UNIVERSITY OF
WOLVERHAMPTON

5CS037

Concepts and Technologies of AI

Name: Swoyam Pokharel

Group: L5CG26

Canvas ID: 2431342

Tutor: Nabin Acharya

Table Of Contents:

Abstract:

Artificial Intelligence (AI) has rapidly become a topic of high relevance in recent years especially in the educational sector with tools like ChatGPT which is widely adopted. This paper questions this adoption by highlighting the critical issues that come with the use of artificial intelligence. This paper touches on the “black box” nature of AI which hinders explainability and justifiability, the lack of emotional intelligence and adaptability. Furthermore, this paper touches on the biases in training data which results in the AI producing stereotypical responses. Additionally this paper highlights concerns about outdated information, unethical training practices, copyright and privacy infringements which challenge AI’s place in the academic field. This report also attempts to create a set of guidelines building on top of already existing guidelines such as the European Union Guidelines to help build better AIs in general.

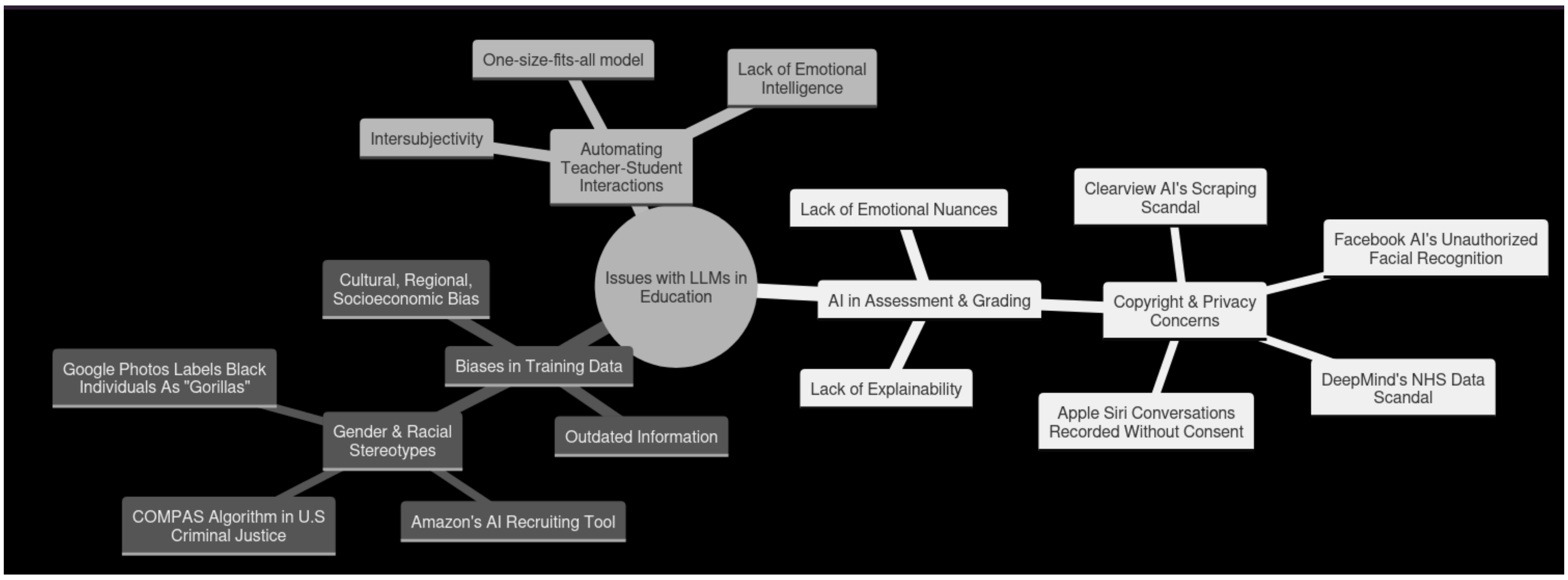
Introduction:

Artificial Intelligence (AI) has been a highly trending topic, especially in the past few years, thanks to tools like Chat-GPT and Midjourney; and rightfully so. These AI tools have had a profound impact in the way we do things, especially in the educational sector. Tools like Chat-GPT have become a one-stop-shop for learning and understanding basically any topic. However, with this widespread adoption, comes some serious concerns regarding the use of AI.

AI works as black boxes, especially in the case of neural networks. While this may sound trivial, it challenges a fundamental human trait. Humans have the ability to think through a problem, and explain their thought process every step of the way, AI lacks that. This simple inability to explain how it came to a certain conclusion makes it so much harder to justify those answers (Trausan-Matu, 2020, p. 6). That along with other concerns raise questions about the use of AI.

The following sections of this report will attempt to highlight more of these dilemmas and provide possible solutions especially in the educational field.

