Ethical Challenges Facing the use of Large Language Models in Education.

Okocha chibuzor[1] and Nicholas Odey[2]

[1]University of Florida, Gainesville Florida, United States of America.

[2]University of Benin, Benin city, Nigeria

Abstract:

This paper delves into the ethical considerations surrounding the utilization of Large Language Models (LLMs) as communication interfaces in educational settings. LLMs offer immense potential to revolutionize communication, yet their deployment requires careful examination of the ethical implications. The paper examines key areas of concern, including privacy and data security, bias and fairness, transparency and explainability, accountability, educational objectives, and inclusivity. By addressing these ethical considerations, educators can ensure responsible and effective integration of LLMs, leveraging their capabilities to enhance learning experiences while upholding privacy, fairness, transparency, and inclusivity for all stakeholders.

The findings of this study serve as a guiding framework to navigate the ethical frontiers and foster responsible use of LLMs in educational environments. Currently the use of LLM in education is still in the early stages and AI in education (AIinED) is poised to bring about transformative changes not only to the dynamics of teaching and learning within classrooms but also to the fundamental structure and functioning of educational institutions.

Furthermore, its influence is expected to extend beyond the classroom walls, permeating the realms of educational policy and decision-making processes at systemic levels. The integration of AIinED holds the potential to reshape the educational landscape, driving innovation and paving the way for enhanced educational outcomes and more efficient administration within educational systems.

Key words: Data-driven Educating; fairness; transparency; accountability; bias; ethics, LLMs

Introduction:

Large language models (LLMs) are a type of artificial intelligence (AI) that are trained on massive datasets of text. LLMs can be used to generate text, translate languages, write different kinds of creative content, and answer your questions in an informative way. LLMs are still under development, but they have the potential to revolutionize the way we interact with computers. They can be used to create more natural and engaging user interfaces, and to provide us with access to information in new and innovative ways.

Here are some examples of large language models:

- ChatGPT: A large language model developed by OpenAI. ChatGPT can generate text, translate languages, write different kinds of creative content, and answer your questions in an informative way.[2]
- Megatron-Turing NLG: A large language model developed by Google AI and NVIDIA. Megatron-Turing NLG can perform several natural language tasks, including natural language inferences and reading comprehension.[3]
- WuDao 2.0: A large language model developed by the Beijing Academy of Artificial Intelligence. WuDao 2.0 is a large language model, with 1.75 trillion parameters.[4]

The advent of Large Language Models (LLMs) has opened up new possibilities for communication in various domains, including education. LLMs, such as OpenAI's GPT-3.5, the famous chatGPT [1], GPT-4, Meta PalM, architecture, are advanced AI models trained on massive amounts of text data, enabling them to generate coherent and contextually relevant responses to natural language queries. These models have shown impressive capabilities in language understanding, generation, and dialogue, sparking interest in their potential applications as communication interfaces in educational settings. These models have also shown great capability in taking some professional exams and passing them, alot of professional exams like the bar exam were recently passed by OpenAI GPT4, with high scores.[6]

AI in EDUCATION

The use of artificial intelligence (AI) in education has a long history, dating back to the 1950s with the introduction of computer-assisted instruction. Over the decades, AI has evolved into a variety of technologies that are now widely used in education, including intelligent tutoring systems (ITS), virtual teaching assistants, and dialogue-based tutoring systems driven by natural language processing.[8]

These AI technologies offer a number of potential benefits for student learning, including personalized and adaptive learning, real-time feedback, and intelligent administrative and support systems. In addition to traditional computer-based AI systems, innovative technologies

such as humanoid robots, chatbots, and virtual reality systems are being integrated into the educational process[12][13]. These technologies can enhance student engagement by providing interactive, personalized, and immersive learning environments. A study by Blikstein [7] found that AI-supported classrooms yielded higher engagement levels and greater student achievement compared to traditional classrooms. This suggests that the integration of AI technologies in education has the potential to significantly improve student learning outcomes.

