm, can such a decision be reversed if the customer can demonstrate their credit-worthiness (Personal Data Protection Commission Singapore, 2019, p. 16)? Organisations using AI need to ensure that the autonomy of AI is restricted and the outcomes are reversible when there is a harm caused (Floridi *et al.*, 2018). AI should be programmed with a condition of reversibility, which ensures controllability and safety of the system: The ability to undo the last action or a sequence of actions allows users to undo undesired actions and get back to the ‘good’ stage of their work” (Clark, 2019).

*4.2.13 Remedy.* When AI holds the possibility of creating harm, there needs to be preemptive steps in place to trace these issues and deal with them in a prompt and responsible manner. Organisations should abide by the “termination obligation”, which states that when a system is no longer under human control, then it must be terminated (Telefónica, 2018). There needs to be specific “red lines” drawn, that when breached, appropriate steps are taken to override the system, terminate it temporarily or indefinitely and remedy any potential issues that may have occurred (PwC, 2019).

*4.2.14 Redress.* In situations where harmful and/or unjust events occur as a result of using AI, those affected should have appropriate and visible measures of redress in a timely manner (FATML, 2016). When decisions made by algorithms create harmful or questionable results, individuals should have the possibility to lodge a complaint and request a justification of the decision (Algo.Rules, 2019). This should be done in a manner that is understandable by those affected and should allow them the opportunity to challenge these decisions (B Debate, 2017). Accountability strategies should be created within companies, with appropriate measures for redress if these internal and external standards are not met (Dawson *et al.*, 2019).

*4.2.15 Challenge.* AI companies should allow for “conscientious objectors, employee organizing and ethical whistleblowers” (AI Now Institute, 2018). There should be clear

policies to protect conscientious objectors, employees to voice their concerns and whistleblowers to feel protected, when it is in the public interest and safety (AI Now Institute, 2018).

*4.2.16 Access and distribution.* AI organisations should ensure that their technologies are fair and accessible among a diversity of user groups within society (Smart Dubai, 2019). Organisations should especially concentrate on “populations that currently lack such access” (AI Now Institute, 2016, p. 3). AI should be accessible to those that are often socially disadvantaged (such as those with vision problems, dyslexia or mobility issues) (Sage, 2017). Wherever possible, organisations should use open data for their AI to ensure access and transparency (NSTC, 2016b).

### *4.3 Non-maleficence*

The principle of nonmaleficence gained attention, resulting from Beauchamp and Childress (1979) ground-breaking *Principles of Biomedical Ethics* and its subsequent editions. In its most basic form, it means to do no harm or avoid doing harm to others. In AI ethics, the avoidance of harm to human beings has been one of the greatest concerns, with some of the most high-profile examples coming from killer robots, autonomous cars and drone technology. It is no surprise that most of the ethics guidelines had a strong emphasis on ensuring no harm comes to citizens, through security and safety of the AI, and precautionary and remedial steps to be taken, if harm occurs.

*4.3.1 Non-maleficence.* AI should be designed with the intent of not doing foreseeable harm to human beings (Personal Data Protection Commission Singapore, 2018). Developers and organisations using AI should receive and incorporate the advice of legal authorities and research ethics boards to ensure that data is retrieved, analysed and used in a manner that does not harm individuals [IPC Ontario (Information and Privacy Commissioner of Ontario), 2017]. Organisations should regularly test their algorithms to determine that no harm results from them (ACM 2017; American College of Radiology, 2019).

*4.3.2 Security.* AI should be robust, secure and safe throughout their life cycle and must function appropriately and not pose unreasonable safety risks (OECD 2019). Organisations must ensure effective cybersecurity so that their AI is protected against attacks (Allistene, 2014). Security must be built into the architecture of the AI (Public Voice 2018) and must be tested before implementation (Algo.Rules, 2019). When security researchers find vulnerabilities or design flaws, they should disclose these findings to be resolved (Internet Society, 2017).

*4.3.3 Safety.* Developers and organisational users should ensure that AI does not infringe on human rights by ensuring their technology’s safety (EGE 2018). They must assess the public safety risks that arise from their AI and implement effective safety controls (Public Voice 2018). Organisations should enforce strict safety measures, ensuring their AI’s manageability and control and that adequate procedures are in place for security breaches (Algo.Rules, 2019). AI should pass quality assurance processes and be tested in real-world scenarios before, during and after deployment (SAP 2018).

*4.3.4 Harm.* The objectives and expected impact of AI must be assessed and documented in the development stage (Algo.Rules, 2019). The effects of these systems must be reviewed on an ongoing basis (Algo.Rules, 2019). Organisations should encourage a form of “algorithmic accountability” and should exercise caution when developing AI that may have negative impacts (ICO, 2017). AI technology that replaces human activity should produce at least a diminution of harm before it is allowed on the market (Federal Ministry of Transport and Digital Infrastructure, 2017). AI should not “cause bodily injury or severe emotional distress to any person” (IIIM, 2015).

---

*4.3.5 Protection.* Developers should implement mechanisms and safeguards to protect user safety (OECD 2019), and AI must be safe and secure throughout their life cycle (IEEE, 2019). AI systems should prioritize the protection of human life (Federal Ministry of Transport and Digital Infrastructure, 2017). External auditors should be allowed to conduct examinations and report negative impacts of the AI without fear of harm or threat by the AI organisations. In addition, the protection of whistle-blowers within AI organisations should also be ensured to allow for effective and legitimate reporting of harms (High-Level Expert Group on AI, 2019, p. 20).

*4.3.6 Precaution.* Those who develop AI must have the necessary skills to understand how they function and their potential impacts (Algo.Rules, 2019), and security precautions must be well documented (Public Voice 2018). AI organisations may receive advice from trained legal professionals, ethicists working in the area and policy analysts. If no consensus can be agreed upon, development of the AI “should not proceed in that form” (High-Level Expert Group on AI, 2019, p. 20). AI systems need to allow for human interruption, or their shutdown, when there is potential harm (Internet Society, 2017).

*4.3.7 Prevention.* An AI system must be manageable throughout the lifetime and its control must be made possible (Algo.Rules, 2019). The reliability and robustness of AI and its reliability with respect to attacks, access and manipulation must be guaranteed (Public Voice 2018). Great effort should be put into ensuring reliability and safety (IEEE, 2019). AI systems should prevent accidents from occurring, whenever possible, and avoid critical situations from occurring in the first place (Federal Ministry of Transport and Digital Infrastructure, 2017).

