Large Language Models in Education: Balancing Innovation with Ethical Responsibility

Student Name & ID: Shubhangad Shrestha (2548317)
Course: Concepts and Technologies of AI (5CS037)
Assignment: Assignment II – Term Paper on Ethical and Responsible AI
Submission Date: 17 January 2026

Abstract

Large Language Models (LLMs) are being used in many schools to help with learning, tutoring, testing, and making content. These technologies have many advantages, such as making things easier to use, faster, and more personalized, but they also raise serious ethical issues. This paper looks at how using LLMs in education can lead to problems with biased and wrong educational content, less human involvement in tutoring and feedback, and fairness and openness in automated assessment systems. It also talks about national and international efforts to promote ethical AI in education and suggests a general ethical AI framework that is specific to LLM-based educational applications. In order to guarantee educational quality, inclusivity, and student well-being, the paper makes the case that the ethical and responsible use of LLMs necessitates strict human oversight, accountability, and transparency.

Key Words:Large Language Models (LLMs),AI in Education,Ethical Issues,Human Oversight,ethical AI

Introduction

Artificial intelligence is really changing things in areas, including the way we learn (Zawacki-Richter et al., 2019). One of the important tools in artificial intelligence is something called Large Language Models (LLMs). These Large Language Models can do a lot of things like write text sounding like a person wrote it, answer questions and help with tasks that make us think. In schools, LLMs are used to make teaching materials help students learn, give students feedback on their work, assist students with their homework and make it easier for teachers to grade assignments and figure out how well students are doing. These apps provide innovation by improving productivity, accessibility, and customization in educational settings.

However, the increasing use of LLMs in education increases significant ethical, social, and moral concerns of different fields. Main ethical challenges include bias included within AI-generated content, lack of transparency in automated decision-making, accountability for AI-driven results, and the potential erosion of human roles in teaching (Khalil et al., 2024). Since education directly shapes knowledge, values, and future opportunities, unethical use of AI in this sector can have long-term consequences for individuals and society.

To address those risks, different ethical norms, guidelines have been suggested; these include fairness, transparency, accountability, privacy, and human oversight (UNESCO, 2021; OECD, 2019). A good ethical system for AI would be one that is reliable, free from discrimination, explainable in decision-making processes, and one where humans are responsible for the use of the system. For these systems to require ethical standards and learning objectives, the developers must oversee the systems' performance.

In this paper, we concentrate on the ethical issues connected with the adoption of LLMs in the educational field. We analyze the issues of bias and accuracy in educational material, the

automation of tutoring, AI grading, and ethical initiatives, concluding with an ethical framework for the use of AI in education.

Thematic Review – Large Language Models (LLMs) and Education

Bias and Accuracy in Educational Content

LLMs have been employed for creating learning content such as explanations, summaries, quizzes, and examples. Although it is beneficial for increasing learning efficiency, it also poses some ethical issues with regard to bias and accuracy. LLMs have been trained using a large amount of data that contains biases and limitations found in web-based sources (Khalil et al., 2024). Thus, AI-based learning content may result in giving priority to dominant major ideas while ignoring minor ideas, cultures, or languages (Zawacki-Richter et al., 2019). It may have a negative impact on learners who belong to various backgrounds.

Accuracy is another important issue that should not be ignored. It is proven that sometimes LLMs produce incorrect or misleading data when conveying it with certainty (Yan et al., 2023). In an academic setting, especially when it comes to subjects such as science, math, or history, this misinformation can result in misleading students and impairing the quality of education being imparted to them.

Ethically, it is the duty and responsibility of educational institutions to make sure that educational resources provided are accurate and credible. Human validation of AI-generated educational resources is thus imperative to prevent negative consequences and preserve integrity.

Automation of Tutoring and Feedback

There has been a personalized feedback feature that is offered by LLM-based tutoring systems that makes them very attractive for online teaching, as well as for large class teaching. Tutoring systems have the potential to assist students outside classroom time. But there are concerns about reducing human teacher participation because of automation in tutoring.

