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?

Introduction

Artificial intelligence (AI) pervades in every aspect of our society, from the critical areas like urban infrastructure, law enforcement, banking, healthcare, and humanitarian aid, to autonomous vehicles. Due to the proliferation of AI in high-risk areas, it has become obligatory to design and govern AI to be accountable, responsible, and transparent. This essay addresses some of the challenges faced when addressing issues related to ethics of AI. It also investigates how to increase awareness among developers as well as end users of AI to enable ethical development, deployment, and governance of AI across radically different contexts of use.

The primary challenge in analyzing ethical issues of AI is to overcome the tendency to anthropomorphise it [1]. By anthropomorphising AI, we risk attaching human behaviours such as trust to AI systems. This in part diverts ethical responsibility from those developing the AI to the AI system itself. To be trusted a trustee must be held accountable for its actions. Based on this theory the burden of responsibility falls on those developing, deploying, and using AI. The main argument against assigning responsibility to the AI developers is it slows progress because they will become more cautionary about what they design and deploy. Users without being concerned about the progress of AI development must ensure that technologies are safe before utilizing it. Autonomous car's driving automation levels are an example of adequate steps taken by the policymakers to ensure AI maturity at each level.

Lack of legal and professional accountability to handle AI ethical issues next major challenge. Excluding privacy violations governed by data protection laws such as GDPR, AI development does not have a professional or legal endorsed accountability [10]. The safety of AI systems is now dependant on self-regulatory frameworks of AI development companies. Industry led guidelines and other self-governance bodies may often only serve to calm critical voices from the public, while the companies continue to maintain the criticized practices. Companies can highlight their presence in associations like "Partnership on AI" [6] whenever a serious commitment to a legal regulation on business activity needs to be stifled. Given this state of "self-regulation", accountability should not be relegated to regulations being applied after the AI systems being deployed, instead accountability must be taken into consideration at design time itself. Clear implementation and review processes should be the norm at a sectoral and organisational level. Owing to the high-risk areas AI gets applied, it may be necessary that an AI licensing mechanism is created that licenses developers working on AI for high-risk areas. The core of the licensing would be focus towards ethical commitments. Furthermore, developers should lobby for changes within their organizations to justify the need to invest in tools to explain their outputs to all stakeholders impacted by the system. Users should hold institutions responsible for the algorithms that they use, even if they are not able to explain how algorithms produced their results. In the event of an AI breach users should demand contrastive explanation such as why output A instead of output B. The more important the decision an AI system makes, the more explainable it should be for the users. A black box or non-explainable system should be treated as inscrutable and there labelled as inadequate technology.

Another troubling trend is use of AI as technical solutionism through the misuse of concepts like 'fairness' and 'discrimination [5]. Referring to complex social concepts to talk about simple statistics

is menacing because it causes confusion to researchers who become oblivious to the difference and policymakers who become misinformed about the ease of incorporating ethical requirements into AI. This runs the risk of narrowing complicated social concepts like fairness to a mere box-ticking exercise. More debate between industry, civil society and academia will be needed to remove conceptual ambiguity on the solutions applied to address ethical issues of AI. To contribute towards resolving this challenge, developers could address some common antipatterns in machine learning such as learning from past without remembering context, making spurious corelations, using inaccurate data, and using available data in place of the needed data. Developers should be capable of identifying flaws and question the use of attributes such as a race gender etc. while training a model. For example, Employees at companies such as Google, Facebook, Amazon, and Microsoft had protested the proposed use of facial recognition systems in criminal justice or military applications. As end users are directly affected by the decisions made by the AI algorithms, they should receive explanations about how the AI algorithm made a decision about them (for e.g., refusal of loan application). This is an idea being debated in Europe [7]. In fact, an important level of explainability and transparency should be to allow individuals to inspect and correct input data about themselves in the AI systems. End users should be able to present additional or alternate evidence for a human decision maker to weigh in conjunction with an AI system.