*4.3.8 Integrity.* Attacks against AI should not compromise the bodily and mental integrity of people by ensuring the reliability and internal robustness of the systems (EGE 2018). AI should “fail gracefully” (e.g. shutdown safely or go into safe mode) (IEEE, 2019).

*4.3.9 Non-subversion.* AI systems should be used to respect and improve the lives of citizens, rather than “subvert, the social and civic processes on which the health of society depends” (Future of LifeInstitute, 2017).

#### *4.4 Responsibility*

Moral responsibility is a very important issue within AI ethics, with a fear that companies will try to obfuscate blame and responsibility onto the autonomous or semi-autonomous system. There may also be incidences where because of this relative autonomy, AI creates a “responsibility gap”, whereby it is unclear who is responsible. Issues of responsibility, accountability, liability and acting with integrity appeared in many of the ethics guidelines that we analysed.

*4.4.1 Responsibility.* Developers are primarily responsible for the design and functionality of the AI, and when there is an error or harm, then the onus of responsibility often lies with them. When the issue is caused by the use and implementation of the technology, the onus is with the organisational user of the AI. There needs to be clear and concise allocation of responsibilities within the organisation using AI, and the creation of potential scenarios and ways to deal with harms when they occur (EGE 2018; FATML, 2016).

*4.4.2 Accountability.* AI organisations need to be aware of the issues involved with using poor data and be held accountable if there are harmful consequences as a result of this. Developers need to be aware that they are accountable for these systems’ impact on the world (IBM, 2018). They need to be open and accountable by means of auditing, monitoring and conducting impact assessments of AI (ICDPPC, 2018). A legal person must always be

---

held accountable for harms caused by AI and this blame cannot be placed on the tools that cause the damage (Algo.Rules, 2019).

*4.4.3 Liability.* There is a need to distinguish between the designer and organisational users of those systems for legal reasons (Cerna Collectif, 2018). To attribute liability in situations of malfunction, error and harms, there needs to be clear attributions of responsibility. Definitive liability should be established for when autonomous systems cause undesired effects (EGE, 2018). This can be achieved through adequate record-keeping, systems for registration, and documentation (IEEE, 2019).

*4.4.4 Acting with integrity.* AI organisations must ensure that their data meets quality and integrity standards at every stage of use (ITI, 2017). If those working with AI discover errors, security breaches or data leaks, then they must report these issues to the relevant authorities, stakeholders, and if relevant, the wider public (University of Montreal, 2017). Ethics training should be implemented to ensure responsible development and deployment of AI (AI for Humanity 2018). AI companies should respect and support the academic and professional integrity of their partners and researchers (Deepmind, 2017).

#### *4.5 Privacy*

Since the GDPR came into force in 2018, privacy has been a hot topic for anyone working in fields where personal data is being used. Particularly, there is a great concern in the development and use of AI, with many of the ethics guidelines strongly featuring privacy and data protection as key tenets in their recommendations. Because of the large abundance of data that is required for AI to work, it is important that individuals' privacy is not jeopardised as a result.

*4.5.1 Privacy.* Some of the steps that AI organisations should take to ensure privacy are the security of databases, storage and AI systems through de-identification, anomaly-detection and effective cybersecurity (IPC of Ontario, 2017); ensuring informed consent is retrieved (EGE, 2018); users should have control and access to data stored about them (IEEE, 2019); follow current data protection regulations (UK Government, 2018) and non-regulatory privacy-by-design frameworks (ICDPPC, 2018) and ensuring that the data retrieved is of a high standard. Organisations purchasing off-the-shelf AI can cultivate a privacy culture by demanding privacy-by-design AI (Datatilsynet, 2018).

*4.5.2 Personal or private information.* The development and use of AI should ensure a strong adherence to the privacy and data protection standards outlined in the General Data Protection Regulation (2018), in addition to non-regulatory frameworks, such as privacy-by-design and privacy impact assessment frameworks (IEEE, 2019; Intel, 2018). Developers and organisational users of AI must place the end user's privacy and personal data at the forefront of the design process, viewing privacy as a human right (Latonero, 2018). The end user's personal data, and data derived or created about them, should be processed in a fair, lawful and legitimate way (UNDG, 2017). Whenever possible, the collection and use of personal data should be kept to a minimum, unless completely necessary and relevant (Datatilsynet, 2018).

#### *4.6 Beneficence*

The principle of beneficence also gained greater acknowledgement and adoption after Beauchamp and Childress (1979) *Principles of Biomedical Ethics*. Beneficence essentially means to do good, to carry out an activity with the intention of benefitting someone or society as a whole. Often, beneficence is overlooked in the AI ethics literature, often being seen as a given that AI will bring benefits. The ethics guidelines we analysed highlighted

---

benevolence to promote the flourishing of individual well-being, ensuring people receive benefits from AI use, or that it should promote peace and the social and common good.

*4.6.1 Benefits.* AI organisations should ensure that their AI is designed to benefit humans (IEEE, 2019). They should clearly map out those benefits and the parties benefiting from them (The Information Accountability Foundation, 2015). AI systems must create greater benefits than their costs for people (Dawson *et al.*, 2019, p. 6) and should benefit as many people as possible (Future of Life Institute, 2017; The Partnership on AI, 2016). AI organisations should “advance scientific understanding of the world, and to enable the application of this knowledge for the benefit and betterment of humankind” (IIIM, 2015).

*4.6.2 Beneficence.* AI organisations should find solutions to some of the world’s greatest problems, such as curing diseases, ensuring food security and preventing environmental damage (Intel, 2017). AI organisations should use data retrieved for the benefit of their customers and society (OP, 2019). Ultimately, AI should “compliment the human experience in a positive way” (Unity Blog, 2018).

*4.6.3 Well-being.* AI organisations should ensure individual well-being and flourishing (IEEE, 2019). They should ensure that their AI is fit-for-purpose and that it does not prohibit individual development and access to primary goods, it ensures human welfare, and allows for the empowerment of individuals around the world (EGE, 2018). AI should be used to compliment those working in the health care sector to provide better care and support the well-being of patients (RCP London, 2018).