Learning and teaching involve not only the transmission of information but also mentorship, emotional support, and motivation (Zawacki-Richter et al., 2019). The lack of emotional intelligence and an ability to relate to students' personal difficulties in learning contexts is a limitation of AI-powered tutors.Overreliance on AI-powered tutoring might undermine student-teacher relationships (Khalil et al., 2024).

Another potential danger could be that of the students becoming too dependent on these AI-powered tools and using them as a means of avoiding work rather than using them as a supplement for learning purposes. The proper implementation of AI tutoring systems can be achieved by having a balanced approach where LLMs can be a supportive tool to teachers

rather than a replacement for them.Maintaining proper oversight of LLMs makes sure AI improves learning without downplaying the role of teachers.

AI-Based Assessment and Grading

The use of LLMs for automated assessment and grading is being increasingly explored. However, such systems pose serious questions from an ethics perspective. AI-based grading systems could adversely affect students who use non-standard language, approach from different cultural backgrounds, or think outside the box (Khalil et al., 2024) and not just within it.

Another issue is the lack of transparency in many AI grading tools that are considered black boxes (NIST, 2023). Students are not in a position to know how their grades have been calculated or contest a grade if it is incorrect (OECD, 2019). Lack of transparency affects the fairness of the grading system. Accountability is also a problem where grades are issued by algorithms despite the ethical duty resting with educators.
To ensure that there are proper ethics involved in the evaluation process, it is recommended that AI evaluation tools be utilized only as a supplement to the evaluation process and not solely relied upon.

Additional Ethical Initiatives and Proposed Framework

Some of the initiatives taken by the global community for ethical use of AI in the field of education include UNESCO's Guidelines for Education with AI, OECD AI Principles, the EU AI Act, and the NIST AI Risk Management Framework (UNESCO, 2021; OECD, 2019; NIST, 2023; European Commission, 2023).

On the basis of these tenets, a generic ethical AI framework for LLMs in education must incorporate the following:

- **Fairness**: Bias Survey and Inclusive Data for Training.
- **Transparency**: Explanation of AI model results and evaluation criteria.
- **Human Oversight**: Teachers maintain control over decisions and outcomes.
- **Accountability**: Institutions are accountable for the outcome
- **Sustainability**: To be responsible and socially conscious about AI usage.

Discussion / Personal Reflection

This research work highlights why it is important for ethics to be taken into consideration when implementing AI and, more specifically, LLMs into educational settings. The reason for this is that education is a critical aspect of forming the knowledge and values of learners and providing them with future prospects, and as such, unethical use of AI would have negative and long-term consequences. Through this research work, I have learned that despite the many benefits that LLMs provide for enhancing accessibility and efficiency, there are also negative aspects such as bias, misinformation, decreased human contact, and unfair assessment.

A major point that needs to be considered is that responsible AI in education is not about avoiding technology but about how it is utilized. The need for human intervention is essential in order to make sure that AI is utilized for learning and not for replacing some essential roles that humans play.

In the long run, adopting ethical AI can ensure that there is equality in the usage of innovations in the field of education, while failure to consider ethics can lead to increased inequalities and poor standards of education. For this reason, ethical AI development should be treated as a core requirement rather than an optional consideration in the educational field.

References

Khalil, M., Prinsloo, P., & Slade, S. (2024). Ethical considerations in the use of generative artificial intelligence in education. Computers and Education: Artificial Intelligence, 5, 100165.

NIST. (2023). Artificial intelligence risk management framework (AI RMF 1.0). National Institute of Standards and Technology.

OECD. (2019). OECD principles on artificial intelligence. Organisation for Economic Co-operation and Development.