Using AI in autonomous vehicles give rise to issues of morality closely associated with ethical behavior. Autonomous vehicles are equipped with AI to make autonomous decisions. Since vehicles can cause considerable harm, the decisions made by them have moral and ethical implications. Programming these vehicles to handle ethical dilemmas is challenge. Ethical dilemma is normally a situation where there is no satisfying decision. Thus, it is impossible to decide among various possible decisions without overriding one moral principle [8]. Ethics is more of a thought process rather than a prescriptive process. A contrived example would be the trolley dilemma where a trolley that can no longer stop is hurling towards people working on a track. These people will die if hit by the trolley unless the trolley changes path on to another track where only one person is working. The challenge in these scenarios is that there is no truly right answer and answers would be based on each person's belief. It would be taxing to program the beliefs of different ethical schools in to an AI system. But the expectation from sophisticated autonomous vehicles will always be ''If humans can do it, why not smart machines? The primary ways ethics and social values are implemented in a society is through legal enforcement and personal choices [9]. For example, if a car does not stop at a stop sign, the driver is penalized through punishments or fines. But other ethical values such as stopping to help a stranded motorist is left to individual choices. Confirming to these values is fostered though individual social controls. As there is no single ethical framework that is sufficient to compute an ethical decision, developers of AI may need to formalize philosophical definitions that are available in natural language into generic concepts that can be programmed in to machines. For example, ethical frameworks such as Deontological ethics (nature of a decision), Utilitarian ethics(consequentialism) and Doctrine of Double Effect (consequentialism, deontology, causality and proportionality) could be used [8]. Use of these ethics frameworks for decision making will help improve explainability for a decision made. More effort may be needed to refine the existing frameworks to suit socially acceptable ethical judgements. Users should be aware of possibility of outlier events of an autonomous vehicle which are rare occurrences whose frequency is bounded due to some incidents that cannot be programmed and foreseen.

AI transparency is key to solving AI ethics issues. Even if AI developers develop systems with well-motivated inspirations, there always will remain the challenge to communicate the understanding of the underlying real-time decision-making functions. AI systems uses philosophical, mathematical, and biologically inspired techniques for building complex artificial systems. Explaining a complex design and the decisions made by it to end users is a challenge. This is the only way the decision-making black box can be opened for analyzing ethical behavior. But AI developers also need to keep their IP (Intellectual Property) secret to retain their competitive edge as well as prevent malicious use by unscrupulous actors from misusing flaws of their system. So, they will always resist any kind source code sharing or review process. These factors contribute to limiting transparency in AI systems. Developers must embrace transparency as a method to facilitate accountability. As a developer

several methods are available to provide accountability while keeping some information hidden, such as software verification to prove mathematically that a piece of software has certain properties, zero knowledge proofs which are cryptographic tools that enable a decision maker to prove that decision policy used had a certain property without revealing the decision policy itself [3]. These methods guarantee procedural regularity. To provide explainabilty interpretable by humans, concepts such as Shapley values can be used. Given a dataset consisting of a set features, Shapley values help to attribute a prediction made by a model to those features. Companies should incentivize engineers to bring incremental improvement of systems rather than continue to work with black boxes. End users can push for certifications of AI systems. Policy makers should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions.

An essential weakness of ethical guidelines for AI is that they are plenty of them and they lack mechanisms to reinforce their own normative claims. As of the beginning of 2020 there are 22 major AI ethics guidelines and enforcement are rather weak and pose no eminent threat [11]. Mechanisms to enforce ethical principles will lead to reputational losses in the case of misconduct or restrictions on memberships in certain professional bodies. Due to this, mechanisms are rather weakened and binding to any legal framework is continuously discouraged. Companies highlight membership in associations such as "Partnership on AI" whenever a need to commit to legal regulation of business activities needs to be stifled [11]. To fulfil commitment to safer AI, AI developers should commit to institutional changes such as adoption of legal framework conditions, establish mechanisms of independent audit of technologies, and institutions which addresses complaints and compensates for harms caused by AI systems. Developers would be able to reason more about issues around ethics if they have the needed education in the field of AI ethics while learning the AI technologies. Universities have a significant role to ensure the right approach based on technological advances, media and information. Ethicists must partly be capable of grasping technical details of the AI development process so that they are able to reflect on the ways data are generated, recorded, curated, processed, disseminated, shared, and on the ways of designing algorithms and code, respectively. To make ethical guidelines more effective guidelines must be more detailed. Splitting AI ethics into machine ethics, computer ethics, information ethics and data ethics would enable to narrow the deep the gap between concrete contexts of research, development, and application on the one side, and ethical thinking on the other.