*4.6.4 Peace.* AI organisations should aim to avoid an “arms race in lethal autonomous weapons” (Future of Life Institute, 2017; see also Smart Dubai, 2019). If AI threatens peace, organisations should collaborate with governments to reduce potential conflicts (OpenAI, 2018).

*4.6.5 Social good.* AI should bring an improvement in beneficial opportunities for society (The Information Accountability Foundation, 2015, p. 10). AI organisations should cultivate a healthy AI industry ecosystem, built on cooperation and healthy competition (Government of the Republic of Korea, 2017, p. 62). The use of AI should not come at a cost of causing a conflict with non-users of these technologies (Ministry of State for Science and Technology Policy, 2019, p. 22).

*4.6.6 Common good.* AI should be developed to support the common good (Future of Life Institute, 2017) and the service of people (AGID, 2018). AI organisations should weigh up the benefits and harms resulting from AI and should take careful consideration to develop ways to mitigate and harms to ensure an overall common good for society (The Information Accountability Foundation, 2015, p. 8). Appropriate steps should be considered to ensure that AI is used for good and that humanity is protected from potentially harmful impacts resulting from it (OpenAI, 2018).

#### 4.7 Freedom and autonomy

Democratic societies place value in freedom and autonomy, and it is important that AI use does not encumber or harm these for us. The ethics guidelines addressed ways to ensure autonomy-promoting and liberty-protecting AI. For example, the AI organisation should ensure that individuals consent to how their data is being used, AI should not harm individuals’ abilities to make choices, or manipulate their self-determination.

*4.7.1 Freedom.* Developers should acknowledge, identify and ameliorate circumstances where AI may create harm against human freedoms. Organisations should ensure that the end users’ freedoms are not infringed upon during the use of AI (High-Level Expert Group on AI, 2019). Developers should ensure that AI does not harm end users through tracking



(freedom of movement), censorship (freedom of expression) or surveillance (freedom of association).

*4.7.2 Autonomy.* AI organisations should ensure that end users are informed, not deceived or manipulated by AI and should be allowed to exercise their autonomy (EGE, 2018). AI organisations need to ensure that the “principle of user autonomy must be central to the system’s functionality” (High-Level Expert Group on AI, 2019, p. 16). Users should be informed actors and have control over their decisions when interacting with AI (Council of Europe, 2019).

*4.7.3 Consent.* The use of personal data must be clearly articulated and agreed upon before its use (UNDG, 2017). If personal data is repurposed, developers should ensure that it is compatible with the original fair processing requirements when consent is given (ICO, 2017), in those cases where consent is the legal basis of data processing. Personal data should not be processed in a way that the data subject considers inappropriate or objectionable (Council of Europe, 2017). The use of personal data should also be done within reasonable expectations and consent of the individuals but must also be used for legitimate purposes (Future Advocacy, 2019).

*4.7.4 Choice.* AI should protect users’ power to decide about decisions in their lives (Floridi *et al.*, 2018). AI should not “compromise human freedom and autonomy by illegitimately and surreptitiously reducing options for and knowledge of citizens” (European Group on Ethics in Science and New Technologies, 2018, p. 17).

*4.7.5 Self-determination.* There needs to be a balance between decision-making power that is freely given by the user to the autonomous systems and when this option is taken away or undermined by the system (Floridi *et al.*, 2018). AI organisations should not manipulate individual’s self-determination, particularly those who may be vulnerable to abuse (Rathenau Institute, 2017, p. 26).

*4.7.6 Liberty.* AI organisations need to ensure that their AI protects individuals’ liberties, as outlined in many human rights legislations, such as the EU’s Charter of Fundamental Human Rights (2000) and the Universal Declaration of Human Rights (1948). Liberty refers to rights such as freedom of speech, freedom of assembly and freedom of movement. During the development of AI, there should be strong adherence to the protection of liberties, outlined in these fundamental human rights documents.

*4.7.7 Empowerment.* AI should be used to empower and strengthen our human rights, rather than curtailing or infringing upon them (ICDPPC, 2018). If decisions are made about individuals that may harm their liberties, they should be empowered with the right to challenge such decisions (ICO, 2017).

#### 4.8 Trust

Trust is such a fundamental principle for interpersonal interactions and is a foundational precept for society to function. Similarly, trust is being acknowledged as a key requirement for the ethical deployment and use of AI. The HLEG (2019) even use it as their defining paradigm for their ethics guidelines, referring to it throughout the entire document. It appears to be a relatively new phenomenon however, with most of the guidelines that make reference to trust coming after 2017.

*4.8.1 Trustworthiness.* AI organisations should prove they are trustworthy and that their technologies are reliable (Digital Decisions, 2019; MI Garage, 2019). End users should be able to justly trust AI organisations to fulfil their promises and to ensure that their systems function as intended (Deutsche Telekom, 2018; Institute of Business Ethics, 2018; Microsoft, 2018; Sony, 2018; NITI Aayog, 2018; and Microsoft, 2017). Building trust should be encouraged by ensuring accountability, transparency and safety of AI (Royal Society,



---

2017). Organisations can cultivate trust by demonstrating the security of their AI (Intel, 2017) and guard the data retrieved from these systems in a responsible way (Unity Blog, 2018).

#### 4.9 Sustainability

Sustainability is a key principle in global discussions at present, and its importance is only set to rapidly increase as a result of climate change predictions and ongoing environmental destruction. All fields and disciplines are affected and need to incorporate sustainability agendas, and AI is no exception. Despite this, it did not appear as an overly pressing concern in the majority of guidelines, demonstrating a greater need to identify how it can be incorporated more effectively.

*4.9.1 Sustainability.* AI organisations need to ensure that they are environmentally sustainable and incorporate environmental outcomes within their decision-making (Special Interest Group on Artificial Intelligence, 2018). There must be an adherence to resource-efficient, sustainable energy-promoting and the protection of biodiversity, by the AI.

*4.9.2 Environment (nature).* Organisations should use AI that has been developed in an environmentally conscious manner (SIIA, 2018). In situations where there is ecological harm caused by AI beyond acceptable levels, steps should be taken to either immediately halt it (temporarily or permanently), identify ways to use it in a non-harmful way or consult the designers for potential solutions and responses. AI should not be used to harm biodiversity (UNI Global 2017).

