International Society of the Learning Sciences

Exploring AI Ethics through educational scenarios with AI generative arts apps

Anna Keune, Santiago Hurtado, and Živa Simšič
anna.keune@tum.de, s.hurtado@tum.de, and ziva.simsic@tum.de
Technical University of Munich

Abstract: Artificial intelligence (AI) tools and technologies have significantly expanded in recent years. Commercial tools that have not necessarily been designed for educational purposes and settings are also entering educational settings. This introduces ethical issues to educational settings that youth and educators need to be aware of. For this paper we designed youth-centered AI-based scenarios that show the possibility of a range of AI, with a focus on AI-generative arts applications, to foster conversations related to AI ethics. To test the scenarios we conducted qualitative and video-based research in a summer camp setting and analyzed how youth participants' practices and discussions reflected engagement with the OECD AI principles with a focus on (1) transparency and explainability and (2) human-centered values and fairness. Findings have implications for facilitating learning with and about AI ethics and the potential of AI generative arts applications for AI ethics learning.

The importance of fostering youth conversations about AI ethics in education

With the recent increase of artificial intelligence (AI) in education, the promise of AI-based educational technology (AI EdTech) for expanding learning opportunities substantially increased as well. The learning sciences have had a long-standing relationship with artificial intelligence (Schank, 1980; Doroudi, 2022), with the field pivoting toward conceptualizing AI literacy, including ethics related to AI (Morales-Navarro et al., 2023; Ng et al., 2021; Nguyen et al., 2022). Educators and young people need to weigh the AI's implications on teaching and learning and make informed decisions about its use as AI also in relation to the ethical issues AI can introduce to educational settings (Ng et al., 2021). To date, AI ethics principles have been discussed as evaluation tools for policy documents (Jobin et al., 2019), basing reviews of principles on how they should be presented and understood by stakeholders, AI developers, and designers (Borenstein & Howard, 2021), and educators (Roschelle et al., 2020). The most prevalent AI ethics principles are the Organization for Economic Co-operation and Development AI principles (OECD, 2021), encompassing inclusive growth, sustainable development and well-being, human-centered values and fairness, transparency and explainability, robustness, security and safety, and accountability.

Yet, current AI ethics frameworks are not designed for educational settings and there is a need of further expanding attention on AI ethics principles related to education to take into consideration how AI changes what can be learned and taught and how (Holmes et al., 2021). Collectively, the work suggests the importance of supporting young people in learning about how to identify and communicate ethical issues related to AI in education (AIED). Notably, Antle et al. (2022) showed that learning about (design) ethics can be done with a hands-on approach and critical reflection. In fact, constructionist approaches to learning highlight the importance of active engagement with the design of personally meaningful and shareable artifacts as a means to internalize complex conceptual knowledge (Papert, 1993). Most recently, constructionist approaches to AI have focused on the tools available to learners for creating and experimenting with AI (e.g., machine learning to design plush toys; Tseng et al., 2021). From a constructionist approach, arts-based approaches that make it possible to create personally meaningful projects with AI generative arts apps seem particularly well suited for investigating AI ethics through first hand and personally meaningful creative practices. However, we know less about how children talk about and identify ethical principles related to AI. This led us to ask: How do young people (age 13-14) practice OECD AI principles while creating projects with AI-generative art?

Methods

This qualitative study facilitated a 4-day (6 hours each) summer camp over the course of four consecutive days at a German public university. The summer camp focused on AI tools and their ethical implications, social media algorithms, and the design of a tool that could help educators and learners be informed about AI tools. The workshop participants were 11 girls (self-identified; 13-14 years old). Of these, 9 girls agreed to participate in the data collection and were divided into two small groups (3-4 participants per group). The workshop covered AI-

generated deepfakes, AI generative writing tools, and other generative AI art with Adobe Firefly. This short paper focuses on AI generative arts apps in educational settings.

