In developing intelligent computing, issues of ethics are central. What are the main challenges around these issues? How can the developer of intelligent systems be aware of and reason about these issues in the development processes? How can the user be aware of these issues and how to deal with breaches of them?

Humans and Intelligent machines CW2

2021

1 Introduction

With AI taking an ever more prominent position in our society, the area of AI ethics has emerged as a response to the possible harm AI systems could have on the individual, or society, due to the potential for unintended consequences. The field of AI ethics aims to deal with any abuse, misuse, poor design or unintended consequences which can result from their uses[1], through the notion of AI for the 'common good'[2]. This notion of 'good' can be interpreted in many different ways depending on where you come from and the ethical principles you apply. There are many ethical theories; the three most prominent ones being consequentialism, deontology and virtue ethics[3]. Consequentialism concerns itself with the outcome being 'good'[3], deontology concerns itself with whether the action is 'good'[3], and virtue ethics considers something as 'good' based upon its virtues[3, 4]. Irrespective of the ethical principles used, the AI systems used are generally designed for common good trying to benefit society as a whole[2].

As Viginia[5] suggests the way in which ethical values are applied is dependent on the socio-cultural context. How these differences are present can be demonstrated by the differing responses to the trolley problem[6]. The differing ethical values of different parts of the world cause problems as what is seen as an ethical issue in one place, may not be seen as an ethical issue in another. An example of this is the implementation of the Chinese social credit system which is considered acceptable in China[7], whereas in Western countries it seen as too much of an invasion of privacy[7]. This difference in perspective is likely due to China's concept of putting the good of the people ahead of the good of the individual[8].

2 What are the main challenges around these issues?

There are many different ethical issues when developing and implementing AI systems, Stahl[3] suggest a potential list of 39 ethical issues created from a study[9], these ethical issues range from "lack of transparency" to "lack of power to frame dialogue". All of these ethical issues have challenges which arise at different points in the development process with the challenge being in how to prevent the ethical issbue from coming to fruition. To understand what the main challenges around ethical issues are for AI systems, we will consider a range of different ethical issues and the challenges which need to be overcome in order to make the AI system ethical.

AI systems are created to learn patterns in data, this can lead to AI systems becoming discriminatory if the input data is of poor quality, due to its age meaning that it no longer describes the population well, or if it insufficiently represents groups in the population to which the AI system is applied. Kodiyan[10] suggests that while this may be the case, AI systems are designed to reproduce pattens in the data which could result in the biasing of AI systems, as was seen with the Amazon hiring algorithm[10]. The removal of what may be preserved as sources of discrimination, such as race or gender, might not prevent discrimination occurring. This has been seen with multiple predictive policing algorithms, such as PredPol[11], which consider 'nuisance crimes' when creating a 'heat map' of crime, leading to a feedback loop being created, whereby the AI system takes on the bias of individual police officers, creating a lack of trust by a group of people as suggested by O'neil[11] due to them perceiving it as though they are being unfairly targeted. AI systems are created by companies which see them as their intellectual property, this means, as Stahl[3] suggests, that it is harder to address the issues of bias and discrimination, due their lack of openness over when and how AI systems are being used. This lack of willingness on the part of the company to be open and transparent prevents the users from being able to understand how the outcomes are achieved[3, 12], and the users' inability to determine when discrimination occurs means that such systems may not be held to account when thing go wrong.

As more and more information about us is available, issues of privacy arise; the privacy of the individual has multiple different aspects which create challenges. The main challenge which we will consider is that of user consent and control of their data; people generate large amounts of data which can be available with or without their knowledge[13] and used without the individual's consent[14]. This challenge of providing the user with control over their information has led to the creation of regulation such as GDPR[15] which is designed to enable the user to regain control of their data. However Brandimarte[16] suggest that this creates a 'control paradox' whereby if an individual wants to use a service, although they may have control over the data they disclose, they have no control over what it is used for. This lack of control for the individual over how their data is used, along with their lack of understanding as to how the data is used[4], means that holding those accountable for their misuses is complicated. The ethical issue of privacy is not limited to the collection of the individual's data but also encompasses its use. As more information about the individual becomes available, this has enabled data to be taken without the user's consent, which has been done to create psychological profiles of the individual, as seen in the Cambridge Analytica scandal[17, 18]. These psychological profiles allow for the targeting of ads so as to 'nudge' the individual into making a decision without them realising it, this was seen in the 2016 US election in attempts to influence the behaviours of individuals[17, 19]. While ad pushing is often seen as ethically permissible, there is a fine line between this and what is no longer seen as ethically permissible as it infringes upon the individual's right to lead a private life[1].

