Academic Year	Module	Assessment Type
2024	Concepts of Technologies of AI	Academic Report

AI: Balancing Innovation with Ethical Integrity: Opportunities and Challenges across various fields

Student Id : 2431328

Student Name : Ugesh Kc

Section : L5CG22

Module Leader : Mr. Siman Giri

Tutor : Mr. Ronit Shrestha

Submitted on : 1/2/2025

Table of Contents

Abstract

In this paper, the research explores the risks or the ethical issues in consideration of AI with special reference to Large Language Models in the education sector. Some of the challenges are manipulation of results by AI, personality of the content generated by AI, impact on the teacher-student relationship, and the issue of fairness in the auto grading system. The described challenges call for the ethical regulation of AI creation and usage.

Harm minimization is another broad conceptual framework of international reporting that is accompanied by principles such as fairness, transparency and accountability. There are four non-negotiable components of ethical AI: inclusiveness, data privacy, and setting up environments that allow human-guided automation rather than automation dominance.

The results stress the role of international agendas, such as UNESCO's Recommendations and the EU's Ethics Guidelines for Trustworthy AI. Ethical Artificial Intelligence can therefore help improve the cooperation among its stakeholders to raise social acceptance and equality in the application of calibrated advancement within education and other sectors.

1.Introduction

Artificial Intelligence is the ability of a computer-controlled system or computer to execute a task that are commonly related with the intellectual processes feature of humans, such as the ability to reason (Copeland, 2024).

The fast adoption of Artificial Intelligence (AI) has given rise to ethical and moral challenges which must be noticed immediately. Major reasons are algorithmic bias for the increase in social inequities and opacity in Ai decision making, also known as “black box” problem. Compromise may be needed for fairness in AI because sometime making things fair at an individual level may lead to unfairness at group level and vice versa (Reuben, 2020).For handling these kind of issues global ethical guidelines and standards such as the Ethics Guidelines for Trustworthy AI from the EU have been created which focus on themes such as fairness, accountability, transparency and respect for human rights. Frameworks like these aim to guide developers for creating system which help to promote societal well-being and avoid the harm (European Commission, 2019).An ethical AI system should always provide fairness by confirming that decisions made by it are without bias and are equitable. For consumers to understand and trust its decisions AI must be reasonable and explainable. Google outlines its principles for AI in its development here, and notes how important it is to avoid unfair bias and build systems that are responsive to people. Google AI principles highlight that avoiding unfair bias and building system that are trustworthy to people is very important (Google AI, 2020). Throughout the AI lifecycle a commitment to these principles is required to develop ethical AI. This includes implementation of privacy safeguard, ensuring data protection and fostering collaboration among stakeholders for promoting responsible AI use (UNESCO, 2021).

2. LLMs and Education: Ethical Implications

Large Language Models (LLMs) are a kind of artificial intelligence that works with big amounts of data to produce text that would be considered realistic by a human reader. Such models can be useful for the individualization of education, administrative task controlling, and tutoring, so they can become the major instrument for educational system change. Nevertheless, as LLMs are gradually becoming a part of the educational process, questions related to possible bias in content, schedule, and interpersonal communication between teachers and students, as well as issues concerning the fairness of using an AI tool for grading, arise (Alkhatib et al., 2021).

One of the main ethical issues of concern with LLMs in education is the issue of bias in AI-generated content. This is because, during training, LLMs work with big data sets that can have inherent prejudice in society. These biases can be carried in AI-generated educational material content which then has a predisposition to favor one group of people while neglecting the other group of people. This is even worse in learning institutions since it reinforces a given stereotype or locks a certain group out. To this end, they must provide diverse samples of training data and follow the AI ethical principles of diversity and fairness (Friedman & Nissenbaum, 2020; Obermeyer, et al., 2021). Ensuring that the output of LLM is free from bias is to ensure each student is offered equal opportunities to learn regardless of their background.

Another teaching/learning environment where LLMs have a major function is in teacher-student interactions which are facilitated by efficient and effective personalized tutoring systems. As this can assist in spreading education as well as providing targeted assistance on student learning, there are concerns that this can reduce human teacher input. Teachers are not only education givers, but also social-emotional providers, which cannot be replaced by AI. One main disadvantage of relying on LLMs is that it will weaken the professional aspect of education, eradicating the human touch of learning that is very important today by providing substandard feedback and advice to students. Ethical consideration includes covering concerns such as how AI would augment teachers and not replace them and assist educators in doing more valuable human-related things in the teaching process (Luckin, et al., 2021). This approach would protect the essential teacher-student relation while improving the learning process with the help of AI.

