Ethical Issues in Artificial Intelligence - Challenges, Awareness and Protections

In recent years Artificial Intelligence (AI) has seen rapid permeations into nearly all aspects of society, delivering both significant advantages and breaches of ethics. Such breaches include racial discrimination, stereotyping, risks to health and safety, intrusion of privacy, amongst others [1]. This paper is divided into three parts, covering:

1. The predominant challenges associated with tackling such ethical issues.
2. How AI developers can be more aware of and reason about ethical issues during development.
3. How users of AI can be aware of and how to deal with breaches of ethical issues.

Challenges of Ethical Issues

Inconsistent Frameworks

Efforts to confront ethical issues in AI have sparked the development of frameworks by organisations and governments alike. Whilst well-intentioned, the resulting plethora of guidelines for stakeholders to digest and subsequently strategise from can be both overwhelming and confusing [2]. For example, the Organisation for Economic Cooperation and Development presents over 600 initiatives for over 60 countries regarding AI frameworks [3], and also within the EU many proposed solutions are deemed to not always coincide, be uniform or even unanimous [4]. A key challenge then is how to apply ethical principles when solutions cannot be prescribed in "a mechanical manner, equally or similarly, in all fields on all subjects" [4].

To an extent, an ethical stance must be taken prior to designing a framework, however, this is a difficult task, as not only do moral judgments vary between societies but also within them, perhaps to the same extent as each other [5]. Furthermore, the question of which parties to involve in the design of guidelines is contested. Some argue for example, that governments should be completely excluded from framework discussions, given recent ethical breaches by state actors purportedly involved in high-profile cybersecurity attacks. It is also argued they would be incentivised to influence investigations and could use their powers to try and steer ethics consortiums away from cases that implicate that state [6].

The Power to Opt Out & Trust

Another challenge is that of individuals not having the power to decide whether or not to use or be affected by AI. From subjection to software-based scrutiny during a job application process, to being monitored using AI-based surveillance systems, external stakeholders often have very little power to opt out [7], even if they do not trust the software or people behind it. Where individuals do wish to opt-out, the challenge is then on controlling the extent to which this begins

to affect their quality of life in terms of the everyday services, health care and possibly even social circles that they might no longer have access to.

Furthermore, in the case of surveillance systems, Wang (2019) deconstructs a predictive policing application reportedly used by Chinese authorities to oppress ethnic groups and finds not only contradictions to Chinese law but also severe violations of internationally guaranteed human rights [8]. In observing cases like these where there isn't even the illusion of choice, a key challenge is therefore how to foster and retain trust between stakeholders involved with AI.

Explainability & Conflicts With Existing Rights

In recent years traditional methods of constructing AI models such as recurrent or convolutional neural networks have been superseded by transformers [9], which are significantly larger, more expensive to train, and in the context of resolving ethical challenges, more complex. The fledgling field of Explainable Artificial Intelligence has grown around the principles of producing more explainable models which advocate human understanding. However, one of the main challenges is that increasingly complex problems require more complex approximations of their related functions, and as a result, there is a consequential decline in the ability to provide both accessible and useful explanations to a broad range of stakeholders [10].

In the case of healthcare clinicians, this also presents a possible conflict with the General Data Protection Regulation because of the entitlement of 'data subjects' to meaningful information about the logic involved in AI systems [11]. This issue is exacerbated when informed developers cannot explain the 'black-box' nature of their products, as reported by Corti, a software company providing cardiac arrest detection software [12]. Therefore another key challenge presented by the rapidly increasing complexities of AI is the extent to which decision-makers (clinicians, insurance providers, etc) have to be sufficiently educated to meet regulations, especially when human lives are potentially at risk.

Economic Pressures

PwC predicts increases in Gross Domestic Product by 10.3% by 2030 as a result of AI, with 1.9% attributable to gains in productivity [13]. The difficulty this presents is how to address ethical issues of AI in economies whose competitive advantage is increasingly derived from its use.

To remain competitive, organisations are turning to AI solutions and, where resource or expertise is not available are likely to resort to off-the-shelf AI [14], and thereby risking incorporating any inherent biases, whether due to lack of care on the part of the developer or as determined by the cultural and ethical differences of the society within which the AI was developed [15].

Societal bias is prevalent across many popular pre-trained transformer models [16], which carry a significant risk of exacerbating racial, gender and other biases [17]. A warning on the HuggingFace website gives a demonstration of Google's "BERT" pre-trained model predicting a woman as working as a nurse or a prostitute, compared to a man working as a doctor or

mechanic [18]. The issue then is how to address bias in readily available AI, when most of the training data is scraped from the internet with little regard for ethical principles [16], and offered for free to developers under increasing pressure to deliver economic gains.