The wilful neglect to consider the ethical issues can be seen as unethical deployment of AI. This may stem from AI ethics being seen as a 'box ticking' exercise[12], however what may be seen as an unethical deployment of AI may also be considered as ethical due to one's interpretation of the ethical issues at hand. As AI ethical issues are understood in different ways around the world, to understand how this is a challenge we will consider China's social credit system. The Chinese social credit system proposes to measure an individual's sincerity, honesty and integrity, and then use these results to help to determine aspects of people's lives[7]. While Síthigh[7] suggest that such systems may create a "culture of integrity", such a system is likely to go against the understanding of the ethical issues of bias and discrimination, and privacy as outlined above. The main challenges for such a system are the prevention of discrimination and invasion of privacy, as the large amounts of data needed can be collected without the individuals being able to prevent it[20]. This invasion of the individual's privacy

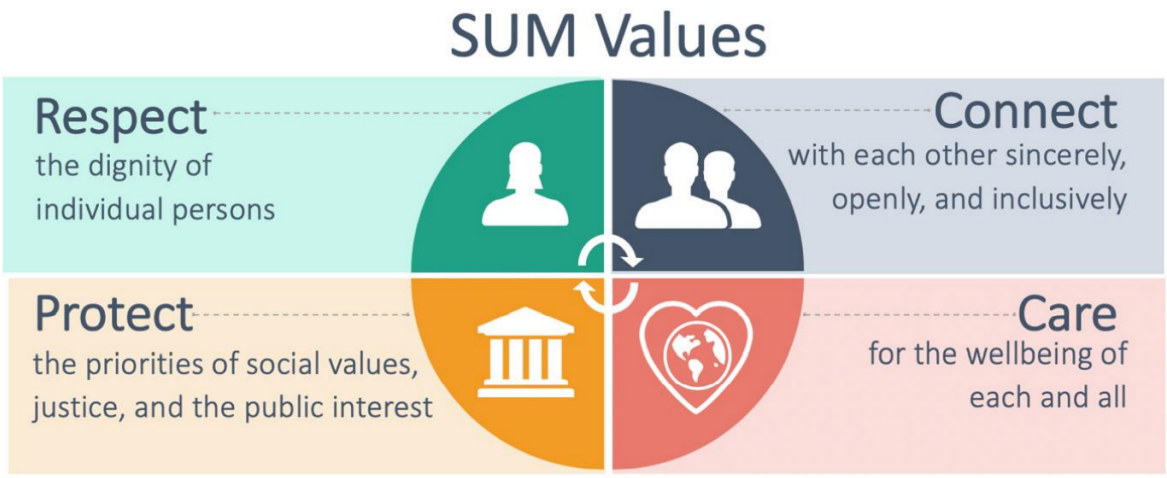

Figure 1: The SUM Values and how they interact with one another. Reproduced form [1].

can have the knock on effect that they could be discriminated against for something that they have done (such as being a member of a trade union)[7]. When issues arise, such systems, involving both Governments and the private sector can result in a blurring of the lines between them creating a lack of accountability[20]. Whilst such systems appear to be accepted by the Chinese population[7] they have been banned in other places such as the EU(([21])), this creates challenges as if the Chinese government decided to extend the use of such systems beyond their borders, applying it to those residing in other countries[7].