The Problem with LLMs in Education:


One-size-fits-all model
Lack of Emotional Intelligence
Intersubjectivity
Automating Teacher-Student Interactions
Clearview AI's Scraping Scandal
Lack of Emotional Nuances
Facebook AI's Unauthorized Facial Recognition
Cultural, Regional, Socioeconomic Bias
Issues with LLMs in Education
AI in Assessment & Grading
Copyright & Privacy Concerns
Google Photos Labels Black Individuals As "Gorillas"
Biases in Training Data
DeepMind's NHS Data Scandal
Lack of Explainability
Apple Siri Conversations Recorded Without Consent
Gender & Racial Stereotypes
Outdated Information
COMPAS Algorithm in U.S Criminal Justice
Amazon's AI Recruiting Tool

The issue with LLMs for generating educational content.

AI simply put, is pattern recognition on a large scale. It trains on a set of data, recognizing a pattern and leverages that pattern to generate further responses. This inherently implies that all knowledge of the AI is based on the data it is trained upon. Xu mentions that LLMs can reflect biases present in the data it was trained upon (Xu et al., 2024, p. 15) such as:

Cultural, Regional and Socioeconomic Bias

LLMs might "accidently" favour a more dominant cultural narrative which can suppress the voices of the minority. For example, historical events might be presented more from a Western viewpoint.

Furthermore, AI is simply not accessible for the lower socioeconomic backgrounds, AI-driven learning tools like Duolingo needs a stable internet access and capable personal devices, which may not be available to many students (Geiger, 2018).

Gender and Racial Stereotypes

Due the the same reason that Xu mentions, (Xu et al., 2024, p. 15) LLMs may produce stereotypical answers in its response,like associating professions with specific genders or simply bias in any way or form; see: :Amazon's AI Recruiting Tool, COMPAS Algorithm in U.S. Criminal Justice, Google Photos Labels Black Individuals as "Gorillas".

Outdated Information:

AI models only know as much information as their training data contains which over time it still fades out of relevancy. This can cause the models to generate outdated or no longer relevant responses.

The Issue With Automating Teacher-Student Interactions

Reduced human involvement in any field, especially one that is as nuanced as education is unjustified to say the least. Education isn't simply "teach stuff", it involves multiple layers such as:

Intersubjectivity

Intersubjectivity, which is the ability to understand each other's perspectives, is fundamental to human learning (Bailey, 2003, p.11). This understanding is possibly what makes us human and it's something that AI simply cannot replicate.

One-size-fits-all model:

Our brains have been trained to recognize patterns; sure AI is better at that than us, but what it lacks is a nuanced and a more emotional understanding. Teachers are so great because they can tailor to each student's learning patterns and their emotional state, which again something AI cannot replicate. This flows nicely to our next point,

Lack of Emotional Intelligence:

Teachers unlike AI can understand cues like body language or tone of voice that can give insight on how or why a student is behaving a certain way. This insight allows a teacher to adjust based on the student's state, which AI can’t do.

The Issue with AI in Assessment and Grading

Lack of emotional nuances and understanding of creative answers

Previously we have already established that AI cannot capture emotional nuances. This might cause it to misjudge creative or unconventional answers albeit correct

Lack of explainability

But, there's another significant flaw: the lack of explainability and transparency. Because AI works as a "black box", there is no one clear explanation as to why it arrived at a certain answer. This lack of explainability makes it a very poor entity to grade any assessment (Trausan-Matu, 2020, p. 6). Also, AI can inadvertently favor certain phrasing simply because it has appeared more frequently in "statistically good" training data, which will skew its grading data.

Copyright and Privacy Concerns:

Information ownership is of utmost importance in academics. AI provides responses based on the data it was trained upon, data collected from multiple individuals. This raises a critical question, who is to be credited to the insights or any breakthroughs made with the help of AI. This concern alone raises questions about whether using AI in the academic field is even appropriate.

Furthermore, is it responsible to use AI that is trained upon unethical data? Incidents like: Clearview AI's Data Scraping Scandal, Facebook AI's Unauthorized Facial Recognition, DeepMind's NHS Data Scandal, Apple Siri Conversations Recorded Without Consent raise important questions. Granted that these issues concern the creators behind the AI rather than the technology itself, they still remain very important questions.

Strategies For Improvement

Having established that AI has clear room to improve, now the question boils down to the how?. While issues like emotional intelligence are hard to tackle, many of the other challenges can be solved with a few strategies. Acknowledging already established regulations such as European Union Guidelines, the following points aim to build upon these strategies to expand those principles especially in the context of education.

- Diverse and Inclusive Training Data: We've shown that AI is prone to stereotypes and biases. However, these flaws can be improved dramatically by fixing the training data. AI extracts its knowledge based on its training data, every stereotype, every bias it produces is because it was prevalent in the training data. Then to solve this, the creators ought to incorporate a diverse range of training data; data that is inclusive, data that is just and data that has been "treated" best it can.