*4.9.3 Energy.* The use of AI should be respectful of energy efficiency, mitigate greenhouse gas emissions and protect biodiversity (University of Montreal, 2017). Those responsible for AI should ensure that its ecological footprint is minimal and all efforts are taken to reduce emission levels (Green Digital Working Group, 2016, p. 7).

*4.9.4 Resources (energy).* AI should be created in a way that ensures effective energy and resource consumption, promotes resource efficiency, the use of renewable materials, and reduction of use of scarce materials and minimal waste (European Parliament, 2017). Resource use and environmental impact should be held in importance in the life cycle impact assessment of AI (COMEST/UNESCO, 2017, p. 55).

#### 4.10 Dignity

Human dignity is the recognition that individuals have inherent worth and that their rights should be respected. It is important that AI does not infringe or harm the dignity of end users or other members within society. Respecting individuals' dignity is a vital principle that should be taken into account within AI ethics guidelines.

*4.10.1 Dignity.* Human beings have intrinsic value and developers/organisational users should ensure that this is respected in the design and use of AI (The Conference toward AI Network Society, 2017). AI should be developed and used in a way that “respects, serves and protects humans” physical and mental integrity, personal and cultural sense of identity, and satisfaction of their essential needs” (High-Level Expert Group on AI, 2019, p. 10). AI needs to be developed and used in a way that makes it clear to the user that they are interacting with AI and not another human being (EGE, 2018). Efforts need to be made to ensure that AI is not confused with human beings, as dignity is a value inherent to human beings (COMEST/UNESCO, 2017, p. 50). Organisations should ensure that their AI does not violate the end-user's dignity and should closely follow the principle of dignity outlined in the first chapter of the EU Charter (Latonero, 2018).

#### 4.11 Solidarity

With the widespread use of AI to disseminate fake news, its potential to surveil and invade individuals' privacy, there is a growing concern that AI may be used to undermine and jeopardise societal relationships and solidarity. It is important to consider if the AI supports rich and meaningful social interaction, both professionally and in private life, and not support segregation and division, within the design and development process. AI should promote social security and cohesion and should not jeopardise societal bonds and relationships.

*4.11.1 Solidarity.* AI should be developed to promote, or avoid harm to, societal bonds and relationships between people and generations (University of Montreal, 2017). AI should facilitate and promote human development, rather than being designed to obstruct or endanger it (ICDPPC, 2018). There should be consideration towards preserving and promoting solidarity and should not undermine existing social structures (Floridi *et al.*, 2018). AI should not create "social dislocation", whereby it adversely harm cultural and social identity, and those organisations that cause it should be held responsible (Accenture, 2019).

*4.11.2 Social security.* Democratic values should not be jeopardised as a result of AI use and citizens should receive accurate and impartial information without interference or manipulation for political purposes (EGE, 2018). AI should not be developed or used to undermine electoral and political decision-making (High-Level Expert Group on AI, 2019). This can be done by ensuring that democratic values are promoted in AI development and implementation (EGE, 2018).

*4.11.3 Cohesion.* AI organisations should promote fair distribution of benefits from AI to ensure social cohesion is not harmed (Koski and Husso, 2018, p. 51). The use of AI should contribute to global justice, in the aim to promote social cohesion and solidarity (European Group on Ethics in Science and New Technologies, 2018, p. 17). AI teams should not develop or use these technologies in a way that knowingly undermines "functioning democratic systems of government" (Unity Blog, 2018). AI organisations should actively develop strategies with academia, civil society and industry partners, to promote social cohesion and knowledge-exchange collaborations (Privacy International/Article 19, 2018, p. 29).

## 5. Discussion and conclusion

Maybe the first impression arising from this long and, as we hope, comprehensive overview of AI ethics guidance is that there is a diversity of ethical principles, issues and concerns that are covered by a large number of guidelines. Even focusing on organisational users and developers and leaving out stakeholder groups like policymakers, as we have done in this paper, the list is impressive. One of the points of criticism sometimes levelled at the dominant approach to principle-based guidelines is that they can oversimplify complex and difficult ethical debates and lead to an appearance of moral consensus where in fact the difficult ethical questions are hidden in the details of the application of principles (Mittelstadt, 2019). We hope that our work goes some way towards addressing this concern. In determining the constituent ethical issues and identifying normative positions arising from these, we provide a rich overview of guidance that is available to developers and users of AI. We believe that this is valuable to academic researchers and individuals who develop or revise ethical guidelines. By providing a comprehensive set of guidelines these stakeholders can now assess the completeness of their work. We are not suggesting that there should only be one set of guidelines that cover everything, but scholars working on guidelines, e.g. for particular application areas, can use our work to validate their work.

---

Furthermore, we believe that the guidelines can be useful to developers and users who would like to understand the ethical challenges they can face and should be prepared to engage with. Our contribution is thus both academic/theoretical and practical.

Having said this, we realise that this paper can only be an intermediary step to a more manageable and practical set of guidance. One step that should now be undertaken is to do a philosophical and conceptual analysis of the guidance provided in these documents. Key questions to be asked include how the individual principles and their constituent issues can be justified from a philosophical perspective. This analysis should form part of a greater check for consistency of the guidelines. We have identified and categorised the components of many existing guidelines, but we have not checked whether and to what degree these are consistent.

A further step will be the exploration of potential conflicts between individual principles. A typical example well discussed in the literature are conflicts between privacy and transparency. Many other conflicts are easily imaginable. A practical set of guidelines that developers and users of AI can apply in practice needs to be aware of such conflicts and provide mechanisms for identifying them and dealing with them in an appropriate way.

On their way to being truly practicable, guidelines also need to go even further into detail than we have and at least provide pointers to ways of realising and implementing normative statements. It is important to understand that one should do X, but at the same time, this is not helpful, if one does not know *how* to do X. There are by now large numbers of tools and initial attempts to collect and categories them that help address various ethical issues of AI ([Morley et al., 2019](#)). What is required now is to map these tools against the ethical guidelines to allow individuals and organisations to adopt these norms in practice.

