Ethics of Artificial Intelligence: Key issues and potential solutions

Group Members:

Muthupriya Shankaran
Shirdi Manjunath Adigarla
Sundara Lakshmi Paramasivam

Abstract:

Research on the ethics of Artificial Intelligence (AI) has grown substantially over the past decade. Due to increase in research interest in AI and its applications (in healthcare, industries and small cities) in everyday life, it has become necessary to identify its ethical impact on moral and societal issues. In the future it is expected that AI would play a vital role in decision making in both small and large (global) level. The impact of decision making would cover all, such as education, health monitoring, marketing and transportation. Under these circumstance it is necessary to evaluate the performance, reliability and accuracy and make changes or corrections if required. In this paper we aim to provide a review of recent developments in the ethical issues of AI. In addition, we highlight the key problems in ethical AI, and shown the importance of high quality data in terms of AI algorithms, also provide some essential solutions to the ethical issues. The ethics of AI can also be fostered between designers and users through some reviews and algorithms.

Keywords: Artificial Intelligence, Ethics, Policymaking, societal issues, Responsibility.

Table of Contents

1. Introduction:

Artificial Intelligence (AI) is a branch of computer science that aims to create intelligent machines. It has become an essential and critical technology that has applications in countless domains like biotech, microeconomics and Internet of things (IoT).However, at the beginning the immense amount of interest was comparatively less in the AI research. Later on, to develop the AI in a societal environment many attempts are carried out to understand better and define the concept of intelligence, which is cognitive capacity of human beings. Also AI traditionally called as an automation of intelligent behaviour. By contrast, AI has fail to come up with an exact definition of intelligence such as in psychology, biology, cognitive science and neuroscience. In other words we can say that it makes the machine able to work and react like humans. From the past fifty years to present we can see the mechanical system produced in Natural Language processing(NLP) (e.g., speech recognition, language translator, face recognition, handwriting recognition and many more) from the AI system. In upcoming years global progress is unpredictable because of versatile in AI technique, in this case AI requires more speed and memory than usual software's and hardware. Moreover, at first, machine learning technique- it involves training program into an expert system which enable computer to learn and perform tasks. There are several machine learning techniques used to train the algorithm such as supervised learning, unsupervised learning and reinforcement learning. On the other hand, Deep learning is more advanced branch of machine learning such as Artificial Neural Network (ANN), Convolution Neural Network (CNN) and Multi Neural Network (MNN)with these inherent learning features of deep learning ,it automatically learn, extract the features from datasets, for example, image, video or text without any traditional code or rules. Despite of these factors, we can see that in the future AI would act as an important role in decision making of small and large scale level in Industries. Hence decision making impact would cover some real life walks such as in education, health monitoring, transportation and many more. According to these, it is necessary to evaluate the performance, reliability, and accuracy of these algorithms and so make changes if corrections requires.

Thus in this paper we aim to provide a review and some key studies in AI ethics, more specifically we highlight how AI has developed gradually over recent years and identify ethical issues in AI in the past years and provide some potential solutions to the ethical issue, also provide some importance of high quality of data for training and testing algorithms which involved some necessary steps to be followed.

2. Literature Review:

2.1 Recently developed in AI Ethics: In this section, we discuss some of the recent developments in the field of Artificial Intelligence. In fact the author [1] said about the motivation of using AI comes in many use cases in modern world of AI. It brings several benefits to make the lives easy such as time consuming, many apps (online shopping, food ordering...) in smartphone and more importantly we can recognize the plentiful numbers of medical applications (example: AI helps the patient to think about the symptoms of various drugs and play major role in medicine field).

2.2 Discussion on some key points related to recent developments in AI Ethics:

In [3], the author have focused on the governance mechanism and regulatory issues. They claim that the lives need to be governed in a transparent manner and accountability must be shared with all stakeholders, also explain how algorithms can maintain and hinder the progress of social contracts.

In other studies [4], the authors provide an overview of tools and methods. They point out over the issue of trust between the machine learning developer community and the end-users, also they argued saying these issues need time to resolve and more efforts are required to transparent AI. Meanwhile the author [5], propose that the research efforts need to explore the tension between practice and principles, by recognizing this, decision making can be improved for specific use cases and guidelines for AI ethics.

In [6], [7], the authors investigate the ethical issues of AI from the perspective of safety applications. They implemented to use multi objective decision-making framework to make AI as more reliable and secure system.