- In: AI evidently is not going to be a drop in replacement for most human endeavors, especially not education. The best use for it, at this day and age, is to reduce the boilerplate work. So, instead of trying to forcefully use it where its not fit, it should rather be seen and used as a tool that supports learning. This approach maintains the essential "human" element to learning.

- Regular Audits and Updates: The creators of the AI should continuously monitor and update their AI to address issues. These issues might range from big mishaps: Gemini Tells User to Die to the simple need for regular updates that adheres to the ever changing societal norms. This ensures that AI remains fair and accurate

- Accountability: The underlying workings of AI should be informed to the general user. Only when the users know how these tools work, will they understand why it produces a certain response. Furthermore, creators of these AI tools shouldn't portray it as a solution to everything. They should accept and advertise the fact that their AI might make mistakes, and when they do, there need to be serious repercussions.

Discussion:

To conclude, as of now, AI is not a viable replacement in the educational sector. While it is great at handling repetitive, tedious tasks or even assisting with education, it falls short when it comes to actual teaching and education as a whole. Its "black box" nature makes it difficult to justify its responses, which in a field like education where challenging opinions promote progress, is a drawback. Furthermore, it's lack of emotional intelligence, empathy and adaptability that teachers have is something crucial when it comes to teaching. Also it goes without saying that AI's generating stereotypical statements doesn't help. Given these drawbacks, rooting for LLMs in Education at this current state, would be an ill-informed statement. Instead, AI should be viewed as an assistive tool.

References:

Bailey, R., 2003. *Learning to be Human: teaching, culture and human cognitive evolution*. [online] UCL Press. Available at: <https://www.researchgate.net/publication/248933138_Learning_to_be_Human_teaching_culture_and_human_cognitive_evolution> [Accessed 27 December 2024].

CBC News, 2021. *Facebook plans to shut down facial-recognition system, delete data*. [online] CBC. Available at: <https://www.cbc.ca/news/world/facebook-face-recognition-shutdown-1.6234288> [Accessed 29 December 2024].

Dastin, J., 2018. Insight - Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. [online] 11 Oct. Available at: <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/> [Accessed 2 January 2025].

Duolingo, 2012. *Learn a language for free*. [online] Duolingo. Available at: <https://www.duolingo.com> [Accessed 2 January 2025].

Europian Union, 2019. *Ethics guidelines for trustworthy AI*. [online] Shaping Europe's digital future. Available at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> [Accessed 27 December 2024].

Geiger, A., 2018. Nearly one-in-five teens can't always finish their homework because of the digital divide. *Pew Research Center*. [online] 26 Oct. Available at: <https://www.pewresearch.org/short-reads/2018/10/26/nearly-one-in-five-teens-cant-always-finish-their-homework-because-of-the-digital-divide/> [Accessed 27 December 2024].

Hart, R., 2024. Clearview AI—Controversial Facial Recognition Firm—Fined $33 Million For 'Illegal Database.' *Forbes*. [online] 3 Sep. Available at: <https://www.forbes.com/sites/roberthart/2024/09/03/clearview-ai-controversial-facial-recognition-firm-fined-33-million-for-illegal-database/> [Accessed 1 December 2024].

News, B., 2015. *Google apologises for Photos app's racist blunder*. [online] BBC News. Available at: <https://www.bbc.com/news/technology-33347866> [Accessed 31 December 2024].

News, B., 2021. *DeepMind faces legal action over NHS data use*. [online] BBC News. Available at: <https://www.bbc.com/news/technology-58761324> [Accessed 29 December 2024].

ProPublica, 2016. *Machine Bias*. [online] ProPublica. Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [Accessed 30 December 2024].

Sky News and Carroll, M., 2024. Google's AI chatbot Gemini tells user to "please die" and "you are a waste of time and resources." *Sky*. [online] 19 Nov. Available at: <https://news.sky.com/story/googles-ai-chatbot-gemini-tells-user-to-please-die-and-you-are-a-waste-of-time-and-resources-13256734> [Accessed 1 January 2025].

Su, J., 2019. Apple Apologizes For Eavesdropping On Customers, Keeping Siri Recordings Without Permission. *Forbes*. [online] 28 Aug. Available at: <https://www.forbes.com/sites/jeanbaptiste/2019/08/28/apple-apologizes-for-eavesdropping-on-customers-keeping-siri-recordings-without-permission/> [Accessed 28 December 2024].

Trausan-Matu, S., 2020. *Ethics in Artificial Intelligence*. [online] Matrix Rom. Available at: <https://www.researchgate.net/publication/351486496_Ethics_in_Artificial_Intelligence> [Accessed 24 December 2024].

Xu, H., Gan, W., Qi, Z. and Yu, P.S., 2024. *Large Language Models for Education: A Survey*. [online] unknown. Available at: <https://www.researchgate.net/publication/380821593> [Accessed 25 December 2024].