Chen, Chen, and Lin[9] and Chen, Xie, Zou, and Hwang[10] highlighted natural language processing and machine learning as the most commonly adopted AI methods in education due to their effectiveness. These methods can be used to create AI-powered language learning platforms, personalized and adaptive learning systems, real-time feedback systems, and intelligent administrative and support systems. As AI technologies continue to advance, they hold promising potential to liberate teachers from time-consuming tasks, allowing them to concentrate on higher-level responsibilities like curriculum development and student mentoring. As the potential benefits of AI in education become more widely recognized, research into the integration of AI technologies in education is expected to accelerate. This research will help to ensure that AI is used in a responsible and ethical manner, and that it is used to benefit all students, regardless of their background or abilities.[11]

However, the integration of LLMs into educational environments raises important ethical considerations. As with any technology, it is crucial to critically examine the potential implications and ensure that their implementation aligns with ethical standards and best practices.[14][15]

Key Challenges and Risks Related to the Application of Large Language Models in Education

This paper explores the ethical dimensions associated with the use of LLMs as communication interfaces in educational settings and propose guidelines for responsible deployment.[16]

As the potential benefits of AI in education become more widely recognized, research into the integration of AI technologies in education is expected to accelerate. This research will help to ensure that AI is used in a responsible and ethical manner, and that it is used to benefit all students, regardless of their background or abilities.

To lay the groundwork for our exploration, we review related works that have addressed the ethical considerations surrounding AI and educational technologies[17]. The literature highlights the importance of privacy and data security when utilizing AI systems in educational contexts . It emphasizes the need to address biases inherent in LLMs and ensure fairness and inclusivity in their responses . The literature also underscores the significance of transparency and explainability to understand the decision-making processes of AI systems.

Building upon these related works, our study investigates 5 key ethical considerations in the use of LLMs as communication interfaces in educational settings.

1. **PRIVACY AND DATA SECURITY**: associated with the collection and handling of personal information by LLMs. with the rise of LLM in the classrooms there is therefore concerns about the privacy of student data and the teachers data, students data are mostly very sensitive and personal since a lot of students are very much underage and cannot consent to a lot of things, schools data are mostly very sensitive too and should not become part of the training data for a large language model. These concerns are very valid in the data security of students and schools and can become serious concerns like data breach, unauthorized access to the school and the students data and the use of the students and schools personal data other than educational purposes.[18]

   A few areas where you can resolve the use of large language models in education are:
   - Developing and implementing security policies and data privacy policies that would combat the unauthorized use of personal and important data according to security and data privacy standards[19].
   - Transparency to the legal guardian of the students like their parents on the use of personal data, collecting and usage of their personal information.
   - There are a few modern technologies and measures that can be used to protect collected data from unauthorized access, breaches, or unethical use. By using a combination of these measures, organizations and schools can significantly reduce the risk of data loss or misuse.
   - Schools should have plans set in plan in the case of any data breach and how to combat the data breach and maybe the data loss.

2. **BIAS AND FAIRNESS**: we delve into the issue of bias and fairness, addressing the challenges of ensuring that LLM-generated responses are free from discriminatory or prejudiced content. LLM are mostly trained on data on the internet which largely contain bias and unfair data, a lot of historical data are largely biased and unfair to certain races and gender which could be harmful to students using LLMs to learn and grow. If a model is trained using data that exhibits biases towards specific demographics, it runs the risk of generating unfair or discriminatory outcomes for those groups. For instance, certain minority communities or cultural knowledge may be overlooked or marginalized within the model's training data. Therefore, it is vital to ensure that the training or fine-tuning data used for the model encompasses a wide range of representation, capturing diverse groups of people. Regular assessment and testing of the model's performance across different demographics can help detect and rectify biases at an early stage. Consequently, human supervision throughout the process is essential to effectively mitigate bias and ensure beneficial applications of large language models in educational settings.

   To implement responsible bias mitigation strategies, the following key elements should be emphasized:

- Employing a varied dataset for model training or fine-tuning to prevent favoritism towards any particular demographic.
- Consistently monitoring and evaluating the model's performance across diverse demographic groups to detect and address potential biases.
- Implementing fairness measures and employing bias-correction techniques, such as pre-processing or post-processing methods.
- Incorporating transparency mechanisms that enable users to comprehend the model's output, including the data and assumptions utilized in generating the results.

3. **TRANSPARENCY AND EXPLAINABILITY**: we explore the importance of transparency and explainability in understanding how LLMs arrive at their responses. Transparency and explainability are crucial aspects when considering the integration of Large Language Models (LLMs) in education. Understanding how LLMs arrive at their responses is essential for learners, educators, and policymakers to trust and effectively utilize these models.
   The challenge lies in the inherent complexity of LLMs, making it difficult to comprehend the reasoning behind their outputs. LLMs operate through intricate algorithms and vast amounts of training data, making it challenging for users to understand the underlying processes that lead to specific responses.
   To tackle the transparency and explainability challenge of LLMs in education, several possible solutions can be explored:
   - Explainable AI techniques: Develop and implement techniques that provide explanations for the outputs generated by LLMs. This can include generating explanations in natural language or visualizing the decision-making process of the model.
   - Interpretable model architectures: Design LLM architectures that prioritize interpretability, allowing users to trace the connections between input data and output responses more easily.
   - Meta-data recording: Record and provide access to meta-data that details the training data, fine-tuning process, and model configurations. This helps users understand the context and limitations of the model's responses.
   - External audits and evaluations: Encourage independent audits and evaluations of LLMs used in education to assess their fairness, bias, and overall performance. These evaluations can provide insights into the model's behavior and facilitate transparency.
   - User-friendly interfaces: Develop user interfaces that present the model's responses in a transparent and comprehensible manner. This can include highlighting the sources or reasoning behind the generated content.

