Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/

(REVIEW ARTICLE)

Check for updates

Ethics in artificial intelligence: Issues and guidelines for developing acceptable AI systems

Sheriff Alim * and Adewale O Adebayo

Department of Computer Science, Babcock University, Ilisan-Remo, Ogun State, Nigeria.

Global Journal of Engineering and Technology Advances, 2022, 11(03), 037–044

Publication history: Received on 09 May 2022; revised on 12 June 2022; accepted on 14 June 2022

Article DOI: https://doi.org/10.30574/gjeta.2022.11.3.0092

Abstract

There is no doubt, that Artificial Intelligent (AI) has a lot of contributions to the overall well-being of humanity and improving productivity, it is worth noting that there is also a negative impact associated with the use of the technology on ethical grounds ranging from privacy issue to lack of transparency. Quite a lot of work has been done to provide best practices in the development and use of these systems to gain the acceptance of the public; this is very essential for nations to exploit the full potential of AI. This work brings to the frontlines some of the risks/issues associated with AI from different pieces of literature and provides commonplace or leading principles and frameworks required to develop responsible AI or autonomous systems. The guidelines encourage stakeholders in the industry of developing AI systems to incorporate facilities that support transparency, privacy protection, accountability, and contestability. These guidelines are not new, but they are contextualized to suit ethics in AI. Above all, AI systems should not be seen as being equal to or greater than man, but rather be seen as augmenting intelligent systems to assist human who is the only moral agents.

Keywords: Artificial Intelligence; Ethics; Principles; Guidelines; Issues

1. Introduction

Artificial intelligence (AI) is primarily about building computing systems capable of solving problems intelligently without being explicitly programmed. It can be considered as the integration of various fields such as computer science, electrical engineering, psychology, mathematics, statistics, biology, and linguistics. Its functioning is premised on optimization and classification or prediction problems [1].

Some opportunities come with the use of AI and AI can improve well-being and overall productivity. Despite all the potentials of AI systems, the concerns, questions raised, and foreseen risks should not be downplayed as they are valid and need to be addressed adequately, and if not, the full potential of AI might not be achieved. There is a need to strike a balance between misuse/excess and appropriate use of AI. There is a strong justification for governance framework(s) to guide the application and degree of use of AI. The framework should inculcate ethical principles into various stages of the lifecycle of the system such as in the analysis, design, implementation, and deployment stages.

Many governments around the world are fully aware of the potential that AI can provide to their societies and economies and are mindful of associated risks which can hinder these benefits because of trust issues from the public. They have come up with several initiatives focusing on ethics in AI. In Germany, the government led the development of an ethical guideline for autonomous vehicles. In New York, Automated Decision Task Force was formed to develop transparency in the use of AI. The minister for communications in Singapore appointed an Advisory Council on the ethical use of AI

* Corresponding author: Sheriff Alim
Department of Computer Science, Babcock University, Ilisan-Remo, Ogun State, Nigeria.

and Data. The Australia Artificial Intelligent Ethics framework is another government drive at maximizing the full potential of AI by addressing fundamental ethical issues [2].

It is also worth noting that professional bodies are not left behind in terms of contributions to AI ethics, IEEE formulated Ethics for Autonomous vehicle.

Inculcating ethics into AI provides dual advantages to society by being enabled to identify and leverage new opportunities, and at the same time, it enables society as well anticipate and avoid or reduce costly mistakes [3].

Various disciplines have contributed to the development of ethical concepts whose wholeness can be seen as a human-rights structure used all over the world. AI ethics is not about re-inventing new laws and ethical standards, but rather, updating the existing ones and making them suitable for the context of AI [2].

The application of AI has the incredible potential of improving society and the overall wealth of nations. The use of AI currently is being faced with strong criticism from the public based on ethical risks or challenges. If the risks are not addressed, the benefits to be derived from the systems will continue to be a mirage. This work focuses on bringing to the forefront these risks or ethical issues surrounding the use of AI and presents principles, guidelines, best practices, and frameworks that can help in dealing with risks and building people's trust in the systems if they are fully embraced by designers and stakeholders in the AI development industry.

Two research objectives for this work are

- To Consolidate ethical issues in AI from various literature with context and
- Present existing guidelines, principles and frameworks needed to develop responsible AI systems.

This work consolidates various ethical issues connected with the design, development and deployment of AI systems and brings into the limelight many guidelines, principles and frameworks that can be adopted in the developing and use of ethical AI applications. For anyone interested in acquitting himself/herself with ethics in AI or researchers who want to embark on AI, this work will serve as a one-stop shop that provides a comprehensive list of ethical concerns in AI and provides relevant guidance required to produce an acceptable result in this field.