3 How can the developer of intelligent systems be aware of and reason about these issues in the development processes?

During the development process of AI systems the developer has the responsibility of considering the ethical issues and their challenges. When considering the developer not only does the individual developer need to be considered but also the company that is developing the AI system. A potential method for the individual developer to be aware of ethical issues would be by subscribing to an ethical code such as that produced by IEEE[22], alternatively they could follow an ethical framework such as 'The Aletheia Framework™' produced by Rolls-Royce[23]. However, as previously mentioned Mittelstadt[12] suggests that companies may make the individual developer prioritise the company's benefits over those of the users. This pressure by companies on the individual developer can lead to AI ethics being seen as an afterthought[12] and therefore these ethical frameworks, or ethical codes, could well be seen as a box ticking exercise. Vayena[13] suggests that ethical issues are context sensitive, for instance, whilst data could be used ethically in a medical setting, to use that same data would be unethical in a non-medical setting; thus developers have to consider their individual fields. Though there are ways of enforcing ethical discussions as is done in medicine through the use of ethics boards, a better method may be through the reflection of the individual by the application of values, such as 'the SUM values' as suggested by Leslie[1]. The SUM values prompt the developer to reflect on what they are creating and its implications (the SUM values are presented in Figure 1 with the values being taken from bioethics and human rights discourse[1]).

4 How can the user be aware of these issues and how to deal with breaches of them?

It is extremely hard for users to be aware of ethical issues as they are occurring in AI systems as these systems are often developed privately and as such users are generally unaware of when and how AI systems are used. This lack of transparency makes it complicated for the users to challenge any decisions made by AI systems or their application, as has been seen in multiple different failures of hiring systems[11, 24] this is generally due to a lack of data or its misinterpretation. As there is a general lack of understanding on the part of the user, companies and governments have a responsibility to ensure that users can be aware of the ethical issues, however Mittelstadt[12] suggests that existing initiatives to establish an ethical code which have been instigated by industry can be seen as 'virtue-signalling'.

As users have very little ability to be aware of the ethical issues, let alone ways in which to deal with any breaches in them, and the means of holding those responsible accountable is a great challenge. This issue of accountability becomes problematic, as without coherent regulation by the government, companies are only liable to suffer reputational risks, which as suggested by Mittelstadt[12] only carry weight while in the public consciousness. Because of this ethical issues are often ignored, meaning that those affected "must rely on the personal conviction of developers[12]". This is slowly changing as regulation is introduced to attempt to increase the transparency/explainability of when, where and how AI systems are used([21]). This change has already begun in the UK and the EU with the introduction of GDPR to regulate data and its uses[15]. While in theory the GDPR regulation gives the user the right to challenge automated decision making, Malgieri[16] suggest that this 'right' will only exist if the user understands how the decision is reached, especially as people argue whether or not the right to an explanation also exists in the GDPR regulation[16]. While GDPR is supposed to give individuals greater control over their data and how it is used, Van Ooijen[25] suggests that while this has happened the scope of the regulation does not extend to cover all types of data, some of which is identifying data.

5 Conclusion

While all ethical issues experience a range of different challenges around them, these challenges may not necessarily be specific to one ethical issue but may instead apply to multiple issues. Two of the largest challenges around ethical issues are those of the accountability and transparency/explainability, as without the AI systems being accountable or transparent/explainable, other issues such as the individual's control of their data or hidden discrimination in outcomes of AI systems will come into play. As a lot of AI systems are seen as intellectual property, this increases the complexity of these challenges. Differing ethical viewpoints around the world also create problems, as while something may be seen as an ethical issue in the West it may not be seen as an ethical issue in other places. The developer has the responsibility to inform themselves of potential ethical issues using one of the many methods available, such as ethical frameworks, however it is complicated for developers to act on these issues as companies may prioritise their interests over the interests of the users.

As AI systems are often seen as intellectual property, it is complicated for the user of the AI system to be aware of and challenge the situations when the ethical issues are breached, as there is often a general lack of transparency/explainability of how the AI systems work and how the results are obtained. To enable the user to be able to be aware of these issues, governments have the responsibility to create regulation so as to provide the users with the information that they require to hold these AI systems to account. However, where regulation

has already been created it has been seen that differing interpretations of it may not lead to an increase in the accountably of AI systems.