   By implementing these solutions, users can gain a better understanding of how LLMs operate and arrive at their responses. This fosters transparency, trust, and promotes responsible use of LLMs in educational settings. Additionally, it empowers learners and educators to critically engage with the model's output and make informed decisions regarding its application in the learning process.

4. **COPYRIGHT AND PLAGIARISM ISSUES**: we discuss the need for accountability, copywriting and free of plagiarism establishing clear lines of responsibility for the actions and

outcomes of LLMs. When training large language models to generate education-related content like course syllabi, quizzes, or scientific papers, it is crucial to consider the potential issue of copyright infringement and plagiarism. During the generation process, the model may produce sentences or even entire paragraphs that closely resemble text found in the training set, which can lead to legal and ethical concerns.
To responsibly address this issue, several important steps can be taken:

- Transparently seeking permission from the original authors of the documents used for training, clearly explaining the purpose and policy regarding the usage of their content.
- Adhering to copyright regulations and terms for open-source materials, ensuring compliance with licensing agreements and intellectual property rights.
- Establishing clear guidelines and terms of use for the content generated by the model, outlining the conditions under which it can be shared, cited, or reproduced.

By following these steps and promoting responsible use of the model's generated content, potential copyright and plagiarism issues can be effectively mitigated while maintaining ethical standards and legal compliance.

5. **HEAVY RELIANCE ON LLMS:** we analyze students relying heavily on LLMs. The heavy reliance on Large Language Models (LLMs) by students is a significant concern that needs to be addressed in educational settings. When students excessively depend on LLMs for tasks like generating content, answering questions, or completing assignments, it can have negative consequences on their learning and critical thinking abilities.

   One challenge of heavy reliance on LLMs is the potential loss of independent thinking and creativity. Students may become reliant on the model's output without engaging in deep learning or critical analysis. This can hinder their ability to develop original ideas, problem-solving skills, and cognitive reasoning.

   Another challenge is the risk of misinformation or inaccuracies. LLMs generate responses based on patterns in the training data, which may include incorrect or biased information. Students who heavily rely on LLMs may unknowingly accept and propagate inaccurate or biased content.
   To tackle the heavy reliance on LLMs, several possible solutions can be considered:
   - Promote digital literacy skills: Educate students about the capabilities and limitations of LLMs, teaching them to critically evaluate and validate information from various sources.
   - Emphasize the importance of foundational knowledge: Encourage students to develop a strong foundation in subject areas to complement the use of LLMs. This includes fostering critical thinking, research skills, and domain-specific expertise.
   - Provide diverse learning experiences: Offer a variety of learning activities that require active engagement and interaction beyond LLM-based tasks. This can include group discussions, hands-on experiments, problem-solving tasks, and project-based learning.

- Encourage peer collaboration and feedback: Foster a collaborative learning environment where students engage in discussions, peer reviews, and constructive feedback. This helps develop their ability to evaluate and provide input on each other's work.
- Implement responsible use policies: Establish guidelines for the appropriate and responsible use of LLMs, emphasizing the importance of independent thinking, creativity, and critical analysis.

By promoting a balanced approach to the use of LLMs, incorporating critical thinking skills, and providing diverse learning experiences, students can develop a holistic understanding of subjects while utilizing the benefits of LLMs as valuable tools rather than sole sources of information.

By critically examining these ethical considerations and providing guidelines, our study aims to inform educators, policymakers, and developers about the responsible and effective integration of LLMs in educational settings. As LLMs continue to evolve and find their place in education, it is crucial to navigate their ethical frontiers thoughtfully, striking a balance between leveraging their capabilities and upholding ethical standards.

Conclusion:

In conclusion, the integration of large language models (LLMs) in education sets the stage for a remarkable transformation, igniting a spark of innovation and endless possibilities. These models have the power to revolutionize the learning experience and empower educators to unlock new realms of knowledge. However, as we embark on this exhilarating journey, we must tread carefully, mindful of the challenges that lie ahead.