Developer Awareness & Reasoning

Existing Literature & Tools

A short term solution to promoting ethical awareness amongst developers of intelligent systems is to self-educate on the issues, challenges, and possible solutions. Efforts have been made to amalgamate the extensive research on the topic into more succinct educational material, such as Morley et al (2020) who provide a typology (to which developers can refer) of relevant tools and methodologies to address ethics at all stages of the development lifecycle [19].

Tools that facilitate ethical reasoning about a particular model are also available on cloud-based computing platforms such as Amazon Web Services, which offers products such as SageMaker Clarify to help developers detect and measure bias as well as improving transparency and explainability to stakeholders [20].

Incorporation Into Education

Developers at present are under no enforceable obligation to proactively pursue knowledge or reasoning about ethical issues in the development process (which may explain why there have been so many reported breaches of ethics [23]). While developers employed by large-scale tech companies are likely to have a clear Corporate Social Responsibility (CSR) charter from which developers can curate an ethical mindset [21], Turyakira (2017) finds there is a lack of general ethical consideration amongst smaller companies as a result of lack of funding and the perceived high costs of operating ethically [22].

To that end, Borenstein & Howard (2021) argue that awareness could be best achieved by the incorporation of AI ethics into various stages of education, such as masters courses or in school curriculums [24]. They also advocate the adoption of a professional mindset, such as by committing to an oath similar to that of the Hippocratic Oath, a sworn ethical commitment by physicians [25].

Modern Auditing

However, some argue that mere knowledge of ethics is insufficient, as it doesn't necessarily exert an influence on actual behaviour. Just as traditional auditing practices seek to enforce sound corporate behaviour and provide professional insight into the operations of a business, algorithmic auditing may achieve similar effects [26]. A more 'official' opinion of the ethicality of the development process could serve to allow developers to better reason about and address ethical issues.

AI Ethicists

Finally, an AI Ethicist is a role involving the determination of right from wrong in the development process of AI systems [27]. Such a position would take much of the onus of digesting the complexities of ethical challenges away from the developer, whilst still providing them with a timely feedback loop and open discussion forum such that they can effectively reason about and implement more ethical AI practices.

User Awareness & Dealing With Breaches

Corporate Social Responsibility

At an organisational level, a CSR policy can incorporate an official stance on ethical issues in AI and should highlight the importance of awareness to budget holders and other stakeholders who might plan to use AI-based products. The CSR is typically introduced to new hires during the induction and onboarding process and can be used in combination with a 'guided tour' of use cases of AI with example scenarios of where systems are biased or demonstrate clear breaches of ethical principles [28]. It can therefore serve as both an awareness-raising exercise as well as a sense check for the potential employee to help decide whether to accept a position or not. Many companies also include a whistleblowing system as part of the CSR policy that individuals can use to anonymously report perceived breaches of ethical principles.

Industry Advice

Externally, users can also research both industry-specific guidelines, such as the UK Office for Artificial Intelligence's guidelines for non-technical public sector workers [29], as well as refer to surveys about the most important ethical principles to guide their learning, such as in Rothenberger et al (2019). In this paper, the authors outline ethical AI focus points relative to each industry, to which the user can refer and obtain what is surveyed as being the most relevant ethical issue [30].

Human Rights

To be able to assess the extent to which an AI-based product or company is acting ethically, it could also benefit the user to be aware of what their rights are concerning their data and how others might be able to use it (public engagement reports have highlighted that many people don't know their rights [31]). Furthermore, if there is a breach of privacy or other concerns the individual can request information directly from the developers about the usage of their data under the Freedom of Information laws which have been adopted by over 60 nations [32]. In the UK, any non-compliance with requests can be escalated to the Information Commissioner's Office [33].

Legal Action

Where rights have been infringed upon, victims can take legal measures if the above actions have not rectified the problem. Crowdfunding organisations can be approached to help raise funds for legal cases, and have been used to some success such as in "The People vs The Snoopers' Charter", a case against the Investigatory Powers Act which allowed authorities to collect and store online activities, which was deemed unlawful by the British High Court [34,35].

Social Media

Finally, social media also has a strong role in raising AI-ethical awareness for the individual user. Although not specifically related to AI, a report by the Pew Research Centre found that 25% of social media users reported becoming more involved in political issues after reading related posts or being involved in online discussions [36]. Furthermore, social media sites can also be used to call attention to ethical breaches, perhaps to reach out to the developers [37], or to highlight direct breaches of ethics by the social media sites themselves [38].