When LLMs are used for operationally assessing and grading students the following several ethical issues of fairness, accuracy, and transparency are associated with it. Special attention should be

paid to the design of LLMs because they might contain bias within the grading process. AI models trained on a biased sample can unfairly assess students from disadvantaged groups. In addition, using many AI systems means it is not easy to know how the inputs and outputs of automated grading systems are related, which can lead to the loss of confidence. On these premises, we should incorporate transparent AI systems with clear grading criteria and monitor the assessment process constantly to avoid unfairness or inaccurate results (Prukaschatkun, et al., 2021; Liu, et al., 2021).

International frameworks and processes have a significant part in helping to increase the ethical use of LLMs in education. For example, the European Commission's Ethics Guidelines for Trustworthy AI mention principles such as transparency or fairness for the creation of an ethical AI system (European Commission, 2019). Similarly, UNESCO's Recommendation on the Ethics of Artificial Intelligence emphasizes the protection of AI Technologies, including LLMs, from being misused that may generate harm or contribute to increasing inequity in education (UNESCO, 2021). These guidelines let developers and policymakers provide direction for the stakeholders to develop and implement proper AI in education with no compromising on ethical values and the public interest.

3.Conclusion

The need to create ethical AI is the need to make technology competent to bring the best in human life while at the same time reducing negative consequences as well as unfair treatment in society. AI can positively impact education from the way information is taught, and coursework is delivered, to admin services and availability of resources particularly with LLMs. However, these systems posed some ethical issues like bias in an AI-created content, lack of interaction between the teacher-student, issue of fairness or bias in grading.

AI solutions must be ethical across its Design, Development, Implementation and operation by operating principles such as equity, simplicity, responsibility, and inclusiveness. To provide equal opportunities to all learners, algorithm bias eliminating is the key to pursuing. One big challenge which is related with use of digital resources in education is that teachers and other participants of the process distrust the decisions made with AI help, especially the grading systems; it is important to make these decisions more transparent.

The EU Ethics Guidelines and UNESCO guidelines call for teacher-developer-policy maker cooperation to produce AI Systems for education that respect human rights and are ethical. These frameworks use fairness equity and humanity to promote the positive use of AI in favor of humanity so that AI will be used to advance education and the general welfare of man.

Bibliography

Alkhatib et al., 2021. Ethics in AI and Education: An Exploration of Bias and Fairness in Educational Tools. *Journal of AI and Education,* 12(45-59), pp. 45-59.

Copeland, B., 2024. *artificial intelligence.* [Online]
Available at: https://www.britannica.com/technology/artificial-intelligence
[Accessed 27 12 2024].

European Commission, 2019. *Ethics Guidelines for Trustworthy AI.* [Online]
Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
[Accessed 28 12 2024].

European Commission, 2019. *Ethics Guidelines for Trustworthy AI.* [Online]
Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
[Accessed 29 12 2024].

Friedman, B. & Nissenbaum, H., 2020. Bias in Machine Learning and AI: Addressing Inequities in Educational Tools. *Journal of Technology in Society,* 25(1), p. 101–115.

Google AI, 2020. *AI principles.* [Online]
Available at: https://ai.google/responsibility/principles/
[Accessed 28 12 2024].

Liu, Y. et al., 2021. Ensuring Fairness in AI-Powered Educational Grading Systems. *Educational Technology Research,* 23(4), pp. 250-265.

Luckin, R. et al., 2021. AI in Education: The Promise and Ethical Considerations of Teacher-Student Interaction Automation. *Journal of Educational Research,* 36(2), pp. 80-95.

Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S., 2021. Dissecting Racial Bias in AI Models Used in Education. *Journal of AI in Education,* 33(2), pp. 120-134.

Prukaschatkun, Y., , & , 2021. Assessing Fairness in AI-Based Educational Systems. *Journal of Artificial Intelligence in Education,* 22(3), pp. 230-245.

Reuben, B., 2020. On the apparent conflict between individual and group fairness. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency.* Barcelona, Spain: Association for Computing Machinery, p. 514–524.

Standford, 2020. *Artificial Intelligence: Definitions and Concepts.* [Online]
Available at: https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf

UNESCO, 2021. *Ethics of Artificial Intelligence.* [Online]
Available at: https://www.unesco.org/en/artificial-intelligence/recommendation-ethics
[Accessed 28 12 2024].