2. Outcome

2.1. Approaches for addressing ethical Issues in AI

A lot has come from various groups such as individuals, professional bodies and governments of different countries in form of principles, professional ethic codes and frameworks. In this section, a review of such efforts will be considered.

2.2. Governmental contribution from various countries of the world to addressing ethics in AI

Many governments of the different countries across the globe are fully aware of the benefits and opportunities that come with AI technologies such as the capability to increase the well-being of nations and efficient utilization of resources, have focused their effort in aligning the design, implementation, and deployment of AI with standard ethical practices and inclusive values. These efforts are essential for the citizen to gain trust in the application of AI technologies and products [2].

Automated Decisions Task Force saddled with the responsibility of formulating policy/framework on transparency and equity in the use of AI was announced by the mayor of New York in 2018 [2]. The Task Force came up with their report in November 2019, added context to Automated Decision System (ADS) as tools used by agencies of New York City for making decisions around services and resources of the city; and tools use data and algorithms and AI to aid decision making. The report contains three broad recommendations which are

- Capacity building for an equitable, effective, and responsible approach to the City's ADS,
- Increasing public discussion on ADS and
- Formalizing management of ADS functions [4].

In Germany, government-led ethical guidelines were released in 2017 by the federal ministry of transportation and digital infrastructure for the development of the autonomous vehicle. The release comprises twenty ethical rules for automated and connected vehicular traffic and the first three are:"

- The primary purpose of partly and fully automated transport systems is to improve safety for all road users. Another purpose is to increase mobility opportunities and to make further benefits possible. Technological development obeys the principle of personal autonomy, which means that individuals enjoy the freedom of action for which they are responsible.
- The safety of individuals takes priority over all other utilitarian considerations. The objective is to lower the level of harm until it is completely avoided. The licensing of automated systems is not acceptable unless it promises to produce at least a diminution in harm compared with human driving, in other words, a positive balance of risks.
- The public sector is responsible for guaranteeing the safety of the automated and connected systems introduced and licensed in the public street environment. Driving systems thus need official licensing and monitoring. The guiding principle is the avoidance of accidents, although technologically unavoidable residual risks do not militate against the introduction of automated driving if the balance of risks is fundamentally positive"[5].

In 2018, Australia funded the development of a national ethic framework called "Australia's Ethics Framework" which contains two fundamental concepts which are the Core Principles for AI and a toolkit for ethical AI. The core principles of AI consist of

- Generating net benefits
- Do not harm
- Regulatory Compliance
- Privacy protection
- Fairness
- Transparency and Explainability
- Contestability and
- Accountability [2].

The toolkit covers

- Impact Assessment
- Internal or External Review
- Risk Assessments
- Best Practice Guidelines
- Industry Standards
- Collaboration
- Mechanism for monitoring and improvement
- Recourse mechanism and
- Consultation [2].

The above toolkit for ethical AI serves as a platform for realizing the core AI principles of the framework in AI systems when properly applied.

The United Kingdom government established, in 2018, the Center for Data Ethics and Innovation to establish measures that will assure ethical and safe innovation around Artificial Intelligent and data; the center was saddled with the responsibility of working with regulators on dealing with ethical challenges that emanated from the use of Artificial Intelligent and data [6]. The European Union appointed a group of experts drawn from civil society, business and academia saddled with the responsibility of making recommendations that will address the opportunities and challenges associated with the use of AI which will cumulate into AI policy recommendations covering ethics, social and legal [7].

Dawson et. al,[2] in their work apart from the toolkit and core principle also cover data governance which explicitly explores

- Consent and privacy issues
- Data breaches
- Bias data and
- Open data source and re-identification.

In June 2018, the Singapore government established an ethics and standard council saddled with the responsibility of advising the government on AI development and its usage. In UAE, the council for AI and the minister of AI worked on the mechanism for the evaluation of AI applications. Other countries of the world such as China, Canada, India and France have also made a tremendous effort on Ethics for AI-related systems [8].

The General Data Protection Regulation (GDPR) by the UK government is also a useful regulation in AI, especially the six principles relating to processing personal data. The six principles of personal data are:

- Fairness, lawfulness, and transparency in processing data,
- Collected data should be used only for the stated, clear, and legitimate purpose,
- Data minimization

Collect adequate and relevant data, restricted to only what is essential to the purposes for which they are processed,

- Accuracy

Personal data must be accurate and, where necessary, must be kept up to date (accuracy),

- Storage limitation

Personal data should not be stored longer than necessary for the purpose for which the data was processed

- Integrity and confidentiality

Data processing should be done in such a way that guarantees the appropriate security of personal data, including protection against unauthorized or unlawful processing [9].