3. Ethics & Philosophy of AI:

In general, as human needs, ethics, laws, norms govern human behaviour and develop reasoning and decision making. In the same way, machines also need some fundamental ethics to work effectively and to avoid damage. In upcoming years, on rising of AI, it is important to develop proper ethics using machine learning algorithms.AI has some potential to make ethical and philosophical decisions such as driverless cars and reliability as a services are pre decided to grow in later 50 years, it may reduce the chance of accident as well. When majority of people chose to save the lives of their own. The main reason for an AI code is to give efficient partner direction when comes with an ethical choice in regards and utilization of AI. There are few ethical issues and philosophical and domains in AI are listed below

i. Unemployment: This is major issues which occurs from physical work to computing system. Loss of jobs to human. For instance, driverless cars runs automatically without drivers, so here driver have to find another job as they are no more wanted.
ii. Commitment to cause: Each person of citizen who works on AI wants to create a safe place where the maximum protection can be provided to save data. To improve safety and security, one of the participants [8] mentioned that approaching the security breaches

for accessing the data is expensive and policymakers want to design a system which protects costs and liabilities.

iii. Biases (Behavioural, population and linking biases): In data bias having many parts idea including desirable attributes, e.g., completeness, rightness and timeliness.

3.1 Developing High-Quality labelled data:

From the above scenario, we know that data collection and how to produce a high quality labelled data.AI instruments can create progressively exact results, in the selection of an appropriate framework which is important to put the social media system into a framework to analyse different aspects, also with important of security of users account. Moreover, the framework is still in the design stage, therefore our aim should be highly to improve a social media system to analyse data more securely and analyse with awareness.

i. Framework design: The architecture of this design includes an inference controller, data manager and risk manager. On using modular approach, architecture checks the users query and puts that into user defining policy to design based on users requirement.

ii. Inference controller: It handles all the access control procedure data sharing. Meanwhile in metadata controller if someone ask for any information then the data controller re-examine and then gives the answer to query.

iii. Query Manager: He is responsible to accept user's requests. To avoid any claim in future query manager allow user to correct the error to be fixed by the data manager.

4. Solution to Ethical issues of AI

In some of commonality analysis, there is a need to provide a reliable solutions and policies for ethical issues of AI. The AI automated system playing a vital role in everyday lives, therefore by increasing our reliability of algorithms we can keep our equality, regulation and legislation as more important. Hence in this section, we provide some primary solution for the ethical issues of AI.

i. Societal and environmental well-being: AI systems should benefit all human beings, including future generations. It must hence be ensured that they are sustainable and environmentally friendly. Moreover, they should take into account the environment, including other living beings, and their social and societal impact should be carefully considered.

ii. Technical Robustness and safety: AI systems need to be resilient and secure. They need to be safe, ensuring a fall back plan in case something goes wrong, as well as being accurate, reliable and reproducible. That is the only way to ensure that also unintentional harm can be minimized and prevented.

iii. Accountability: Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes. Auditability, which enables the assessment of algorithms, data and design processes plays a key role therein, especially in critical applications. Moreover, adequate an accessible redress should be ensured.

iv. Diversity, non-discrimination and fairness: Unfair bias must be avoided, as it could have multiple negative implications, from the marginalization of vulnerable groups, to the exacerbation of prejudice and discrimination. Fostering diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle.

5. Conclusion:

Thus we conclude that the implementation of ethics in AI has become one of the liveliest topics in philosophy of technology. Though the ethical issues of AI can slow the process of development of AI solutions on a commercial level, there are several guidelines and possible solutions to the ethical usage of AI algorithms. In this paper we presented some key insight of AI ethics and highlighted the relevant issues. More importantly we have point out showing that there is a lack of study focusing on ethics for designing AI algorithms. In this regard, some fundamental steps for improving the quality of the data have been provided. Then, some key solutions to the ethical issues of AI have been presented along with some future research directions. The insight provided in this paper would be helpful for future work.

References:

[1] F. Jameel, U. Javaid, B. Sikdar, I. Khan, G. Mastorakis, and C. X. Mavromoustakis, “Optimizing Blockchain Networks with Artificial Intelligence: Towards Efficient and Reliable IoT Applications,” in Convergence of Artificial Intelligence and the Internet of Things. Springer, 2020, pp. 299–321.

[2] B. Mittelstadt, “Ai ethics–too principled to fail?” Available at SSRN 3391293, 2019.

[3] I. Rahwan, “Society-in-the-loop: programming the algorithmic social contract,” Ethics and Information Technology, vol. 20, no. 1, pp. 5–14, 2018

[4] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, “From what to how. An overview of AI ethics tools, methods and research to translate principles into practices,” arXiv preprint arXiv: 1905.06876, 2019.

[5] J. Whittlestone, R. Nyrup, A. Alexandrova, and S. Cave, “The role and limits of principles in AI ethics: towards a focus on tensions,” in Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. ACM, 2019, pp. 195–200.

[6] P. Vamplew, R. Dazeley, C. Foale, S. Firmin, and J. Mummery, “Human-aligned artificial intelligence is a multi-objective problem,” Ethics and Information Technology, vol. 20, no. 1, pp. 27–40, 2018.

[7] A. Critch, “Toward negotiable reinforcement learning: shifting priorities in Pareto optimal sequential decision-making,” arXiv preprint arXiv: 1701.01302, 2017.

[8] T. Jameel, R. Ali, and S. Ali, “Security in Modern Smart Cities: An Information Technology Perspective,” in 2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE), March 2019, pp. 293–298.