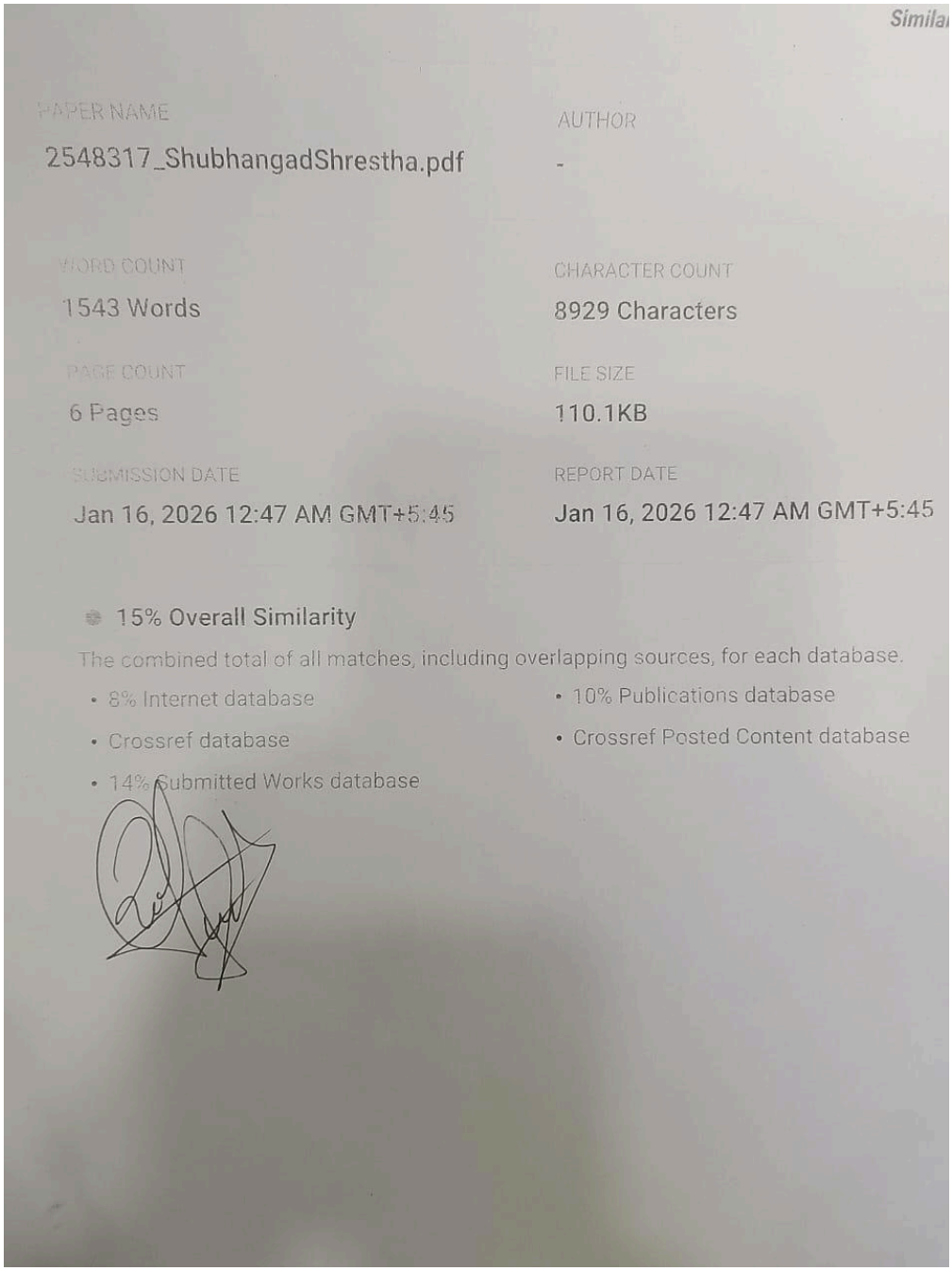
UNESCO. (2021). Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific and Cultural Organization.

Yan, D., Li, Y., & Wang, J. (2023). Hallucinations and factual errors in large language models: Implications for education. Educational Technology Research and Development, 71(5), 2381–2396.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16(39).

European Commission. (2023). Regulation (EU) 2023/… laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).

Appendix



Simila

PAPER NAME

2548317_ShubhangadShrestha.pdf

AUTHOR

-

WORD COUNT

1543 Words

CHARACTER COUNT

8929 Characters

PAGE COUNT

6 Pages

FILE SIZE

110.1KB

SUBMISSION DATE

Jan 16, 2026 12:47 AM GMT+5:45

REPORT DATE

Jan 16, 2026 12:47 AM GMT+5:45

15% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 8% Internet database
- 10% Publications database
- Crossref database
- Crossref Posted Content database
- 14% Submitted Works database