And, finally, there needs to be ways to integrate the guidelines as presented here to address ethical issues in AI. They may, for example, find their way into standards, they can form parts of corporate or industry governance mechanisms, they can be reflected in legislation and regulation and be enforced by regulators. We tried in this paper to provide a detailed account of guidance that is available to developers and users, but we realise that these are unlikely to have much practical effect, if they simply remain aspirational documents which, to exacerbate matters, are long, wordy and difficult to digest.

It is thus clear that this paper can only be one step in a longer journey towards a more comprehensive approach to dealing with the ethics of AI. We hope to have shown, however, that this step is a crucially important one that is required to progress both theoretically and practically. In this spirit we hope that the paper finds a broad audience and can provide the input into the next steps that are no doubt required.

## Notes

1. Throughout the paper, we will simply refer to developers and users of AI systems as “AI organisations” for convenience’s sake.
2. The additional nine guidelines were ADMA 2013; [Algo.Rules 2019](#); [B Debate 2017](#); [Council of Europe 2017](#); [IPC Ontario 2017](#); [OECD 2019](#); [Personal Data Protection Commission Singapore 2019](#); UK Government, Department for Digital, Culture, Media & Sport 2018 and [UNDG 2017](#).

## References

- Accenture (2019), “Responsible AI and robotics: an ethical framework”.
- ACM (2017), “Statement on algorithmic transparency and accountability”.

- Association for Data-driven Marketing and Advertising (ADMA) (2013), "Best practice guideline: Big Data".
- AGID (2018), "L'Intelligenza artificiale al servizio del cittadino".
- AI for Humanity (2018), "For a meaningful artificial intelligence: towards a French and European strategy".
- AI Now Institute (2016), "The AI now report: the social and economic implications of artificial intelligence technologies in the near-Term".
- AI Now Institute (2017), "AI Now 2017 report".
- AI Now Institute (2018), "AI Now report 2018".
- Algo.Rules (2019), "Rules for the design of algorithmic systems".
- Allistene (2014), "Éthique de la recherche en Robotique".
- AMA (2018), "Policy recommendations on augmented intelligence in health care H-480.940".
- American College of Radiology (2019), "Ethics of AI in radiology: european and North American multisociety statement".
- Amnesty International/Access Now (2018), "The Toronto declaration: protecting the rights to equality and non-discrimination in machine learning systems", available at: [www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration\\_ENG\\_08-2018.pdf](http://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf)
- Aronson, J. (1995), "A pragmatic view of thematic analysis", *The Qualitative Report*, Vol. 2, pp. 1-3.
- Asilomar Conference (2017), "Asilomar AI principles", Future of Life Institute, available at: <https://futureoflife.org/ai-principles/> (accessed 10 February 2018).
- B Debate (2017), "Barcelona declaration for the proper development and usage of artificial intelligence in Europe".
- Beauchamp, T.L. and Childress, J.F. (1979), *Principles of Biomedical Ethics*, Oxford University Press, Oxford.
- Berendt, B. (2019), "AI for the common good?! pitfalls, challenges, and ethics pen-testing", *Paladyn, Journal of Behavioral Robotics*, Vol. 10 No. 1, pp. 44-65, available at: <https://doi.org/10.1515/pjbr-2019-0004>
- Braun, V. and Clarke, V. (2006), "Using thematic analysis in psychology", *Qualitative Research in Psychology*, Vol. 3 No. 2, pp. 77-101.
- Bynum, T.W. (2010), "The historical roots of information and computer ethics", in Floridi, L. (Ed.), *The Cambridge Handbook of Information and Computer Ethics*, Cambridge University Press, pp 20-38.
- Bynum, T.W. and Rogerson, S. (2003), *Computer Ethics and Professional Responsibility: Introductory Text and Readings*, Wiley Blackwell.
- Capurro, R. (2008), "On Floridi's metaphysical foundation of information ecology", *Ethics and Information Technology*, Vol. 10 Nos 2/3, p. 167.
- Cerna Collectif (2018), "Research ethics in machine learning".
- Clark, R. (2019), "Principles for AI: a SourceBook".
- COMEST/UNESCO (2017), "Report of COMEST on robotics ethics".
- Council of Europe (2017), "Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data".
- Council of Europe (2019), "European ethical charter on the use of artificial intelligence in judicial systems and their environment".
- Datatilsynet (2018), "Artificial intelligence and privacy".
- Dawson, D. et al. (2019), "Artificial intelligence: Australia's ethics framework, Australian Government".

- 
- DeepMind (2017), “DeepMind ethics and society principles”.
- Demiaux, V. and Abdallah, Y.S. (2017), “How can humans keep the upper hand? the ethical matters raised by algorithms and artificial intelligence”, Report on the public debate led by the French Data Protection Authority (CNIL) as part of the ethical discussion assignment set by the Digital Republic Bill. CNIL, Paris.
- Deutsche Telekom (2018), “Guidelines for artificial intelligence”.
- Digital decisions (2019), “Center for democracy and technology”, available at: <https://cdt.org/issue/privacy-data/digital-decisions/>
- EPSRC (2011), “Principles of robotics”.
- European Group on Ethics in Science and New Technologies (2018), “Statement on artificial intelligence, robotics and ‘autonomous’ systems”.
- European Parliament (2017), “Report with recommendations to the commission on civil law rules on robotics”.
- Executive Office of the President (2016a), “Preparing for the future of artificial intelligence. Executive office of the president national science and technology council committee on technology”.
- Executive Office of the President (2016b), “Artificial intelligence, automation, and the economy. Executive office of the president national science and technology council committee on technology”.
- FATML (2016), “Principles for accountable algorithms and a social impact statement for algorithms”.
- Federal Ministry of Transport and Digital Infrastructure (2017), “Ethics commission: automated and connected driving”.
- Floridi, L. (2019), “Establishing the rules for building trustworthy AI”, *Nature Machine Intelligence*, Vol. 1 No. 6, available at: <https://doi.org/10.1038/s42256-019-0055-y>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... Schafer, B. (2018), “AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations”, *Minds and Machines*, Vol. 28 No. 4, pp. 689-707.
- Future Advocacy (2019), “Ethical, social, and political challenges of artificial intelligence in health”, in Fenech, M., Strukelj, N. and Buston, O. (Eds).
- Future of Life Institute (2017), “Asilomar AI principles”.
- Gilbert, B. (2019), “Women leading in AI: 10 principles of responsible AI”, Towards Data Science, available at: <https://towardsdatascience.com/women-leading-in-ai-10-principles-for-responsible-ai-8a167fc09b7d>
- Government of the Republic of Korea (2017), “Mid- to Long-Term master plan in preparation for the intelligent information society: Managing the fourth industrial revolution”.
- Green Digital Working Group (2016), “Position on robotics and artificial intelligence”.
- Haenlein, M. and Kaplan, A. (2019), “A brief history of artificial intelligence: on the past, present, and future of artificial intelligence”, *California Management Review*, Vol. 61 No. 4, pp. 5-14.
- High-Level Expert Group on AI (2019), “Ethics guidelines for trustworthy AI”.
- HoL (2018), “AI in the UK: ready, willing and able? select committee on artificial intelligence”, London.
- House of Commons Science and Technology Committee (2016), “Robotics and artificial intelligence”.
- IBM (2017), “Transparency and trust in the cognitive era”, available at: [www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/](http://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/)
- IBM (2018), “Everyday ethics for artificial intelligence”.
- International Conference of Data Protection and Privacy Commissioners (ICDPPC) (2018), “Declaration on ethics and data protection in artificial intelligence”, available at: [https://icdppc.org/wp-content/uploads/2018/10/20180922\\_ICDPPC-40th\\_AI-Declaration\\_ADOPTED.pdf](https://icdppc.org/wp-content/uploads/2018/10/20180922_ICDPPC-40th_AI-Declaration_ADOPTED.pdf)
- ICO (2017), “Big data, artificial intelligence, machine learning and data protection”.