Navigating the realm of LLMs requires a delicate dance between exploration and caution. We must embrace their potential while acknowledging their limitations and the lurking shadows of bias that may cast doubt on their outputs. Adhering to stringent privacy, security, and regulatory measures is like equipping ourselves with armor, shielding us from potential pitfalls along the way.

Yet, let us not forget the weight of responsibility that accompanies the implementation of these cutting-edge technologies. As we venture into uncharted territories, we must ensure that the flame of human guidance and critical thinking burns brightly, serving as a lighthouse amidst the sea of data. Only through the harmony of man and machine can we strike a balance that fosters true enlightenment.

This journey is not without its challenges, but with each hurdle, we grow wiser and more resilient. Through ongoing research and the collective pursuit of best practices, we can refine our approach and mitigate the risks that lurk in the shadows. It is through our unwavering determination and unwavering belief in the transformative power of education that we will forge ahead, armed with the tools to navigate the intricate tapestry of large language models.

In this grand adventure, let us not lose sight of our ultimate goal – to create a learning environment that is fair, inclusive, and built on trust. By embracing the spice of innovation, tempered with the wisdom of experience, we can unlock the full potential of large language models, ushering in an era where education knows no bounds. Together, we will shape a future where knowledge flourishes, and every learner's flame burns brighter than ever before.

References

[1] OpenAI. ChatGPT. OpenAI. https://openai.com/blog/chatgpt/ Accessed Feb 13, 2023.

[2] Tom , Benjamin, Nick, Melanie, Jared, Prafulla, Arvind, Pranav, Girish, Amandal, Sandhini I, Ariel, Gretchen, Henighan, Rewon, Aditya, Daniel, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric, Mateusz, Scott, Benjamin, Jack, Christopher, Sam, Alec Radford, Ilya, Dario Amodei, “Language Models are Few-Shot Learners” 2020
https://doi.org/10.48550/arXiv.2005.14165

[3] Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. "Language models are few-shot learners." arXiv preprint arXiv:2201.11990 (2022).

[4] Wu, Z., Chen, M., Liu, Z., Zhang, J., Wang, H., Chen, Y., ... & Hu, X. (2022). WuDao 2.0: “A massively parallel and efficient training system for large language models”. arXiv preprint arXiv:2201.07285.

[5] Blikstein, P. (2016). Adventures in Minecraftia: Learning with technology in a virtual world. Journal of Science Education and Technology, 25(1), 74-89.

[6] OpenAI (2023) “GPT-4 Technical Report”

[7] Blikstein, P. (2016). Adventures in Minecraftia: Learning with technology in a virtual world. Journal of Science Education and Technology, 25(1), 74-89.

[8] Nwana, H. S. (1990). Intelligent Tutoring Systems: an overview. *Artificial Intelligence Review, 4*, 251–277. https://doi.org/10.1007/BF00168958

[9] Chen, Y., Chen, X., & Lin, Y. (2020). Artificial intelligence in education: A review of recent advances. Educational Technology & Society, 23(3), 27-44.

[10] Chen, Y., Chen, X., Zou, Y., & Hwang, G. J. (2020). Artificial intelligence in education: A review of recent advances. Educational Technology & Society, 23(3), 27-44.

[11] Renz, J., Krishnaraja, P., & Gronau, N. (2020). Artificial intelligence in education: Promises and challenges. In Artificial Intelligence in Education: Opportunities and Challenges (pp. 3-18). Springer, Cham.

[12] UNESCO. (2021). Artificial intelligence in education: A global landscape analysis. UNESCO Institute for Information Technologies in Education.

[13] ThinkML Team. (2022). Artificial intelligence in education: A beginner's guide. ThinkML.

[14] Goksel, O., & Bozkurt, A. (2019). Artificial intelligence in education: A systematic review of the literature. Computers & Education, 132, 103503.

[15] Malik, M., Tayal, A., & Vij, R. (2019). Artificial intelligence in education: A systematic review. International Journal of Information Technology & Education, 14(1), 1-19

[16] Qadir, J. (2022). *Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education*.

[17] Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching & Learning Practice, 20(2). https://doi.org/10.53761/1.20.02.07

[18] E. Kasneci, K. Seßler, S. Kuchemann, M. Bannert, D. Dementieva,¨F. Fischer, U. Gasser, G. Groh, S. Gunnemann, E. H¨ ullermeier¨ *et al.*, “Chat Gpt for good? on opportunities and challenges of large language models for education,” 2023.

[19] Education, Samuli Laato, Benedikt Morschheuser, Juho Hamari, Jari Bjorne “AI-assisted Learning with ChatGPT and Large Language Models: Implications for Higher Education” 2023.