Concluding Remarks

This short discussion has highlighted some of the challenges in tackling ethical issues associated with AI, as well as offering guidance to both developers and other stakeholders around becoming more proficient in awareness, reasoning, and how to resolve related breaches. It seems that there is no choice but to accept AI into our daily lives, but what is also clear is that there is a significant global movement to make the transition to an AI-based society fairer for everyone.

Given that the source of many problems stem from a lack of education or willingness on the part of the developer, one area that appears to be under-served is that of the professional accreditation in AI. In many professions where there is a significant degree of responsibility to act ethically there exist global professional bodies whose members must demonstrate not only in-depth knowledge but also that they can uphold certain ethical values. Even after membership is obtained, members are required to prove that they are following set standards via Continuing Professional Development (CPD) schemes [39]. In researching this discussion it became apparent that no such global body for AI professional standards exists, and I believe society could be better served by its introduction.

References

[1]Stahl, B., 2021. Ethical Issues of AI. SpringerBriefs in Research and Innovation Governance, pp.35-53.

[2] Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review, 1. https://doi.org/10.1162/99608f92.8cd550d1

[3] The Organisation for Economic Cooperation and Development, accessed 26/06/21, https://oecd.ai/

[4] Robles Carrillo, M., 2020. Artificial intelligence: From ethics to law. Telecommunications Policy, 44(6), p.101937.

[5] Graham, J., Meindl, P., Beall, E., Johnson, K. and Zhang, L., 2016. Cultural differences in moral judgment and behavior, across and within societies. Current Opinion in Psychology, 8, pp.125-130.

[6] Davis, John S. II, Boudreaux, B. , Welburn, J., Aguirre, J., Ogletree, C., McGovern, G., and Chase, M., 2017. Stateless Attribution: Toward International Accountability in Cyberspace, Santa Monica, Calif.: RAND Corporation https://www.rand.org/pubs/research_reports/RR2081.html

[7] Vakkuri, V., Kemell, K., Jantunen, M. and Abrahamsson, P., 2020. “This is Just a Prototype”: How Ethics Are Ignored in Software Startup-Like Environments.

[8] Human Rights Watch, 2018. China’s Algorithms of Repression. Reverse Engineering a Xinjiang Police Mass Surveillance App. [online] Available at: https://www.hrw.org/report/2019/05/01/chinas-algorithms-repression/reverse-engineering-xinjiang-police-mass [Accessed 26 June 2021]

[9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I., 2017. Attention is all you need. arXiv preprint arXiv:1706.03762.

[10] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F., 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, pp.82-115.

[11] Gerke, S., Minssen, T. and Cohen, G., 2020. Ethical and legal challenges of artificial intelligence-driven healthcare. Artificial Intelligence in Healthcare, pp.295-336.

[12] Vincent J. AI that detects cardiac arrests during emergency calls will be tested across Europe this summer. Verge, https://www.theverge.com/2018/4/25/17278994/ai-cardiac-arrest-corti-emergency-call-response; 2018 [Accessed 21 June 2021].

[13] PricewaterhouseCoopers LLP, 2017. The economic impact of artificial intelligence on the UK economy. [online] Available at: https://www.pwc.co.uk/economic-services/assets/ai-uk-report-v2.pdf [Accessed 26 June 2021].

[14] Anderson, J., 2020. Europe has an artificial-intelligence skills shortage. [Blog] Available at: https://www.bruegel.org/2020/08/europe-has-an-artificial-intelligence-skills-shortage/ [Accessed 26 June 2021].

[15] Brookings Institution, 2019. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. [online] Available at: https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/ [Accessed 26 June 2021].

[16] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., et al. 2019. Huggingface's transformers: State-of-the-art natural language processing. ArXiv, abs/1910.03771.

[17] Lovering, C., Jha, R., Linzen, T. and Pavlick, E., 2020. Predicting Inductive Biases of Pre-Trained Models. OpenReview, [online] Available at: https://openreview.net/forum?id=mNtmhaDkAr [Accessed 26 June 2021].

[18] Huggingface.co. 2021. Transformer models - Hugging Face Course. [online] Available at: https://huggingface.co/course/chapter1/8?fw=pt [Accessed 26 June 2021].