- IEEE (2017), "Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent systems", Version 2.
- IEEE (2019), "Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent systems", Version 1.
- IIIM (2015), "Ethics policy".
- IPC Ontario (Information and Privacy Commissioner of Ontario) (2017), "Big data guidelines".
- Institute of Business Ethics (2018), "Business ethics and artificial intelligence".
- Intel (2017), "Artificial intelligence: the public policy opportunity".
- Intel (2018), "Intel's AI privacy policy white paper: protecting individuals' privacy and data in the artificial intelligence world".
- Internet Society (2017), "Artificial intelligence and machine learning: policy paper".
- ITI (2017), "ITI AI policy principles".
- Japanese Society for Artificial Intelligence (2017), "The Japanese society for artificial intelligence ethical guidelines".
- Jobin, A., Ienca, M. and Vayena, E. (2019), "The global landscape of AI ethics guidelines", *Nature Machine Intelligence*, Vol. 1 No. 9, pp. 389-399, available at: <https://doi.org/10.1038/s42256-019-0088-2>
- Johnson, K., Pasquale, F. and Chapman, J. (2019), "Artificial intelligence, machine learning, and bias in finance: toward responsible innovation", *Fordham Law Review*, Vol. 88, p. 499.
- Kaplan, A. and Haenlein, M. (2019), "Siri, Siri, in my hand: Who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence", *Business Horizons*, Vol. 62 No. 1, pp. 15-25.
- Koski, O. and Husso, K. (2018), "Work in the age of artificial intelligence, ministry of economic affairs and employment".
- Latonero, M. (2018), "Governing artificial intelligence: upholding human rights and dignity", Data and Society.
- Leaders of the G7 (2018), "Common vision for the future of artificial intelligence".
- MI Garage (2019), "Ethics framework".
- Microsoft (2017), "Microsoft AI principles", Microsoft.
- Microsoft (2018), "Responsible bots: 10 guidelines for developers of conversational AI".
- Ministry of Economic Affairs and Employment (2018), "Work in the age of artificial intelligence".
- Ministry of State for Science and Technology Policy (2019), "Report on artificial intelligence and human society: Unofficial translation".
- Mittelstadt, B. (2019), "Principles alone cannot guarantee ethical AI", *Nature Machine Intelligence*, available at: <https://doi.org/doi:10.1038/s42256-019-0114-4>
- Moor, J.H. (1985), "What is computer ethics", *Metaphilosophy*, Vol. 16 No. 4, pp. 266-275.
- Morley, J., Floridi, L., Kinsey, L. and Elhalal, A. (2019), "From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices", *Science and Engineering Ethics*, available at: <https://doi.org/10.1007/s11948-019-00165-5>
- NITI Aayog (2018), "Discussion paper: National strategy for artificial intelligence".
- NSTC (2016a), "Preparing for the future of artificial intelligence".
- NSTC (2016b), "The national artificial intelligence research and development strategic plan".
- O'Neil, C. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Publishers.
- OECD (2019), "Recommendation of the council on artificial intelligence".



- 
- OP (2019), “Commitments and principles”.
- OpenAI (2018), “OpenAI charter”.
- Personal Data Protection Commission Singapore (2018), “Discussion paper on AI and personal data — fostering responsible development and adoption of AI”.
- Personal Data Protection Commission Singapore (2019), “A proposed model artificial intelligence governance framework”.
- Privacy International/Article 19 (2018), “Privacy and freedom of expression in the age of artificial intelligence”.
- PwC (2019), “The responsible AI framework”.
- Rathenau Institute (2017), “Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality”.
- RCP London (2018), “Artificial intelligence (AI) in health”.
- Royal Society (2017), “Machine learning: the power and promise of computers that learn by example”.
- Ryan, M. (2019a), “Ethics of using AI and big data in agriculture: the case of a large agriculture multinational”, *ORBIT Journal*, Vol. 2 No. 2, available at: <https://doi.org/10.29297/orbit.v2i2.109>
- Ryan, M. (2019b), “Ethics of public use of AI and big data”, *ORBIT Journal*, Vol. 2 No. 1, available at: <https://doi.org/10.29297/orbit.v2i1.101>
- Ryan, M. (2019c), “The future of transportation: ethical, legal, social and economic impacts of self-driving vehicles in the year 2025”, *Science and Engineering Ethics*, available at: <https://doi.org/10.1007/s11948-019-00130-2>
- Ryan, M. and Gregory, A. (2019), “Ethics of using smart city AI and big data: the case of four large European cities”, *ORBIT Journal*, Vol. 2 No. 2, available at: <https://doi.org/10.29297/orbit.v2i2.110>
- Sage (2017), “The ethics of code: Developing AI for business with five core principles”.
- SAP (2018), “SAP’s guiding principles for artificial intelligence (AI)”.
- SIIA (2017), “Ethical principles for artificial intelligence and data analytics”.
- Smart Dubai (2019), “Artificial intelligence principles and ethics”.
- Sony (2018), “Sony group AI ethics guidelines”.
- Special Interest Group on Artificial Intelligence (2018), “Dutch artificial intelligence manifesto”.
- Stix, C. (2019), *A Survey of the European Union’s Artificial Intelligence Ecosystem*, Leverhulme Centre for the Future of Intelligence, Cambridge.
- Telefónica (2018), “AI principles of telefónica”.
- The Conference toward AI Network Society (2017), “Draft AI R&D guidelines for international discussions”.
- The Future Society (2018), “Science, law and society (SLS) initiative”.
- The Information Accountability Foundation (2015), “Unified ethical frame for big data analysis: IAF big data ethics initiative, part A”.
- The Partnership on AI (2016), “Tenets”.
- The Public Voice (2018), “Universal guidelines for artificial intelligence”.
- Tieto (2018), “Tieto’s AI ethics guidelines”.
- UK Government (2018), “Department for digital, culture, media and sport, data ethics framework”, available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/737137/Data\\_Ethics\\_Framework.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/737137/Data_Ethics_Framework.pdf)
- UNI Global Union (2017), “Top 10 principles for ethical AI”.
- United Nations (1948), “Universal declaration of human rights, (general assembly resolution 217 a)”.