Conclusion

A few challenges in achieving ethical and safe AI has been discussed and possible approaches from developers and user's standpoint was discussed. There are far more challenges associated with dealing ethical issues of AI which will need in depth discussions. Also, there were limited references on the role of users in identifying breaches of AI ethics. The fast pace development of AI could be one reason for this, nevertheless this area has scope for further research. The need for guidelines to be broken down for better implementation by developers and a better understanding by AI users and policy makers was also discussed. Expectation is that clear guidelines will augment development of ethical AI further. Challenges around the transparency of AI in addressing AI ethical challenges was discussed. Transparent AI is key to AI accountability and necessary to making AI decision making free from ethical issues. The concept of combining trust with autonomous systems and ethical issues arising from them was discussed. Autonomous systems are expected to be incorporated with decision processes based on advanced ethical frameworks and extensive testing mechanisms in future to deal with user's ethical concerns. The challenge of using AI to solve complex social concepts such as fairness and discrimination was discussed. This leads to a misguided assumption that complex social issues can be solved through some algorithm and debates on the issue itself will be oversimplified. Expectation is that when we begin to translate and implement our lofty social principles, we will discover the actual ethical challenges of AI. The lack of legal and professional accountability for AI was discussed. And finally, it paramount that users of AI systems be made aware that decisions made by AI systems are based on human made algorithms and that AI systems do not have the ability to

make ethical choices autonomously. Exposing flaws in AI systems is a possible research area for the future which would help in this area. As AI systems continues to become widespread and users continue to become increasingly conscious about ethical issues of AI, it is expected that the need for legal and professional accountability will be demanded by the users.

References

[1] Ryan, Mark. "In AI We Trust: Ethics, Artificial Intelligence, and Reliability." Science and Engineering Ethics 26.5 (2020): 2749-767. Web.

[2] Matthews, Jeanna. "Patterns and Antipatterns, Principles, and Pitfalls: Accountability and Transparency in Artificial Intelligence." The AI Magazine 41.1 (2020): 82. Web.

[3] Lepri, Bruno, Lepri, Bruno, Oliver, Nuria, Oliver, Nuria, Letouzé, Emmanuel, Letouzé, Emmanuel, Pentland, Alex, Pentland, Alex, Vinck, Patrick, and Vinck, Patrick. "Fair, Transparent, and Accountable Algorithmic Decision-making Processes." Philosophy & Technology 31.4 (2018): 611-27. Web.

[4] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-of-access (Accessed: 02 Dec 2020)

[5] Cath, Corinne. "Governing Artificial Intelligence: Ethical, Legal and Technical Opportunities and Challenges." Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical, and Engineering Sciences 376.2133 (2018): 20180080. Web.

[6] https://www.partnershiponai.org (Accessed: 02 Dec 2020)

[7] Steels, Luc, and De Mantaras, Ramon Lopez. "The Barcelona Declaration for the Proper Development and Usage of Artificial Intelligence in Europe." Ai Communications 31.6 (2018): 485-94. Web.

[8] Bonnemains, Vincent, Saurel, Claire, and Tessier, Catherine. "Embedded Ethics: Some Technical and Ethical Challenges." Ethics and Information Technology 20.1 (2018): 41-58. Web.

[9] Etzioni, Amitai, and Etzioni, Oren. "Incorporating Ethics into Artificial Intelligence." The Journal of Ethics 21.4 (2017): 403-18. Web.

[10] Mittelstadt, Brent. "Principles Alone Cannot Guarantee Ethical AI." (2019). Web.

[11] Hagendorff, Thilo. "The Ethics of AI Ethics: An Evaluation of Guidelines." Minds and Machines (Dordrecht) 30.1 (2020): 99-120. Web.