

Fairness, Accountability, Transparency, and Empowerment: Establishing Ethical Principles for AI in Higher Education

Dr. Anna Bougia, Dr. Anastasios Liapakis, Mr. Konstantinos Chardavelas

New York College Athens

Introduction

- The integration of Artificial Intelligence (AI) into higher education is transforming the way educational institutions operate, but this transformation comes with significant ethical challenges. This study emphasizes the need to address these ethical challenges, ensuring that AI tools in education are deployed responsibly and with clear ethical oversight
- Following the European Union's guidelines for AI ethics, our research aims to develop an ethics framework specifically for AI integration in education, focusing on student well-being, transparency, and autonomy
- This presentation will explore key ethical dimensions including student agency, privacy, prevention of harm, and explainability in the context of AI in education
- The focus of data collection will be four topics: Student Agency, Privacy Concerns, Prevention of Harm, and Explainability, which are topics proposed as Ethics focus points for Artificial Intelligence in education by Nguyen *et al.*, (2023)

Introduction

Key Questions:

- 1. Ethical concerns about AI in higher education
- 2. Perception of harm prevention with AI
- 3. Interpretation of ethical AI implementation

Project Aim:

- - Establish a student-centered ethical framework for AI use in higher education

Objectives:

- 1. Identify student ethical concerns on AI.
- 2. Examine views on harm prevention during AI implementation.
- 3. Understand student perspectives on ethical AI use.



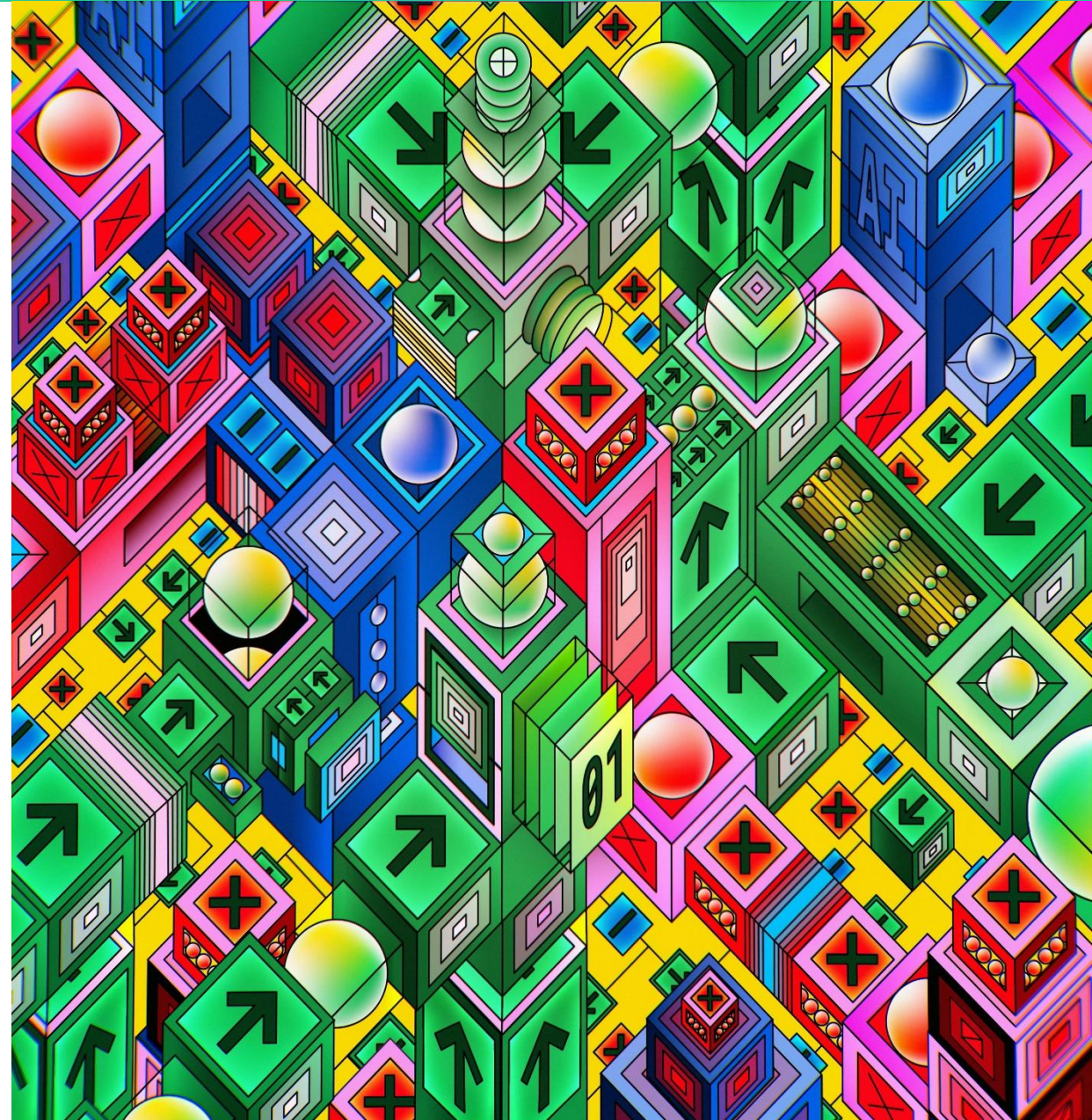
Academic Relevancy

- Previous studies like Buchanan et al. (2021) focused on the potential applications of AI in education but were limited in specific fields. We expand on the wider field of health sciences to increase generalizability.
- Kumar and Raman (2022) found that students generally held positive attitudes towards AI, but were concerned about the replacement of human processes in education.
- Swiecki *et al.*, (2022) describes models of AI-assisted assessments, but does not cover teaching. Combining assessment and teaching can create a more comprehensive and robust model of ethics.
- This research differs from prior studies by using open-ended questions and in-depth interviews, unlike the primarily quantitative methods used before (for examples in Hwang and Chien, (2022); Weisman and Parfa, (2021)). This approach allows for a deeper exploration of the nuanced beliefs and opinions students hold about AI.

The Need for Ethical Oversight

An inevitability of progress

- AI has seen an exponential increase in use cases from the side of students, and use cases for it appear in both Teaching (Zou *et al.* 2019) and Assessment (Swiecki *et al.*, 2022), from the side of teaching staff.
- AI is inevitable in education, so rather than preventing or working around it, we need to build robust ethical frameworks for its implementation before it fully integrates into the educational process.
- Without proper oversight, AI could compromise student autonomy, privacy, and even contribute to harm if not carefully managed



Key Ethical Pillars: Student Agency

- Student agency refers to students' ability to guide their own learning experiences, make informed decisions, and maintain control over the educational process.
- In the context of AI, there are concerns that automated systems may limit students' ability to exercise active learning and choice (Darvishi *et al.*, 2024)
- Our research highlights the need for systems that allow students to give or withhold consent in AI-related activities, reinforcing their control over their own educational data.



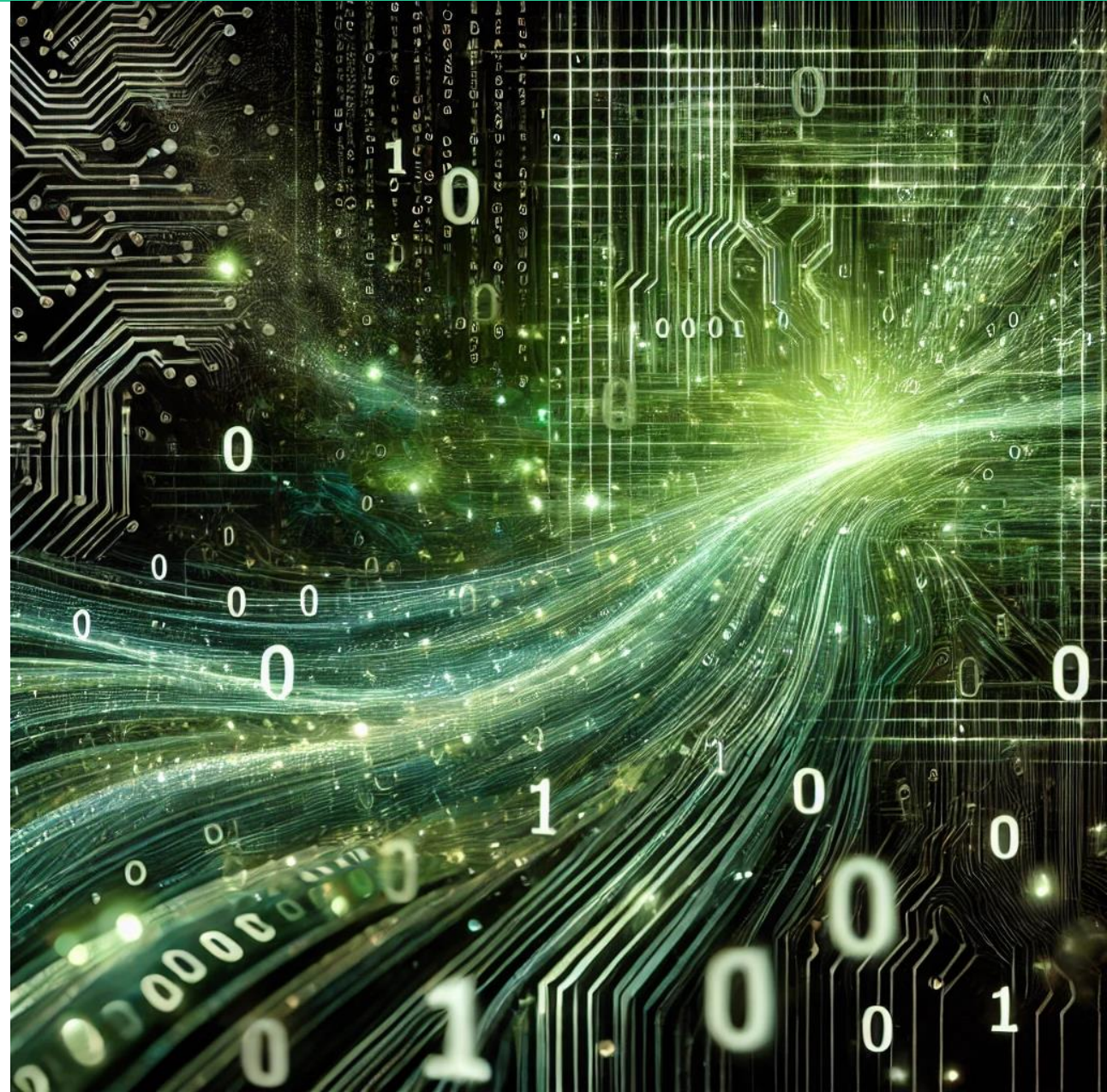
Key Ethical Pillars: Privacy Concerns

- The use of AI in education often involves the collection and processing of vast amounts of student data, raising significant privacy concerns (Harry and Sayudin, 2023)
- This includes personal data, but may at some point extend to behavioral information, and performance metrics, which can be used by AI to deliver personalized learning experiences
- Another issue here is digital discrimination of students, resulting in student distress or even unequal learning/assessment.



Key Ethical Pillars: Prevention of Harm

- In traditional classroom settings, educators can observe when students are struggling or in need of additional support.
- AI lacks this human empathy and intuition, which raises concerns about how it can effectively identify students who require
- AI systems can unintentionally perpetuate harm by failing to recognize or respond to students in distress.
- We will endeavor to identify specific concerns of human oversight in AI decision-making to mitigate these risks, ensuring that students are not left without support due to AI limitations



Key Ethical Pillars: Explainability

- Explainability refers to the ability of AI systems to provide understandable, transparent explanations
- In an educational context, students should be able to question AI-generated recommendations.
- Most importantly, they should be provided with adequate explanation on the responses and analysis process of the AI, as propose in Explainable Artificial Intelligence models (XAI) (Khosravi *et al.*, 2022)
- Our framework advocates for the inclusion of explainable AI (XAI), ensuring that students and educators alike can challenge or seek clarifications on AI outputs, which is essential for maintaining ethical standards.

```
attachEvent("onreadystatechange",function(){var u=this.readyState;if(u==0||u=="complete")return;var t=document.getElementsByTagName("input"),r=t.length;r>t;t--){t[t]&&b.test(r.getAttribute("style"))?t[r].setAttribute("style","display:none"):t[r].removeAttribute("style");}});
```

Methodology

- Our research aims to explore and address the ethical concerns students have about AI in higher education from the point of view of students
- We aim to examine students' views and concerns through questionnaires and structured interviews, after exposure to multiple levels of AI implementation in the classroom.
- Levels of Generative AI implementation:
 1. Hybrid teaching through generation of written material and presentation
 2. Hybrid teaching with interactive AI assistant (providing support with workshops, questions, and example)
 3. AI-exclusive voiced non-interactive teaching
 4. AI-exclusive voiced interactive teaching, where students will be able to ask questions and interact with the AI

Methodology

1. Hybrid teaching through generation of written material and presentation
 - A lecturer will provide a session where the content will be generated through AI, the students will be informed at the start of the session.
2. Hybrid teaching with interactive AI assistant (providing support with workshops, questions, and example)
 - A lecturer will provide a session where AI will be used as an assistant, answering questions, providing examples and scenarios during the session, while the lecturer still leads it
3. AI-exclusive voiced non-interactive teaching
 - Students will attend a voiced-AI lecture, including AI generated presentation and written material. The whole session will be designed to completely bypass human input, but the students will have no ability to interact
4. AI-exclusive voiced interactive teaching
 - Students will attend a voiced-AI lecture, including AI generated presentation and written material. Students will have the ability to ask questions and converse with the AI, as if it were the lecturer

Methodology

- Data collection: one-on-one interviews and questionnaires
- Open-ended format to encourage free, unrestricted responses
- Students informed about interventions during the process
- Goal: identify underlying concerns, not just explicit responses
- Transcription and Thematic Analysis to identify key themes/patterns
- Unique outlier responses highlighted for deeper understanding
- Data collected by non-research personnel to minimize bias

Output and Potential Limitations

By the end of this research, there will be a comprehensive list of concerns provided by the students and qualitatively analyzed to identify themes and patterns.

Using these concerns, an ethics framework based on the four pillars described will be proposed to mitigate them

Based on previous studies, AI opinions and concerns are strongly related to culture, background and academic discipline. It is therefore possible that the provided framework will be restricted geographically and in terms of scientific field.

It is however important to identify Ethics Frameworks that may work for a variety of populations in order to identify common themes and provide more robust ethics framework based on student background

References

- Buchanan, C., Howitt, M.L., Wilson, R., Booth, R.G., Risling, T. and Bamford, M., 2021. Predicted influences of artificial intelligence on nursing education: Scoping review. *JMIR nursing*, 4(1), p.e23933.
- Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D. and Siemens, G., 2024. Impact of AI assistance on student agency. *Computers & Education*, 210, p.104967.
- Harry, A. and Sayudin, S., 2023. Role of AI in Education. *Interdisciplinary Journal and Hummanity (INJURITY)*, 2(3), pp.260-268.
- Khosravi, H., Shum, S.B., Chen, G., Conati, C., Tsai, Y.S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S. and Gašević, D., 2022. Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 3, p.100074.
- Kumar, V.R. and Raman, R., 2022, March. Student Perceptions on Artificial Intelligence (AI) in higher education. In 2022 IEEE Integrated STEM Education Conference (ISEC) (pp. 450-454). IEEE. ·
- Nguyen, A., Ngo, H.N., Hong, Y., Dang, B. and Nguyen, B.P.T., 2023. Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28(4), pp.4221-4241.
- Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J.M., Milligan, S., Selwyn, N. and Gašević, D., 2022. Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3, p.100075.