- 
- United Nations Development Group (UNDG) (2017), "Data privacy, ethics and protection. Guidance note on big data for achievement of the 2030 agenda".
- Unity Blog (2018), "Introducing unity's guiding principles for ethical AI".
- University of Montreal (2017), "Montreal declaration for a responsible development of artificial", "Intelligence", available at: [www.montrealdeclaration-responsibleai.com/the-declaration](http://www.montrealdeclaration-responsibleai.com/the-declaration)
- WEF (2018), "White paper: How to prevent discriminatory outcomes in machine learning".
- Wiener, N. (1954), "The human use of human beings", Doubleday.
- World Wide Web Foundation (2018), "Artificial intelligence: Open questions about gender inclusion".

### Further reading

- Australian Government (2019), "Artificial intelligence: Australia's ethics framework".
- Bertelsmann Stiftung and iRights.Lab (2019), "Algo.Rules – Rules for the design of algorithmic systems".
- Center for Democracy and Technology (2019), "Digital decisions".
- CNIL (2017), "How can humans keep the upper hand? the ethical matters raised by algorithms and artificial intelligence".
- Commission Nationale de l'Informatique et des Libertés, European Data Protection Supervisor and Garante per la protezione dei dati personali (2018), "Declaration on ethics and data protection in artificial intelligence".
- European Union (2000), "Charter of fundamental human rights (2000/C 364/01)".
- Future of Life Institute (2018), "National and international AI strategies".
- Google AI (2019), "Artificial intelligence at google: our principles".
- GOV UK (2018), "Data ethics framework".
- GOV UK (2019), "Initial code of conduct for data-driven health and care technology".
- HLEG on AI HEG on AI (2019), "Ethics guidelines for trustworthy AI. European Commission - Directorate-General for communication", Brussels.
- Towards Data Science (2019), "Women leading in AI: 10 principles of responsible AI".

### Corresponding author

Mark Ryan can be contacted at: [mryan@kth.se](mailto:mryan@kth.se)

Appendix

Name of Document/Website	Issuer	Country
Artificial Intelligence. Australia's Ethics Framework: A Discussion Paper	Department of Industry Innovation and Science	Australia
Best practice guideline: Big Data	Association for Data-driven Marketing and Advertising (ADMA)	Australia
Montréal Declaration: Responsible AI	Université de Montréal	Canada
Big Data Guidelines	IPC Ontario (Information and Privacy Commissioner of Ontario)	Canada
Work in the Age of Artificial Intelligence. Four Perspectives on the Economy, Employment, Skills and Ethics	Ministry of Economic Affairs and Employment	Finland
Tieto's AI Ethics Guidelines	Tieto	Finland
Commitments and Principles	OP Group	Finland
How Can Humans Keep the Upper Hand? Report on the Ethical Matters Raised by AI Algorithms	French Data Protection Authority (CNIL)	France
For a Meaningful Artificial Intelligence. Towards a French and European Strategy	AI for Humanity	France
Ethique de la Recherche en Robotique	CERNA (Allistene)	France
AI Guidelines	Deutsche Telekom	Germany
SAP's Guiding Principles for Artificial Intelligence	SAP	Germany
Automated and Connected Driving: Report	Federal Ministry of Transport and Digital Infrastructure, Ethics Commission	Germany
Rules for the Design of Algorithmic Systems	Algo.Rules	Germany
Ethics Policy	Icelandic Institute for Intelligent Machines (IIIM)	Iceland
Discussion Paper: National Strategy for Artificial Intelligence	National Institution for Transforming India (NITI Aayog)	India
L'intelligenza Artificiale al Servizio del Cittadino	Agenzia per l'Italia Digitale (AGID)	Italy
The Japanese Society for Artificial Intelligence Ethical Guidelines	Japanese Society for Artificial Intelligence	Japan
Report on Artificial Intelligence and Human Society (unofficial translation)	Advisory Board on Artificial Intelligence and Human Society (initiative of the Minister of State for Science and Technology Policy)	Japan
Draft AI R&D Guidelines for International Discussions	Institute for Information and Communications Policy (IICP), The Conference toward AI Network Society	Japan
Sony Group AI Ethics Guidelines	Sony	Japan
Human Rights in the Robot Age Report	The Rathenau Institute	Netherlands
Dutch Artificial Intelligence Manifesto	Special Interest Group on Artificial Intelligence (SIGAI), ICT Platform Netherlands (IPN)	Netherlands
Artificial Intelligence and Privacy Discussion Paper on Artificial Intelligence (AI) and Personal Data—Fostering Responsible Development and Adoption of AI	The Norwegian Data Protection Authority Personal Data Protection Commission Singapore	Norway Singapore

(continued)