2.3. Organizational and Institutional Ethics Framework

A couple of organizations and institutions have also documented ethical guidelines in AI among several of them are the Institute of Electrical and Electronics Engineers, The Public Voice Coalition, Google, Microsoft, and IBM just to mention a few.

The IEEE [10] released a report regarding the design and implementation of AI that highlighted core principles which are:

- human rights
- Accountability
- Well-being,
  Transparency
- Data agency
- Effectiveness
- Awareness of misuse
- Competency.

The Universal Guidelines for Artificial Intelligence issued by "The Public Voice" [11] emphasized that the primary responsibility of AI lies with organizations that fund, develop, and deploy AI systems. The guidelines cover the following:

- Right to Transparency
- Right to human determination
- Identification of obligation
- Fairness Obligation
- Assessment and Accountability
- Accuracy, Reliability and Validity Obligations
- Data Quality Obligation
- Public Safety Obligation
- Cybersecurity Obligation
- Secret Profiling Prohibition

- Unitary Scoring Prohibition and
- Obligation Termination.

3. Discussion

With reviews of various contributions, one can see significant overlap in principles or guidelines and an attempt is made below to generate a consolidated or master list of relevant principles with a short description of what each stands for.

3.1. Generate profit

Benefits derived from the use of AI systems must be more than it total cost.

3.2. Do no harm

AI systems must not be designed to harm or deceive humans or result in negative outcomes.

3.3. Regulatory and legal compliance

The AI system must conform strictly to the existing regulations and legal framework of all relevant countries i.e., the country where the design and implementation are done and the country of use.

3.4. Privacy Protection

With AI, the personal information of individuals will be collected, and a process must be put in place to ensure such information is protected from unauthorized access or unauthorized use or misuse.

3.5. Fairness

The system must be devoid of any form of bias, and must not be seen as discriminating against any group e.g., sex, ethnicity etc.

3.6. Transparency and Explainability

People must be aware when AI is being used and the rationale used by the system should be explainable in the language people or stakeholders can relate with.

3.7. Contestability

Embedded in the design of AI applications, there should be a provision for people to challenge the decision made by the underlying algorithm. To prevent human prejudice and biases from being reflected in the system, design fairness, data fairness, and outcome fairness must be guided jealously during the design and implementation stages.

3.8. Accountability

In the design of any AI system, consideration must be given to the end-to-end facility needed to support answerability and audibility. As part of the identification obligation, people who are responsible for the design and implementation of the AI systems must be identified in the scheme and held accountable for the impact of system decisions and actions.

3.9. Cybersecurity Obligation

The AI systems must be secured against cybersecurity threats.

3.10. Accuracy, Reliability and Validity Obligation

The designers and manufacturers of AI systems must ensure the systems produce accurate and reliable outcomes.

All the above principles must be incorporated in all stages of the AI system life cycle, and this refers to the analysis, design, implementation, deployment, and monitoring stages. To achieve this, consideration should be given to the concept of ethics by design, ethics in Design and ethics for Design [12].

Ethics by Design implies that the reasoning capabilities of AI should be infused with the identified principles which will serve as a guide in the decision-making of the autonomous systems. Ethics by Design is concerned with the analysis and

evaluation of ethical consequences of AI systems supported by engineering and regulatory methods. And lastly, Ethics for Design is primarily about codes of conduct, standards, and certification process.

4. Review of related work

Leslie [13] in his work identified seven potential harms caused by AI systems and they are:

- Bias and discrimination
- Denial of individual autonomy, recourse and rights
- Non-transparency, unexplainable or unjustifiable outcomes
- Invasion of privacy
- Isolation and disintegration of social connection and
- Unreliable, unsafe or poor quality outcomes.

Leslie [13] went ahead to recommend a governance architecture for the ethical design and deployment of an AI system which when adopted as an ethical platform can help deal with the identified potential harms that AI systems can cause. The governance architecture consists of values, principles and processes that must guide every life cycle stage of AI. The values are: Respect, Connect, Care and Protect while the principles are fairness, accountability, sustainability and transparency.

Some of the challenges associated with the application of AI as identified by Dawson et al., (2) are the

- Privacy issue
- Data breaches
- Bias in data
- Black box and transparency issues
- Automation bias issues and
- Bias, prediction, and discrimination issues.

Embracing a set of principles ethical toolkit in AI practice can help in the delivery of ethical AI systems. This principle comprises

- Fairness
- Do no harm
- Regulatory and legal compliance
- Privacy protection
- Generate net benefits
- transparency and explainability
- contestability and
- Accountability.

The tool kit for ethical AI comprises impact assessment, internal or external review, risk assessment, best practice guidelines, and consultation just to mention a few [2].