[19] Morley, J., Floridi, L., Kinsey, L. et al. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. Sci Eng Ethics 26, 2141–2168 (2020). https://doi.org/10.1007/s11948-019-00165-5

[20] Amazon, "Amazon SageMaker Clarify". https://aws.amazon.com/sagemaker/clarify/ [Accessed 22/06/21]

[21] Google LLC, "Artificial Intelligence at Google:Our Principles". https://ai.google/principles/ [Accessed 22/06/21].

[22] Turyakira, P.K., 2018, 'Ethical practices of small and medium-sized enterprises in developing countries: Literature analysis', South African Journal of Economic and Management Sciences 21(1), a1756. https://doi.org/10.4102/sajems.v21i1.1756

[23] Sinha, D., 2021. Top 5 Most Controversial Scandals in AI and Big Data. [online] Analyticsinsight.net. Available at: https://www.analyticsinsight.net/top-5-most-controversial-scandals-in-ai-and-big-data/ [Accessed 26 June 2021].

[24] Borenstein, J., Howard, A. Emerging challenges in AI and the need for AI ethics education. AI Ethics 1, 61–65 (2021). https://doi.org/10.1007/s43681-020-00002-7

[25] Nlm.nih.gov. 2021. Greek Medicine - The Hippocratic Oath. [online] Available at: https://www.nlm.nih.gov/hmd/greek/greek_oath.html [Accessed 26 June 2021].

[26] McNamara, A., Smith, J. and Murphy-Hill, E., 2018. Does ACM's code of ethics change ethical decision making in software development?. Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering,

[27] Gambelin, O., 2020. Brave: what it means to be an AI Ethicist. AI and Ethics, 1(1), pp.87-91.

[28] Bogina, V., Hartman, A., Kuflik, T. and Shulner-Tal, A., 2021. Educating Software and AI Stakeholders About Algorithmic Fairness, Accountability, Transparency and Ethics. International Journal of Artificial Intelligence in Education,.

[29] GOV.UK. 2021. A guide to using artificial intelligence in the public sector. [online] Available at: https://www.gov.uk/government/publications/a-guide-to-using-artificial-intelligence-in-the-public-sector [Accessed 26 June 2021].

[30] Rothenberger, L., Fabian, B., Arunov, E.. 2019. "RELEVANCE OF ETHICAL GUIDELINES FOR ARTIFICIAL INTELLIGENCE – A SURVEY AND EVALUATION". In Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden, June 8-14, 2019. ISBN 978-1-7336325-0-8 Research-in-Progress Papers.

[31] The Royal Society, 2017. Data governance: public engagement review. [online] Available at: https://royalsociety.org/~/media/policy/projects/data-governance/data-governance-public-engagement-review.pdf?la=en-GB [Accessed 26 June 2021].

[32] Ackerman, J., Sandoval-Ballesteros, I.. "THE GLOBAL EXPLOSION OF FREEDOM OF INFORMATION LAWS." Administrative Law Review, vol. 58, no. 1, 2006, pp. 85–130. JSTOR, www.jstor.org/stable/40712005. [Accessed 23 June 2021].

[33] Information Commissioners Office. https://ico.org.uk/ [accessed 23/06/21]

[34] Stoughton, C., 2018. The People vs the Snoopers' Charter. [online] crowdjustice.com. Available at: https://www.crowdjustice.com/case/snoopers-charter/?referer=stories [Accessed 26 June 2021].

[35] CHARTER, S., 2021. SNOOPERS' CHARTER - Liberty. [online] Liberty. Available at: https://www.libertyhumanrights.org.uk/fundamental/mass-surveillance-snoopers-charter/ [Accessed 26 June 2021].

[36] Lee, J., & Myers, T. (2016). Can Social Media Change Your Mind? SNS Use, Cross-cutting Exposure and Discussion, and Political View Change. Journal of Social Media Studies, 2, 87–97. https://doi.org/10.15340/2147336622992

[37] CommBox. 2021. The Role of Social Media in Customer Service, a Social Media Guide | CommBox. [online] Available at: https://www.commbox.io/the-role-of-social-media-in-customer-service-a-social-media-guide/ [Accessed 26 June 2021].

[38] MIT Technology Review. 2021. TikTok changed the shape of some people's faces without asking. [online] Available at: https://www.technologyreview.com/2021/06/10/1026074/tiktok-mandatory-beauty-filter-bug/ [Accessed 26 June 2021].

[39] Cimaglobal.com. 2021. CIMA - CPD. [online] Available at: https://www.cimaglobal.com/Members/CPD/ [Accessed 26 June 2021].