**Table A1.**  
Ethics guidelines  
used

Name of Document/Website	Issuer	Country
A Proposed Model Artificial Intelligence Governance Framework	Personal Data Protection Commission Singapore	Singapore
Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society	Government of the Republic of Korea	South Korea
AI Principles of Telefónica	Telefónica	Spain
Barcelona Declaration for the Proper Development and Usage of Artificial Intelligence in Europe	B Debate	Spain
AI Principles and Ethics	Smart Dubai	UAE
Principles of robotics	Engineering and Physical Sciences Research Council UK (EPSRC)	UK
The Ethics of Code: Developing AI for Business with Five Core Principles	Sage	UK
Big Data, Artificial Intelligence, Machine Learning and Data Protection	Information Commissioner's Office	UK
DeepMind Ethics and Society Principles	DeepMind Ethics and Society	UK
Business Ethics and Artificial Intelligence	Institute of Business Ethics	UK
AI in the UK: Ready, Willing and Able?	UK House of Lords, Select Committee on Artificial Intelligence	UK
Artificial Intelligence (AI) in Health	Royal College of Physicians	UK
Initial Code of Conduct for Data-Driven Health and Care Technology	UK Department of Health and Social Care	UK
Department for Digital, Culture, Media and Sport, Data Ethics Framework	UK Government	UK
Ethics Framework: Responsible AI	Machine Intelligence Garage Ethics Committee	UK
The Responsible AI Framework	PricewaterhouseCoopers UK	UK
Responsible AI and Robotics. An Ethical Framework.	Accenture UK	UK
Machine Learning: The Power and Promise of Computers that Learn by Example	The Royal Society	UK
Ethical, Social, and Political Challenges of Artificial Intelligence in Health	Future Advocacy	UK
Unified Ethical Frame for Big Data Analysis. IAF Big Data Ethics Initiative, Part A	The Information Accountability Foundation	USA
The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term	AI Now Institute	USA
Statement on Algorithmic Transparency and Accountability	Association for Computing Machinery (ACM)	USA
AI Principles	Future of Life Institute	USA
AI—Our Approach	Microsoft	USA
Artificial Intelligence. The Public Policy Opportunity	Intel Corporation	USA
IBM's Principles for Trust and Transparency	IBM	USA
OpenAI Charter	OpenAI	USA

Table A1.

(continued)

Name of Document/Website	Issuer	Country
Our Principles	Google	USA
Policy Recommendations on Augmented Intelligence in Health Care H-480.940	American Medical Association (AMA)	USA
Everyday Ethics for Artificial Intelligence. A Practical Guide for Designers and Developers	IBM	USA
Governing Artificial Intelligence. Upholding Human Rights and Dignity	Latonero <i>et al.</i>	USA
Intel's AI Privacy Policy White Paper. Protecting Individuals' Privacy and Data in the Artificial Intelligence World	Intel Corporation	USA
Introducing Unity's Guiding Principles for Ethical AI—Unity Blog	Unity Technologies	USA
Digital Decisions	Center for Democracy and Technology	USA
Science, Law and Society (SLS) Initiative	The Future Society	USA
AI Now 2018 Report	AI Now Institute	USA
Responsible Bots: 10 Guidelines for Developers of Conversational AI	Microsoft	USA
Preparing for the Future of Artificial Intelligence	Executive Office of the President; National Science and Technology Council; Committee on Technology	USA
The National Artificial Intelligence Research and Development Strategic Plan	National Science and Technology Council; Networking and Information Technology Research and Development Subcommittee	USA
AI Now 2017 Report	AI Now Institute	USA
Position on Robotics and Artificial Intelligence	The Greens (Green Working Group Robots)	EU
Report with Recommendations to the Commission on Civil Law Rules on Robotics	European Parliament	EU
Ethics Guidelines for Trustworthy AI	High-Level Expert Group on Artificial Intelligence	EU
AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations	AI4People	EU
European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment	Council of Europe: European Commission for the Efficiency of Justice (CEPEJ)	EU
Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data	Council of Europe: European Commission for the Efficiency of Justice (CEPEJ)	EU
Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems	European Commission, European Group on Ethics in Science and New Technologies	EU
Artificial Intelligence and Machine Learning: Policy Paper	Internet Society	International
Report of COMEST on Robotics Ethics	COMEST/UNESCO	International
Ethical Principles for Artificial Intelligence and Data Analytics	Software and Information Industry Association (SIIA), Public Policy Division	International

(continued)

Table A1.

Name of Document/Website	Issuer	Country
ITI AI Policy Principles	Information Technology Industry Council (ITI)	International
Ethically Aligned Design. A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2	Institute of Electrical and Electronics Engineers (IEEE), The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	International
Top 10 Principles for Ethical Artificial Intelligence	UNI Global Union	International
The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation	Future of Humanity Institute; University of Oxford; Centre for the Study of Existential Risk; University of Cambridge; Center for a New American Security; Electronic Frontier Foundation; OpenAI	International
White Paper: How to Prevent Discriminatory Outcomes in Machine Learning	WEF, Global Future Council on Human Rights 2016–2018	International
Privacy and Freedom of Expression in the Age of Artificial Intelligence	Privacy International and Article 19	International
The Toronto Declaration: Protecting the Right to Equality and Non-discrimination in Machine Learning Systems	Access Now; Amnesty International	International
Charlevoix Common Vision for the Future of Artificial Intelligence	Leaders of the G7	International
Artificial Intelligence: Open Questions About Gender Inclusion	W20	International
Declaration on Ethics and Data Protection in Artificial Intelligence	ICDPPC	International
Universal Guidelines for Artificial Intelligence	The Public Voice	International
Ethics of AI in Radiology: European and North American Multisociety Statement	American College of Radiology; European Society of Radiology; Radiology Society of North America; Society for Imaging Informatics in Medicine; European Society of Medical Imaging Informatics; Canadian Association of Radiologists; American Association of Physicists in Medicine	International
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), The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	International
Recommendation of the Council on Artificial Intelligence	OECD	International
Data Privacy, Ethics and Protection. Guidance Note on Big Data for Achievement of the 2030 Agenda	United Nations Development Group (UNDG)	International
Tenets	Partnership on AI	N/A
Principles for Accountable Algorithms and a Social Impact Statement for Algorithms	Fairness, Accountability, and Transparency in Machine Learning (FATML)	N/A
10 Principles of Responsible AI	Women Leading in AI	N/A

**Table A1.**