References

1. Leslie D. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. Available at SSRN 3403301 2019
2. Berendt B. AI for the Common Good?! Pitfalls, challenges, and ethics pen-testing. Paladyn, Journal of Behavioral Robotics 2019; 10:44–65
3. Stahl BC. Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies. Springer Nature, 2021
4. Schrader DE and Ghosh D. Proactively protecting against the singularity: Ethical decision making in AI. IEEE Security & Privacy 2018; 16:56–63
5. Dignum V. Ethics in artificial intelligence: introduction to the special issue. 2018
6. Gold N, Colman AM, and Pulford BD. Cultural differences in responses to real-life and hypothetical trolley problems. Judgment and Decision making 2014; 9:65–76
7. Mac Síthigh D and Siems M. The Chinese social credit system: A model for other countries? The Modern Law Review 2019; 82:1034–71
8. Steele LG and Lynch SM. The pursuit of happiness in China: Individualism, collectivism, and subjective well-being during China's economic and social transformation. Social indicators research 2013; 114:441–51
9. Santiago N. Santiago N (2020) Shaping the ethical dimensions of smart information systems: a European perspective. SHERPA Delphi Study—Round 1 Results. SHERPA project 2020. Available from: https://www.project-sherpa.eu/wp-content/uploads/2020/03/sherpa-delphi-study-round-1-summary-17.03.2020.docx.pdf
10. Kodiyan AA. An overview of ethical issues in using AI systems in hiring with a case study of Amazon's AI based hiring tool. Researchgate Preprint 2019
11. O'neil C. Weapons of math destruction: How big data increases inequality and threatens democracy. Crown, 2016
12. Mittelstadt B. Principles alone cannot guarantee ethical AI. Nature Machine Intelligence 2019; 1:501–7
13. Vayena E, Salathé M, Madoff LC, and Brownstein JS. Ethical challenges of big data in public health. 2015
14. Custers B, Calders T, Schermer B, and Zarsky T. Discrimination and privacy in the information society. Studies in applied philosophy, epistemology and rational ethics 1866; 3
15. Parliament E and European Union the Council of the. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union 2016; 119:1–88
16. Malgieri G. Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations. Computer law & security review 2019; 35:105327

17. Shipman FM and Marshall CC. Ownership, privacy, and control in the Wake of Cambridge Analytica: The relationship between attitudes and awareness. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020 :1–12

18. Isaak J and Hanna MJ. User data privacy: Facebook, Cambridge Analytica, and privacy protection. Computer 2018; 51:56–9

19. Bessi A and Ferrara E. Social bots distort the 2016 US Presidential election online discussion. First monday 2016; 21

20. Liang F, Das V, Kostyuk N, and Hussain MM. Constructing a data-driven society: China's social credit system as a state surveillance infrastructure. Policy & Internet 2018; 10:415–53

21. European Comission. Regulation of the European Parliament and of the Council - Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending Certain Union Legislative Acts. 2021; COM(2021) 206 final

22. IEEE-CS/ACM Joint Task Force on Software Engineering Ethics and Professional Practices. Code of Ethics. Available from: https://www.computer.org/education/code-of-ethics [Accessed on: 2021 Dec 10]

23. Rolls Royce. The Aletheia Framework^{TM} v2.0. Available from: https://www.rolls-royce.com/sustainability/ethics-and-compliance/the-aletheia-framework.aspx [Accessed on: 2021 Dec 10]

24. Mujtaba DF and Mahapatra NR. Ethical considerations in ai-based recruitment. *2019 IEEE International Symposium on Technology and Society (ISTAS)*. IEEE. 2019 :1–7

25. Ooijen I van and Vrabec HU. Does the GDPR enhance consumers' control over personal data? An analysis from a behavioural perspective. Journal of consumer policy 2019; 42:91–107