Cowls et al., [3] in their work pointed out four opportunities offered by the use of AI and four risks associated with AI. The opportunities are enabling human self-realization, enabling human agency, increasing societal capabilities and cultivating social cohesion. The possible risks are devaluing human skills, eliminating human responsibility, decreasing human control and eroding human self-determination.

5. Conclusion

Ethics in AI is not about developing new principles as ethics is mature with contributions from various fields of studies, it rather focuses on aligning the existing work to fit the new needs that arise due to the new technology; AI ethics is all about contextualizing existing ethics principles/framework for AI.

The core principles or guidelines to be used in the design, implementation, deployment and evaluation of AI systems or autonomous decision-making systems are

- The AI application generates profit or value
- The system should not do any harm (do no harm)
- The system should be compliant with the existing regulatory and legal framework
- Individual privacy should be protected
- The system should demonstrate fairness and transparency in decision making which must be explainable in a language stakeholders can understand
- The system's decisions and actions should be contestability
- There should clear accountability structure, sustainability
- Cybersecurity obligation of the system and data
- Accuracy, Reliability and Validity Obligating (the system must not produce a poor quality outcome).

Recommendation

AI should be treated as augmented intelligent systems and should never be positioned in competition with humans. Despite the obvious fact that cars are faster than the fastest human, they have never been seen to be equal to or better than man, rather, they are seen as tools for improving human wellbeing. This philosophy should serve as guidance when assessing the social impacts of the application of such AI/autonomous systems.

The current and future designers of AI systems should be thoroughly taught all relevant areas of ethics, especially students of computer science, electrical engineering, computer engineering as well as statistics and mathematics and in all higher institutions so that they will be properly equipped with the necessary tools in prioritizing ethical considerations in designs, implementation, deployment, and monitoring of autonomous/AI systems.

Many countries of the world are fully aware of the positive contributions AI is already delivering and will continue to deliver to societies, increasing overall productivity and well-being; like any other technology, it can be subjected to poor design, implementation and misuse which will result in negative social consequence whose impact and degree can at best be imagined; therefore, the government of such countries are taking bold steps, driving effort of formulation of best practices, guidelines and framework targeted as forestalling the foreseen risks. Nigeria being the most populous nation in Africa should take a cue from other nations to formulate an AI ethical framework that is best for the country, a form of governance for the use of AI or autonomous related solutions in the country.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflicts of interest.

References

[1] Jonathans S. Artificial Intelligence and Ethics:Ethics and the dawn of decision-making machines.Harvard Magazine 2019.

[2] Dawson D, Schleiger E, Horton J, McLaughlin J, Robinson C, Quezada G, et al. Artificial Intelligence: Australia's Ethics Framework (A Discussion Paper). Artif Intell Aust Ethics Framew. 2019;

[3] Cowls J, Floridi L. Prolegomena to a White Paper on an Ethical Framework for a Good AI Society. 2018;

[4] New York City Automated Decision Systems. Automated Decision Report Task Force Report. 2019.

[5] Ethics Commission Automated and Connected Driving. Fed Minist Transp Digit Infrastruct. 2017;

[6] DCMS. Centre for Data Ethics and Innovation . Uk Gov [Internet]. 2018;(June). Available from: https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation

[7] Digibyte. Commission appoints expert group on AI and launches the European AI Alliance | Digital Single Market. Eur Comm [Internet]. 2018;(April 2018). Available from: https://ec.europa.eu/digital-single-market/en/news/commission-appoints-expert-group-ai-and-launches-european-ai-alliance

[8] Samir Saran, Nikhila Natarajan MS. In Pursuit of Autonomy: AI and National Strategies. BMC Public Health [Internet]. 2018; Available from: https://www.orfonline.org/research/in-pursuit-of-autonomy-ai-and-national-strategies/

[9] Guide to the General Data Protection Regulation (GDPR). Guid to Gen Data Prot Regul [Internet]. 2019;(May):n/a. Available from: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/

[10] IEEE. Ethically Aligned Design: Version 2 - For Public Discussion. IEEE Stand [Internet]. 2017;1–263. Available from: https://standards.ieee.org/industry-connections/ec/ead-v1/

[11] The Public Voice. Universal Guidelines for Artificial Intelligence[internet]. 2018. Available from: https://thepublicvoice.org/ai-universal-guidelines/memo/

[12] Dignum, V. (2018). Ethics in artificial intelligence.Ethics and Information Technology: introduction to the special issue. Ethics and Information Technology, 20(1), 1–3. https://doi.org/10.1007/s10676-018-9450-z

[13] Leslie D. Understanding artificial intelligence ethics and safety. arXiv Comput Sci [Internet]. 2019; Available from: https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf