AI in Education: Ethical and Epistemic Perspectives

6 – 7 March

Book of Abstracts

hosted by
Eindhoven University of Technology
Eindhoven Center for the Philosophy of AI

AI in Education: Ethical and Epistemic Perspectives
Book of Abstracts

6 – 7 March 2024
Eindhoven University of Technology and online
https://ephil.ai/event/workshop-ai-education/

Invited Speakers
Imre Bárd, Radboud University Nijmegen
Gunter Bombaerts, Eindhoven University of Technology
Karolina Doulougeri, Eindhoven University of Technology
Marko Galjak, Institute of Social Sciences, Belgrade
Silvia Milano, University of Exeter
Sandro Radovanović, University of Belgrade
Carlos Zednik, Eindhoven University of Technology

Program Committee
Myrthe Blösser, University of Amsterdam
Tim Fütterer, University of Tübingen
Mihály Héder, Budapest University of Technology and Economics
Theodore Lechterman, IE University
Diana Martin, University College London
Federica Russo, Utrecht University
Constantin Vică, University of Bucharest

Organizing Committee
Marina Budić, Institute of Social Sciences, Belgrade
Gavrilo Marčetić, Eindhoven University of Technology
Vlasta Sikimić (Chair), Eindhoven University of Technology
Aleksandra Vučković, University of Belgrade

Contents

Cognition Tech and Cognitive Skills: Rights, Obligations, and Human Dignity
Dina Babushkina 3

Embedding Ethics in EdAI Development – Early Impressions of Challenges and Opportunities
Imré Bard 4

ChatGPT in Education: Extended Cognition, Cognitive Artifacts and Cognitive Abilities
Guido Cassinadri 5

When Machines Judge: Exploring the Real Costs of AI Grading
Thomas Corbin & Gene Flenady 7

The Role of AI in Self-Regulated Learning in Engineering Students
Karolina Doulougeri 8

AI in Education: Visions, Challenges, and Strategies for Tomorrow
Marko Galjak 9

What Does Generative AI Actually Know?
Daniel L. Golden 10

"Must be tough living your life according to a couple of scraps of paper.": Using Memento to Teach Extended Epistemology to 15-18-year-olds and Cultivate Epistemic Virtues Required for the Responsible Use of Large Language Models
William Gopal & Michael Quinn 11

A Top-Down-Bottom-up Approach to Implement AI in Education
Marije Goudriaan, Anouschka Van Leeuwen, Ünal Aksu & Momena Yousufzai 12

What is the Place of Emotion AI in Moral Education?
Charlie Kurth 14

Leveraging Artificial Intelligence for Assessing Mental Health in Education: An Innovative Approach
Chitaranjan Mahapatra & Ashish kumar Pradhan 15

Dual-Process Theory and Artificial Intelligence: LLMs as Type-1 Processors
Joshua Mugg 16

Plagiarism or Extended Belief
Hadeel Naeem 17

Chat Bots, AI, and the Perils of the Homogenization of Education: Pluralism as an Epistemic Virtue
José Antonio Pérez-Escobar & Deniz Sarikaya 18

Simulating Equity: Modelling Affirmative Action in Education
Sandro Radovanović 19

Against Impersonal Evaluation: A Philosophical Discourse on AI for Student's Evaluation
Giovanni Russo 20

EquiLearn: A Pioneering Framework for Ethical AI in Education
Sahaj Vaidya & Stefan Bauschard 21

Ethical Implications of AI in K-12 Education: A Systematic Literature Review
Michał Wieczorek, Mohammad Hosseini & Bert Gordijn 22

Re-Aligning Higher Education in the Age of Generative AI
Carlos Zednik & Gunter Bombaerts 23

Cognition Tech and Cognitive Skills: Rights, Obligations, and Human Dignity

Dina Babushkina, University of Twente

Putting aside the excitement about and possible benefits of such LLM as ChatGPT for education (and, arguably, there are some), I will focus on the disruptive potential of ChatGPT epistemic agency, both of learners—whose cognitive skills are still under development—and the educator—as a formed epistemic agent practicing her cognitive skills (especially in high-order theorising, such as philosophy). This is not with the intent to demonize ChatGPT, but in order to find a way to use this technology for learning in a way that does not undermine cognitive agency and respectful epistemic interactions between human agents. I will approach ChatGPT an element of the industry/technology of cognition and explore it's implications for human epistemic agency and intersubjectivity.

I will outline conditions under which the use of ChatGPT in learning constitutes outsourcing cognitive operations and I will conceptualize faking epistemic states as a form of cheating. I will connect this with jeopardizing the development of the epistemic skills. I will look at the problem from two perspectives: dignity and epistemic rights. As a cognition technology, ChatGPT aims to turn into a product and substitute an array of epistemic skills, such as: the textual analysis, critical review, summarizing, retelling, interpreting, categorising, drawing conclusions. It comes with the implicit devaluation of "owning an ability" or "being able to X" by the human agent herself. The risks of such approach in learning is that the agent will not be able to perform—or not perform in a satisfactory manner—those actions on her own, independently of the technology. In this light, we are forced to ask the question about the ideal of human epistemic agency (What cognitive skills and abilities one ought to develop? What cognitive skills are desirable to preserve in a human agent?) and the epistemic dimensions of human worth (e.g.: Hindering which cognitive abilities and practices constitute undue treatment of a person?).

I will argue that despite the implicit message of cognition technology, there is value in the development and possessing epistemic skills by the agent herself. As a result, from the moral point of view, when outsourcing substitutes the development of own skills, we can talk about the violation of learner's epistemic rights (compare to Watson 2021), i.e. the right of a person to access and possess certain epistemic goods (such as skills related to knowing and communicating). Likewise, we can talk about the inhibition of the educators' epistemic obligations to provide such epistemic goods to the learners. Hindering the development of certain epistemic skills comes with high costs: undermining learner's cognitive autonomy, limiting and disrupting certain intersubjective interactions (which rely on the application of own epistemic skills), the proliferation of faking and deception.

Embedding Ethics in EdAI Development – Early Impressions of Challenges and Opportunities

Imré Bard, Radbound University Nijmegen

This presentation explores the Dutch National Education Lab AI (NOLAI), a research initiative at Radboud University's Faculty of Social Sciences. Funded by the Dutch Growth Fund, NOLAI collaborates with strategic partners to enhance the quality of primary and secondary education through intelligent technologies, while considering the pedagogical, societal, and ethical impacts of educational AI. We develop AI prototypes addressing educational needs and promote responsible AI use in education.

NOLAI's work is divided into two main programs: the co-creation program, which develops AI applications with schools, students, education scientists, and businesses; and the scientific program, which focuses on pedagogical-didactical knowledge, technical AI, data and infrastructure, teacher professionalization, and ethics.

The ethics team employs an embedded ethics approach, spanning the entire lifecycle of AI development. This includes problem definition, prototyping, implementation, assessment, and scaling. NOLAI's dual focus on scientific and practical goals involves studying design challenges and value conflicts, advancing embedded ethics methodology, and implementing responsible AI practices. NOLAI's unique construction and 10-year mandate represent a powerful opportunity to advance the field of educational technology by fostering a sustainable, ethical, and evidence-based integration of AI into educational practices, potentially transforming the landscape not just in the Netherlands, but perhaps as a model internationally.

I will provide an overview of the approach and processes which NOLAI's ethics focus area has sought to establish across the project's several work streams and will also address the challenges we have encountered over the first year, along with our attempts at responding to these. For example: managing multi-stakeholder and multi-disciplinary collaboration, anticipating ethical impacts in an extremely fast-moving environment, as well as avoiding the risk of techno-solutionism.

ChatGPT in Education: Extended Cognition, Cognitive Artifacts and Cognitive Abilities

Guido Cassinadri, Scuola Universitaria Superiore Sant'Anna, Pisa

Given the proliferation of technological tools such as ChatGPT for solving cognitive tasks, how should educational practices incorporate the use of such tools without undermining the cognitive abilities of students? Pritchard (2013, 2014, 2016) argues that it is possible to properly solve this 'technology-education tension' (TET) by combining his virtue epistemology framework with the theory of extended cognition (EXT). He argues that since EXT enables us to consider tools as constitutive parts of the cognitive system of students, in some cases the diminishment of brain-based cognitive abilities does not imply a diminishment of the cognitive character of students. The aim of this article is to offer a complementary and more encompassing framework of tool-use compared to the one presented by Pritchard to address the TET, applying it to the educational uses of ChatGPT.

To do so, in section 1.1, I present Pritchard's framework of cognitive character and virtue responsibilism applied in education. In section 1.2 I present and highlight the problems of using EXT as a solution to address the TET. First, in the literature there are no clear conditions for positing extended cognitive system and it is not practically clear when these conditions apply in real-world scenarios. Moreover, the conditions for cognitive extension would apply to a limited set of cases. Finally, Pritchard (2014) assumes a simplistic dichotomic view of cognitive scaffolding according to which without embracing EXT students' reliance on technology would imply a form of cognitive 'diminishment' (Pritchard 2016, p. 9).

In section 2.1, I combine Pritchard's framework of cognitive character with Fasoli's taxonomy of cognitive artifacts (2017; 2018). This taxonomy distinguishes 'substitutive cognitive artifacts', which enable the agent to exert a minimum degree of cognitive agency for completing a task by delegating most of the necessary work to the artifact; 'complementary cognitive artifacts', that jointly contribute to the agent's cognitive capacities to complete a cognitive task, and 'constitutive cognitive artifacts', which offer a necessary contribution to the completion of the cognitive task, that could not be completed by the agent without such a contribution. I offer a refinement of Fasoli's framework that can account for the complex, heterogenous, and multilevel ways in which cognitive artifacts are integrated into our cognitive abilities, enhancing and diminishing them. In section 2.2, I propose to combine Pritchard's virtue responsibilism with Fasoli's refined framework of cognitive artifacts to address the TET, using a context-sensitive, case by case approach. An educator committed to this approach may, in one case, accept that the use of digital tools may undermine, in the long run, some specific students' cognitive abilities, while in another case he may prefer to preserve other kinds of on-board cognitive abilities. I argue that my framework is more informative than Pritchard's and enables to evaluate the trade-offs between internally developed skills and externally delegated ones.

To conclude, in section 3.1, I present some epistemically virtuous uses of ChatGPT in educational contexts by considering it as a substitutive, complementary, or constitutive cognitive artifact (Fasoli 2017) involved within a cognitive ecology (Hucthins 2010).

References:

Fasoli, M. (2017). Substitutive, Complementary and Constitutive Cognitive Artifacts: Developing an Interaction-Centered Approach. *Review of Philosophy and Psychology*, 9, 671-687. https://doi.org/10.1007/s13164-017-0363-2

Fasoli, M. (2018). Super Artifacts: Personal Devices as Intrinsically Multifunctional, Metarepresentational Artifacts with a Highly Variable Structure. *Minds and Machines*, 28(3), 589-604.

Hutchins, E. (2010). Cognitive Ecology. *Topics in Cognitive Science*, 2, 705-715. https://doi.org/10.1111/j.1756-8765.2010.01089

Pritchard, D. (2013). Epistemic Virtue and the Epistemology of Education. *Journal of Philosophy of Education*, 47, 236-247. https://doi.org/10.1111/1467-9752.12022

Pritchard, D. (2014). Virtue Epistemology, Extended Cognition, and the Epistemology of Education. *Universitas: Monthly Review of Philosophy and Culture*, 478(2014), 47-66.

Pritchard, D. H. (2016). Intellectual Virtue, Extended Cognition, and the Epistemology of Education. *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*, Routledge, edited by J. Baehr, 113-127.

When Machines Judge: Exploring the Real Costs of AI Grading

Thomas Corbin, Macquarie University & *Gene Flenady*, Monash University

This paper engages in a critical analysis of a looming yet currently undiscussed shift in higher education: the adoption of generative AI (GenAI) for grading practices. Our analysis is structured into three interconnected sections, each examining a different facet of this emerging phenomenon.

Firstly, we explore the technological underpinnings and operational implications of AI grading systems. This section delves into the capabilities and limitations of GenAI in assessing student responses, addressing both the technological sophistication and the challenges inherent in automating such nuanced tasks. We discuss how AI grading can range from simple objective assessments to complex evaluative tasks, underscoring the implications for educational assessment and teaching practices.

The second section shifts focus to the economic and institutional motivations driving the adoption of AI technology in higher education. We analyze the "productivity promise" of AI, highlighting the efficiency gains in grading processes and the consequential impacts on educators and university administrations. This discussion extends to the broader economic landscape of higher education. We assess how the increasingly corporatised university affects budget allocations, staffing decisions, and the prioritization of research over teaching responsibilities. Ultimately, we aim to demonstrate what we feel to be a simple and conservative claim, that the integration of AI in grading processes within higher education presents a range of economic benefits that will become increasingly hard for institutions to ignore or resist.

Finally, we provide a critical examination of the broader implications of AI in grading, particularly from ethical, professional, and pedagogical perspectives. We aim to pre-empt discussions on the economic incentives to adopt AI-assisted marking by presenting here an overview of several of the most compelling potential counterarguments. For example, we explore the potential erosion of the human element in teaching and assessment, discussing the moral and ethical dilemmas posed by AI grading. We also cover the impact on the academic career pipeline and public perception, delving into how AI integration in grading could reshape the reputation and credibility of educational institutions. Finally, but also most importantly, we look to the negative pedagogical implications of AI-assisted grading.

Through the above analysis, the paper serves as a pre-emptive intervention in a crucial yet currently unpublicized dialogue about the impact of GenAI on higher education. We aim to broaden the discourse around AI in education, emphasizing the need to consider not just operational efficiency, but also the nuanced and long-lasting implications of GenAI in grading across technological, economic, and ethical dimensions.

The Role of AI in Self-Regulated Learning in Engineering Students

Karolina Doulougeri, Eindhoven University of Technology

In this contribution, we focus on self-regulated learning (SRL) as one key principle of innovative engineering curricula. We will discuss important considerations when using AI, specifically ChatGPT, in educational practice and how this can influence students' SRL.

In engineering, the ability of students to manage their learning process is crucial. SRL empowers engineering students to take control of their learning, fostering skills such as goal setting, time management, and reflection. As engineering problems become increasingly complex, a way to prepare students is by using real-world and open-ended challenges that require understanding of complex concepts and applying theoretical knowledge as driver for student learning. In such context, the importance of self-regulated learning becomes evident in equipping students with the tools needed to navigate open-ended challenges and develop innovative solutions.

Artificial Intelligence (AI) integration in engineering education is receiving significant attention for its potential to enhance personalized learning experiences, teaching methodologies, and student engagement. AI applications, ranging from student grading to intelligent tutoring, offer many possibilities to influence student learning. One example of AI used in education is the integration of chatbots like ChatGPT into learning environments. From the students' perspective, ChatGPT can be valuable resource that supports and guides their learning activities, helping them understand complex theoretical concepts, define their objectives and tasks, and develop plans when working with open-ended challenges. However, existing examples of integration of AI in higher education reveal a need for educational frameworks to guide educators in effectively integrating these tools into teaching practice.

This contribution will review existing literature on the interplay of AI and self-regulated learning in higher education. Understanding how individual factors shape the relationship between the use of AI tools and students' self-regulated learning is important. We will discuss student characteristics, such as student agency, motivation, and perceptions about learning activities and tasks and how they can influence students' interaction with AI and the development of SRL. By understanding how individual factors interact with contextual aspects of the learning environment, educators can tailor the use of AI to meet the learning needs of engineering students, fostering self-regulation. Building upon the insights from the state-of-the-art literature, we aim to contribute towards a comprehensive pedagogical framework where AI can be used to foster a student-centered, motivating, and active learning environment.

AI in Education: Visions, Challenges, and Strategies for Tomorrow

Marko Galjak, Institute of Social Sciences, Belgrade

The incorporation of Artificial Intelligence (AI) in educational systems signals a transformative phase in teaching and learning methods. In the presentation we examine the interplay between AI developments and educational practices, envisioning how this synergy might unfold in three future contexts. In one scenario, where AI capabilities expand gradually, we foresee AI augmenting education, as seen with examples like Khanmigo – Khan Academy's virtual tutor. Recent advancements in Large Language Models highlight AI's potential as an adaptive tutor, capable of tailoring educational content to individual learners' preferences and needs. This opens possibilities for personalized learning, potentially revitalizing traditional teaching methods and aligning education more closely with the needs of a workforce increasingly shaped by AI technologies. However, this prospect raises ethical concerns, particularly regarding the use of AI in early education, where its impact on young minds and cognitive development must be carefully managed. Alternatively, a rapid leap in AI capabilities could fundamentally challenge our current educational structures, with AI excelling in both curriculum customization and teaching, potentially outperforming human educators. Here, we explore how educational systems might adapt to remain relevant and effective in the face of such swift AI advancements while also considering the potential psychological and societal risks involved. The third scenario explores the concept of technological singularity, envisioned as a moment where technological advancement accelerates uncontrollably, leading to profound and unpredictable impacts on humanity. Identifying common threads in all these scenarios is essential to promote a deliberate and mindful integration of AI in education, ensuring that technological progress is in harmony with human ethics and values. The integration of AI in education is inevitable, but its direction is not preordained. Our presentation aims to examine possible directions for AI in reshaping education. We advocate for extensive experimentation with AI applications within educational settings, underscored by innovation and a firm commitment to student welfare and ethical considerations. This includes empirical studies involving pilot programs and longitudinal research, with educators, technology experts, and policymakers engaging in proactive, long-term, and inventive discussions. Our objective must be to ensure that the future of education is not only technologically proficient but also ethically responsible and centered on human development and values.

What Does Generative AI Actually Know?

Daniel L. Golden, HUN-REN Research Centre for the Humanities, Institute of Philosophy

Keywords: epistemology; knowledge acquisition; knowledge production; reliance; responsible use

Recent developments in the field of artificial intelligence (AI) show that we are on the threshold of a new kind of coexistence of human beings and autonomous machines which will certainly bring along profound changes in society, economy and culture. The exceptional capacities of Large Language Models (LLM) and applications based on them becoming accessible to everyone present extraordinary challenges in various fields.

In the context of education, the growing usage of generative tools creates deep concerns regarding reliance and responsible use. It seems that our concepts and practices about learning and knowing should be carefully reexamined and possibly transformed. In order to be able to design those procedures properly, first of all we have to pose some fundamentally epistemological questions. What should be counted as knowledge in the case of a generative AI? What can we say about the groundings of that kind of knowledge? What role the human-machine dialogue has in the formation of this knowledge? What types of transfer processes are going on during the use of generative AI? What impacts the immersion into these epistemological frameworks will have on the functioning of our cognitive systems, will their abilities and capacities change significantly?

The special perspectives of educational settings may help to redirect philosophical discourses around AI from the traditional simulation paradigm to a new approach focusing on the interactions between human and machine intelligences. In my presentation I shall give an epistemological analysis of what kind of knowledge LLM systems do actually provide. As new technological developments naturally provoke classical descriptions of AI presented by Turing (1950), Minsky (1974), Dennett (1978), Searle (1980), Harré (1990) and Brooks (1991), they can mutually inform each other for conceptual revisions. Moreover, confrontations of real life user cases in knowledge acquisition and knowledge production with epistemological frameworks can shed new light even on some of the age-old issues of educational practitioners with individual accountability and performance assessment.

References

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47(1-3), 139-159.

Dennett, D. C. (1978). Artificial intelligence as philosophy and as psychology. *Brainstorms. Philosophical essays on mind and psychology*, Bradford Books, edited by D. C. Dennett, 109-126.

Harré, R. (1990). Vigotsky and artificial intelligence: What could cognitive psychology possibly be about? *Midwest Studies in Philosophy*, 15(2), 389-399.

Minsky, M. (1974). A Framework for Representing Knowledge, *MIT-AI Laboratory Memo*, 306, June.

Searle, J. S. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences*, 3(3), 417- 424.

Turing, A. (1950). Computing Machinery and Intelligence. *Mind* 59(236), 433-460.

“Must be tough living your life according to a couple of scraps of paper.”: Using Memento to Teach Extended Epistemology to 15-18-year-olds and Cultivate Epistemic Virtues Required for the Responsible Use of Large Language Models

William Gopal & *Michael Quinn*, University of Glasgow

Pedagogy in computing is focused on equipping students with the technical skills required to use ICTs, viz., framing problems computationally and building programs to solve them. Given recent developments in AI, specifically LLMs, attention has shifted toward responsible use. We propose a pedagogical shift to equip students with the necessary interdisciplinary skills to use LLMs responsibly if they choose to do so.

We take an aim of education to be the cultivation of epistemic virtue, which, as a pedagogical approach, has been implemented at the Intellectual Virtues Academy (Curren, 2019). From this perspective, the worry that students’ long-term use and possible reliance on LLMs leads to de-skilling (e.g., essay-writing skills) is understood as a novel virtue-epistemic risk to students. Such worries are amplified, given the possibility of cognitive extension (Carter, 2018). Furthermore, whilst education is considered a remedy to the epistemic risks of cognitively extending with digital technologies (Heersmink, 2016), there is little focus on how best to teach students to use cognitive extensions virtuously. We propose a method of teaching 15-18-year-olds to identify (a) what epistemic virtues they should develop for responsibly using LLMs-as-cognitive-extension and (b) what types of intellectual actions they should perform to develop these virtues and ‘epistemic phronesis’ (Johnson, 2021).

First, we provide a justification for using Memento as a teaching tool, drawing on notions of moral and epistemic exemplars. Next, we elaborate on how, within the teaching unit, Memento can be used to illustrate the extended mind thesis and (extended) epistemic virtue, suggesting that Leonard’s Polaroids and tattoos are a more accessible update to the Otto and Inga thought experiment. That is, the study of Memento is used to pedagogically scaffold students’ (i) development of philosophical analysis and (ii) their virtue-epistemic reflections on using cognitive extensions.

Second, we turn to the latter half of the teaching unit, where we introduce examples of people using ChatGPT's new customizability feature, allowing it to function as a “tutor”, “writing-coach”, or “proof-reader”. Students evaluate whether the examples constitute cases of (virtuous) cognitive extension or abstaining from use would be preferable. Third, we propose that students are assessed on the quality of their reflection (through a learning journal) regarding what virtues they need to responsibly use LLMs-as-cognitive-extensions and what actions they should (not) take to develop them.

Finally, we explore the upshots of our approach. One positive learning outcome is students gaining the knowledge to identify that relying on LLMs as cognitive extensions could inculcate epistemic vice. This aids the identification and critique of techno-optimistic rhetoric used to justify the widespread use of AI which students, and their teachers, will face. From this, we present tools for practitioners to reflect on how, if left unchallenged, this rhetoric (e.g., AI as “democratising” education) tactically fails to acknowledge AI’s technical deficiencies nor anticipates unexpected outcomes which could exacerbate the attainment gap. Likewise, we support educators in addressing educational governance bodies' decision to ban the use of LLMs, guided by techno-pessimism, which pre-emptively disregards the possible benefits of virtuous use.

A Top-Down-Bottom-up Approach to Implement AI in Education

Marije Goudriaan, Anouschka Van Leeuwen, Ünal Aksu & Momena Yousufzai,
Utrecht University

In the last decade, there has been an increase in research and societal use of artificial intelligence (AI), including in the educational sector (Sheikh et al., 2023). When applications of AI in education (AIED) are used to 'understand and optimize learning and the environments in which it occurs by using data about learners and their contexts', it falls under learning analytics (LA) (Siemens & Gasevic, 2012).

Using data derived from educational technology for further analyses, such as LA and AIED, gives rise to ethical challenges. Key principles of data ethics are that those involved are treated respectfully, fairly, equally, and are not harmed (Dama International, 2017; Holmes et al., 2022). Part of these challenges are covered in the general data protection regulation (GDPR)(European Union, 2016) and the AI act (European Union, 2021). It is the responsibility of the educational institutions who are employing LA or AIED to ensure that the data that is entrusted to them is handled ethically and according to these regulations (Dama International, 2017).

At our Dutch university we tackled these challenges by combining a top-down and bottom-up approach in a roadmap (Perez-Sanagustin et al., 2022). From the top-down, a policy for LA and AIED initiatives was created, based on a national LA and AI reference framework, literature (Tsai & Gasevic, 2017), and conversations with stakeholders. Besides the pedagogical goals and the legal framework, five ethical values are described:

i. Stakeholders are informed about the processing of their data.
ii. University is responsible and accountable for processing data.
iii. Interests of all stakeholders will be considered.
iv. University ensures that those who are handling the data are able to conduct, understand, and improve analyses.
v. Humans are always in the loop.

From the bottom-up, projects are initiated at the staff level within faculties, thereby ensuring stakeholder input. The roadmap has five steps that all LA and AIED initiatives go through. This helps to ensure that each project complies with the pedagogical goals, legal framework, and ethical values defined in the LA policy:

1. Orientation phase: determine if the project is in line with the LA policy and is feasible.
2. Educational check: assess if the project has a clear and relevant pedagogical goal and if the proposed evaluation is appropriate.
3. Privacy and security: check if the project adheres to the GDPR (and AI-act) and make sure the used data is protected accordingly.
4. Start of the project: prepare a privacy statement for those whose data will be processed and start with the technical realization of the project.
5. Evaluation: assess if the project reached its intended goal.

So far, our experience with the roadmap has been promising. It provides clarity and transparency regarding policy and responsibilities for each LA and AIED project. During the workshop, we will further elaborate on the development of our roadmap and our initial findings related to user experiences.

References

Dama International. (2017). *DAMA-DMBOK Data management body of knowledge* (2nd ed.). Technics Publications LLC.

European Union. (2016). *Regulation 2016/679—General Data Protection Regulation*. https://gdpr-info.eu/

European Union. (2021). *EU AI Act. Laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, 32(3), 504-526. https://doi.org/10.1007/s40593-021-00239-1

Perez-Sanagustin, M., Hilliger, I., Maldonado-Mahauad, J., & Perez-Alvarez, R. (2022). Building Institutional Capacity for Learning Analytics: Top-Down & Bottom-Up Initiatives*. IEEE Revista Iberoamericana de Tecnologias Del Aprendizaje*, 17(3), 281-289. https://doi.org/10.1109/RITA.2022.3191413

Sheikh, H., Prins, C., & Schrijvers, E. (2023). *Mission AI: The New System Technology*. Springer International Publishing. https://doi.org/10.1007/978-3-031-21448-6

Siemens, G., & Gasevic, D. (2012). *Guest Editorial—Learning and Knowledge Analytics*. 15(3), 1-2.

Tsai, Y.-S., & Gasevic, D. (2017). Learning analytics in higher education — challenges and policies: A review of eight learning analytics policies. P*roceedings of the Seventh International Learning Analytics & Knowledge Conference*, 233-242. https://doi.org/10.1145/3027385.3027400

What is the Place of Emotion AI in Moral Education?

Charlie Kurth, Helsinki Collegium for Advanced Studies

Cultivating one's emotions—learning to feel anger, say, at the right time and in the right way—has long been viewed as central to moral education. Recently, educators, philosophers, and entrepreneurs have pointed to "emotion nudges" and other forms of emotion-focused AI (EAI) as powerful, but under-utilized tools for emotion cultivation (Valor 2016, Suh 2016). And the initial results are intriguing: merely placing "watching-eye" icons in online chatrooms can prompt feelings of anxiety that help curb vicious posting (Park 2022), and virtual reality simulations can engage stereotype-challenging empathy (Bedrik 2017). But while there's a growing body of research examining ethical issues surrounding the use of AI for education and emotions in general, there is little that looks specifically at the ethics of using EAI for moral education. My talk aims to address this need by focusing on three issues.

(1) Bad answers to basic questions. Designing effective EAI tools requires the ability to both identify what emotion a person is feeling and assess whether that emotion is being experienced in an appropriate manner. However, we know that people are very concerned to keep their emotional lives private (Roemmich et al 2023, Khameneh 2023), and this suggests that they will work hard to conceal what they are feeling, thus undermining EAI-based tools. Moreover, even if these issues can be addressed, effective EAI also requires an understanding of what emotion cultivation interventions actually work. Here too there is trouble. Given what we know about how EAI algorithms operate, it appears that "successful" interventions are likely to come with high costs (e.g., homogenizing the emotional diversity we value; favoring the emotional experiences of neuro-typical individuals).

(2) Deskilling of crucial meta-emotional capacities. A central—if not essential—aspect of what it is to be a morally mature person is the ability to recognize and resist problematic feelings and impulses. This is why we grow concerned when someone fails to reassess (and let go of) misplaced anger or excessive pride. But our ability to develop these meta-emotional capacities is threatened by EAI, for these technologies are designed work precisely by giving technology a central role in assessing and reshaping an individual's emotions. Crucially, unlike the deskilling of our ability to remember phone numbers that has come with the use of mobiles, what's at risk here is at the foundation of what it is to be human.

(3) Problematic incentives. One might hope that frank and open discussion will allow us to work through the issues that underlie (1)-(2). But such hopes are challenged by powerful financial and political pressures that threaten to distort these conversations. We see evidence of this in both software companies' interest in developing "moral" EAI so that it can be sold to corporate training consultancies (Bailenson 2020) and politicians who are weaponizing talk of social emotional learning for electoral gains (Gross 2022).

While the overall tenor of my discussion is negative, I conclude by highlighting two places were EAI could have an important role to play in moral education.

Leveraging Artificial Intelligence for Assessing Mental Health in Education: An Innovative Approach

Chitaranjan Mahapatra, University of Paris-Saclay, French National Centre for Scientific Research (CNRS) & *Ashish kumar Pradhan*, Indian Institute of Sciences, Bangalore

Our comprehensive review article examines the fundamental change in thinking that occurs when Artificial Intelligence (AI) is incorporated into the evaluation of mental health in educational settings. The research explores the possibility of using AI to transform conventional assessment methods in response to the increasing prevalence of mental health issues among students. Through the analysis of many datasets, including academic records, digital communication patterns, and behavioral data, AI can offer a detailed and comprehensive insight into the well-being of students.

The background part emphasizes the constraints of traditional methods, which necessitates the development of novel solutions. The ability of AI to analyze large datasets is recognized as a significant benefit, since it enables the detection of subtle patterns and anomalies that may not be detected using conventional examinations. Natural Language Processing (NLP) is presented as a transformative element that allows for a more thorough examination of written and spoken language in order to uncover insights into the emotional states of students.

The article highlights the potential of AI in mental health screening for early intervention. AI can recognize early signs of mental health issues, allowing for prompt support and action to prevent the worsening of these concerns. The versatility of AI is demonstrated by its capacity to offer tailored assistance, addressing the specific requirements of each student and promoting a more comprehensive educational setting.

Furthermore, the paper discusses the necessity of enhancing accessibility in mental health assistance. AI-driven evaluations provide a scalable solution in major educational institutions, where traditional approaches may require a lot of resources. This allows for a wider reach among the student population.

The paper addresses problems and ethical dilemmas related to privacy concerns, biases in AI models, and the importance of human oversight. It is recommended to implement strong security measures and ethical rules to safeguard student privacy. Additionally, it is emphasized that fair and equitable outcomes should be prioritized to prevent the continuation of current inequities. The essay emphasizes the need for a harmonious incorporation of AI and human supervision to accurately understand intricate emotional states, hence guaranteeing empathic evaluations.

Ultimately, the article foresees a future in which AI, under the guidance of responsible methods, revolutionizes mental health assistance in the field of education. The potential advantages of AI, such as timely intervention, tailored assistance, and enhanced accessibility, are emphasized as crucial in fostering more supportive and inclusive educational settings. The essay emphasizes the importance of a collaborative approach in educational institutions as they incorporate AI technology. This approach should prioritize the well-being of students and ethical issues. By doing so, it will enable a comprehensive and transformative influence on mental health assessment in educational settings.

Dual-Process Theory and Artificial Intelligence: LLMs as Type-1 Processors

Joshua Mugg, Park University

This paper's central move is to bring Dual-Process Theory (DPT) to bear on discussion of LLMs. Doing so, we suggest, is helpful because it situates discussions on the epistemic and moral value (and perils) of LLMs within existing literature on the epistemic and moral discussion of human cognition. In the first section of this paper, we suggest thinking of LLMs as engaged in type-1 processing, where type-1 processing is thought to be associative, fast, and automatic. While DPT has been criticized (Gigerenzer 2010, Kruglanski 2013, Keren and Schul 2009), nothing we say in this paper requires thinking that DPT to be a good model of human cognition. Indeed, most defenders of DPT today do not think of type-1 processing as associative (e.g. Evans and Stanovich 2013, De Neys 2023).

There are a number of interesting similarities to normative upshots if one accepts this suggestion. Type-1 as defined here has been characterized as generally overconfident. For example, Kahneman (2011: 84) suggests that lack of confidence in one's answer is evidence that this answer arises from the more reliable type-2 processing. This normative point is even more pronounced in one of the most interesting and notable applications of DPT in the moral psychology literature. For example, Jonathan Haidt (2001, 2012) famously argued that system 1 generates the vast majority of our moral judgments, and the primary role of system 2 is to produce post-hoc rationalizations for those judgments, just as we saw in Kahneman's cognitive architecture. The analogy he uses is an elephant with a human rider on its back: the elephant (system 1) decides where to go; the human rider (system 2) merely comes up with reasons for why the direction chosen by the elephant is, in fact, the right direction.

Suppose that LLMs are Type-1 processors. What are the educational upshots? We suggest that higher education can only realistically be leveled at Type-2 processing, which we define here as rule-based, slow, and effortful. Insofar as education can impact Type-1 processing, it does so indirectly through its making novel Type-2 processing automatic over time through habituation. Therefore, we suggest that education should largely treat outputs of LLMs in the same way that we would treat Type-1 processing if one accepted this version of DPT. Additionally, insofar as type-1 processing is reliable, it is reliable only in an ecological way because it was trained through millions of years of evolution. Greene (2014) argues that an upshot is that we cannot trust our type-1 processing in novel circumstances. While Dale (2020) criticizes this account on evolutionary grounds, we argue Greene's upshot holds for LLMs, leading to circumspect helpfulness of LLMs.

Plagiarism or Extended Belief

Hadeel Naeem, Käte Hamburger Kolleg: RWTH Aachen

My research examines how we interact with AI systems to form beliefs. I am interested in a specific kind of interaction, one that is characterized by a reciprocal and continuous exchange of information between an agent and an AI system. Such back-and-forth of information with an AI system may generate beliefs that are partially realized in the agent's brain and partially in the AI system.

In this context, it is interesting to understand how students utilize large language models (LLMs) to do their assignments. On the one hand, it is clear if a student asks a GPT model to write an essay on Indian Philosophy and then submits this essay as her homework that she has plagiarized and passed some work as hers when it clearly isn't. On the other hand, consider a student who picks a specific debate in Indian Philosophy, asks a GPT model to recommend the newest and most controversial literature on the debate, then proposes several thesis statements and asks the GPT model to evaluate them, and then asks the model to suggest a few ways to structure an essay on the particular thesis statement. Suppose a few more steps of back and forth between the student and the GPT model. It is now unclear if this qualifies as plagiarism. In the context of the literature on extended cognition and extended knowledge, one can argue that the essay represents the student's own work.

The extended knowledge thesis (Pritchard 2010; Palermos 2014; Carter et al. 2018) (which stems from the extended cognition theory (Clark and Chalmers 1998)) is the idea that we sometimes employ external resources such that our epistemic process occurs at least partially outwith the boundaries of our body. Such an extended epistemic process can produce a belief that is also extended, which may be extended knowledge. Whether a process extends will depend on how an agent employs a process. Extension typically requires that the agent approach the process seamlessly and fluently, just as we employ our internal faculties.

When the external resource is an AI system that does far more of the information processing that realizes the belief, it's not clear why such an extended belief should be attributed to us. For example, it is straightforward that a thought you wrote in a notebook (that extends your epistemic process) and later retrieved is your thought, but it's unclear how a GPT model can generate information that may be your belief.

In my paper, I examine the extended knowledge literature to determine how to distinguish between a student's original work and plagiarized work produced with an LLM. I explore the conditions that must be met for the student to be credited with the LLM-generated belief and identify some potential issues with this approach.

References

Carter, J. A., Clark, A., Kallestrup, J., Palermos, S. O., & Pritchard, D. (Eds.). (2018). *Extended epistemology*. Oxford University Press. https://doi.org/10.1093/oso/9780198769811.001.0001

Clark, A., & Chalmers, D. (1998). The extended mind. *analysis*, 58(1), 7-19. http://www.jstor.org/stable/3328150

Palermos, S. O. (2014). Knowledge and cognitive integration. *Synthese*, 191(8), 1931-1951. https://doi.org/10.1007/s11229-013-0383-0

Pritchard, D. (2010). Cognitive ability and the extended cognition thesis. *Synthese*, 175(Suppl 1), 133-151. https://doi.org/10.1007/s11229-010-9738-y

Chat Bots, AI, and the Perils of the Homogenization of Education: Pluralism as an Epistemic Virtue

José Antonio Pérez-Escobar, University of Geneva & *Deniz Sarikaya*, Vrije Universiteit Brussel

In today's technological landscape, the integration of advanced AI-powered chat bots into education is reshaping the way knowledge is imparted and acquired. However, this talk seeks to illuminate a potential peril that lurks within these adaptations. While the focus is primarily on mathematics, will also briefly consider other areas like biology. This discussion centers around the potential hazards of, among others, standardized video lectures, particularly popular after the Covid-19 pandemic, and their possible role in undermining the principle of pluralism within academia. The central argument posits that these technological advancements could potentially have adverse effects on both the practice of mathematics and the broader societal landscape.

We argue that the mathematical undergraduate curriculum should not succumb to excessive codification and formalization through the overuse or misuse of online teaching tools. The talk contends that while technology undoubtedly has its place, the rigid structure of such tools risks harming the diversity of thought and pedagogical approaches that has traditionally characterized the academic landscape.

In order to underscore the importance of maintaining a diverse array of research practices for fostering mathematical progress, we will present a historical discussion on the role of the Tripos in creating a specific mathematical culture in pre-globalization Cambridge, and the important results that would not have been obtained otherwise. This historical context serves as an example to assess the potential impact of current technological trends on the future of mathematics education and research.

After the historical overview, we return to modern times and draw parallelisms between the standardization of education and the concept from economics “winner-takes-it-all”: it refers to an economic system where the best performers in a competitive environment outrun the competition and monopolize the rest of the market. This is a term that was studied especially in the context of modern internet-based business models, thus serving our analysis well.

Last, we discuss philosophical research stressing how mathematical pluralism is important, from different points of view. This includes not just pure mathematics but also the mathematization of other areas: for instance, there is a variety of alternatives to mathematize biology, each with its own virtues. In doing this, we can better assess what is at stake with an improper homogenization of education, and what the most vulnerable and resilient areas are. This characterization offers us hints to balance technological innovation and the preservation of intellectual richness. One key notion here is productive ambiguity: a certain type of ambiguity is desirable from an epistemic perspective, but the digitalization and homogenization of mathematical notions may undermine it.

Simulating Equity: Modelling Affirmative Action in Education

Sandro Radovanović, University of Belgrade

This talk focuses on evaluating the effectiveness and efficacy of affirmative action policies in promoting diversity, equity, and inclusion within educational systems, particularly for groups that have historically been underserved or marginalized. Affirmative action, as a subset of social inclusion policies, addresses issues arising from historical, social, and economic exclusion, whether direct or indirect. However, these policies have also sparked debates over reverse discrimination, stigmatization, challenges to meritocracy, and their long-term effectiveness. The core of this talk is to review and categorize affirmative action policies from around the world, grouping them based on their similarities. Using agent-based simulations, the talk presents the simulated impacts of these policies. Specifically, it presents the outcomes of simulations applied to three distinct affirmative action strategies that have been historically implemented: 1) quota-based policies - policies that set specific quotas for marginalized groups, 2) institutional affirmative actions - actions taken by educational institutions or government to foster inclusivity, and 3) indirect affirmative action policies – seemingly neutral, but "purposefully inclusionary" measures. Hopefully, the talk will provide a different viewpoint (from computer science stand) on understanding of the effectiveness of different affirmative action policies in education.

Against Impersonal Evaluation: A Philosophical Discourse on AI for Student's Evaluation

Giovanni Russo, University of Bologna

This proposal problematizes the desirability of using AI to evaluate student performance. Acknowledging its potential advantages, the analysis delves into the inherent limitations, specifically focusing on how AI obscures the subjectivity of educators – an indispensable element for fostering student identity. Drawing on Axel Honneth's philosophy of recognition as a normative framework, the discussion wants to underscore the significance of authentic intersubjective encounters within educational settings, wherein the evaluation process assumes a pivotal role.

In brief, the utilization of AI to evaluate student performance is commonly perceived as a desirable goal, promising a fair, objective, and efficient judgment, due to its impersonal nature. In this framework, the impersonality of evaluative judgment reflects a specific set of desiderata, with educator subjectivity identified as the crux of the issue. Consequently, when the use of AI in the evaluation process is criticized, the focus is not on its end – pursuing an impersonal judgment – but on its means, for instances its privacy, transparency, and explicability problems. In contrasts, this proposal aims to question not the means but the end itself, i.e., the impersonality of evaluative judgment.

As a methodological starting point, this discussion distinguishes a normative standpoint from an empirical one. Positioned within a normative framework, the empirical perspective can assume a critical direction. Following Honneth, a genuine intersubjective encounter takes place between a subject and an alterity that, through a dialectical recognition process, can disclose a personal identity. In this struggling process for self-affirmation, the subject seeks recognition from otherness. The student-educator relationship embodies these normative aspirations, culminating in the evaluative moment as the natural conclusion of this recognition process.

Considering the normative direction set by recognition theory, it is important to analyze, from an empirical standpoint, the practical application of AI in the context of student evaluation. Within this sphere, the educator's role is hindered, giving rise solely to the impersonality of AI judgment. AI's impersonal judgment undermines the educator's subjectivity. There is not intersubjective relationship, crucial for student development. In essence, the impersonal evaluative relationship contradicts the normative framework, negating the condition of possibility of identity formation, especially as the grading moment is construed as the apex of the recognition process.

In conclusion, this proposal argues that the implementation of AI should strive to fortify the subjectivity of educators rather than supplanting it. In other terms, AI ought to function as a supportive element for subjectivity, avoiding an impersonal substitution.

EquiLearn: A Pioneering Framework for Ethical AI in Education

Sahaj Vaidya & *Stefan Bauschard*, New Jersey Institute of Technology

In the dynamic landscape of AI integration in education, a pressing concern emerges—the pervasive bias within AI systems that compounds educational disparities among students. This challenge is exacerbated by the opacity of AI decision-making processes, inadvertently favoring certain demographic groups. Prevailing solutions often focus on corrective measures post hoc, neglecting a holistic approach for proactive prevention. Our groundbreaking solution introduces a comprehensive algorithmic fairness framework meticulously crafted for educational environments. Emphasizing transparency, accountability, and stakeholder involvement in AI system design, our approach goes beyond rectifying bias; it proactively prevents its emergence. This forward-thinking strategy sets our solution apart, ensuring fair and equitable AI-driven education for all. As part of a larger initiative, our algorithmic fairness solution contributes to the development of an open-source, public AI in Education Model, employing a Mix of Experts approach to support academic content instruction and advance best practices in education.

Ethical Implications of AI in K-12 Education: A Systematic Literature Review

Michał Wieczorek, Dublin City University, *Mohammad Hosseini*, Northwestern University & *Bert Gordijn*, Dublin City University

Background: This paper provides a systematic review of the literature discussing the ethics of using artificial intelligence in primary and secondary education (AIED). Although recent advances in AI have led to increased interest in its use in education, discussions about the ethical impacts of this new development are dispersed. Moreover, while computer scientists and education scholars have engaged in interesting and promising work, they consider their discussions as exploratory and cite lack of ethical expertise as a significant limitation.[1,2] At the same time, policymakers and international organizations mention that the lack of systematic studies on the ethics of AIED makes it difficult to address new developments at the policy level.[3] Our literature review responds to some of these challenges by consolidating discussions that occurred in different epistemic communities interested in AIED and by offering an in-depth ethical analysis of the debate.

Methods: We explicitly focus on the use of AI in primary and secondary education since this level of education is compulsory in most countries. The review is conducted using the PRISMA-ETHICS guidelines and has resulted in the inclusion of 48 manuscripts published between 2016 and 2023.

Results: Using a thematic approach, we subsumed ethical issues under twenty categories, with five outlining potential positive developments and fifteen dealing with perceived negative consequences. Retrieved works cover, for example, the impact of AI on students' skills, knowledge and wellbeing; the changes to teachers' jobs and educational practices; concerns related to privacy, bias and security; analyses of power relations and the influence of private companies; as well as inherent limitations of AI in an educational setting.

Discussion: We argue that in-depth engagement with ethical theory and philosophy of education is needed to adequately address certain challenges brought by AIED. We also encourage researchers to devote more attention to the ethics of AIED, because the published literature disproportionately focuses on higher education.

References

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., ... & Koedinger, K. R. (2021). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 1-23.

Porayska-Pomsta, K., Holmes, W., & Nemorin, S. (2023). The ethics of AI in education. In *Handbook of Artificial Intelligence in Education* (pp. 571-604). Edward Elgar Publishing.

UNESCO. (2019). *Beijing consensus on artificial intelligence and education*. Unesco Paris.

Re-Aligning Higher Education in the Age of Generative AI

Carlos Zednik & *Gunter Bombaerts*, Eindhoven University of Technology

Generative AI threatens the alignment between learning objectives, teaching activities, and assessment methods in higher education. In the past, knowledge of argumentative writing and computer programming--among others--were taught by demonstrating good writing and programming technique in class, by assigning essays and programming exercises at home, and by evaluating the quality of corresponding end products. Today, students can employ generative AI systems such as ChatGPT to complete the relevant assignments. As a result, traditional learning objectives appear out of date; traditional learning activities appear to be a waste of time; and traditional assessment methods appear ineffective. In this talk, we explore the possibility of restoring constructive alignment in the age of generative AI. First, we reflect on learning objectives: which knowledge and skills will citizens actually need in a society in which generative AI is increasingly powerful and prevalent? Second, we review some recent attempts to tune assessment methods to these new learning objectives. Third, we consider the possibilities of promoting the relevant knowledge and skills by incorporating generative AI in the classroom and at home.