To support youth in discussing ethical issues related to AI tools, we developed scenarios with hypothetical educational settings. The scenarios were printed documents that presented the possibility of using different AI-based tools in educational settings in student-centered ways. For instance, the use of AI-generative arts to augment a drawing for a presentation. We designed the scenarios with the aim of fostering conversations related to AI ethics. During the workshops, youth participants explored the Adobe Firefly AI-arts tool. On the third day of the workshop, the focus of this paper, we introduced three scenarios to the participating youth: (1) AI-generative arts app to create an homage to an artist, (2) AI-generative arts app to augment students' drawings, and (3) AI-generative arts apps to illustrate local celebrations. In this paper, we focus on scenarios 1 and 2. Scenario 1 presented one original pencil drawing on beige paper of a woman with a blue bird attached to her head, and three AI-generated versions to the youth. Scenario 2 contained a simple drawing of a boy as if drawn by a child and an AI-generated version of the image. Youth engaged in interacting with an AI image generative tool and the material prompts by comparing them and trying out different prompts. Youth edited themselves in pictures, generated content through prompts, and tried replicating images presented in the scenarios.

The video recorded workshops resulted in 27.6 hours of video data, filmed with two cameras, each focusing on one of two small groups of youth participants (3-4 youths per group). The iterative and thematic analysis coded the youth's verbal utterances and interactions based on the OECD AI principles. In collaboration with public policy and governance project partners, the iterative analytical process served the development of a codebook for recognizing the OECD AI principles in action. While we coded discussions related to all OECD AI principles, in the interest of space, we will highlight how the principles of (1) transparency and explainability and (2) human-centered values and fairness were discussed by youth in relation to AI-generative arts scenarios. Transparency and explainability include understanding outcomes of AI systems and challenging system outcomes, where young people demonstrate how the AI provides information or not in simple enough terms about how it arrived at an outcome. Human-centered values and fairness include non-discrimination, equality, diversity, social justice, and fairness, where young people demonstrate how AI involves biases that may mis or over-represent depending on data sets it is trained on and that AI can influence learning outcomes and processes when accomplishing tasks.

Findings

The analysis of the workshop data showed that youth engaged in discussions of the OECD AI principles based on their experiences of creating personally meaningful AI generative arts projects.

AI-generative arts app to create an homage to an artist

The youth were introduced to an original drawing by a local artist, together with three AI-generated replicas of the artist's original drawing (Figure 1). Conversations related to transparency and explainability unfolded when participants commented on the styles and motifs of the original and AI-generated artworks. For example, Sarah (all names are pseudonyms) highlighted how the images appeared as if created by different people without providing an explanation about how the styles entered the artworks. Sarah said: "[…] *like they are from a different artist,*" and later on, "[…] *you can't predict what the AI would generate for you* […] *you can't explain why AI chose to show certain generated images*." Sarah noticed that the AI changed the original artwork into different styles and that these styles could be associated with entirely different artists. However, the AI did not transparently explain which artist's style inspired these changes and under which conditions the changes took place. Participants identified that the AI tool did not provide meaningful information and context to support understanding of the AI system and its outcomes, such as how the AI-generated artworks were created. Additionally, the youths deepened their critique of the AI's explainability as Vera, another youth in the same group, noted that the source motifs of the AI-generated images were not clear. Vera said: "*Who says that AI doesn't pick out an image that is already existing and just creates it in another style, but it's still, for example, just your face from an Instagram post*". Vera explained that it was not transparent whether the AI-generated image was based on an image the AI pulled from the internet and altered (compared to inventing one) and whether or to what extent an image that was created by a person and posted online (e.g., on social media) was altered by the AI. As neither the level and kind of manipulation nor the source of the manipulation were made accessible, it was not transparent whether the image was generated at all.

Figure 1
Youth interacting with AI-generative arts app to create an homage to an artist.

Youth also talked about the principle of human-centered values and fairness in relation to their AI-generated art productions in relation to how it could be used for learning. Vera commented: "*The student could lie to the teacher and say it's his artwork when it was generated by AI*." With this quote, Vera implies that the artwork generated by the AI tool is not the same as artwork generated by a human hand holding a pencil and, therefore, should be disclosed. This relates to the fairness aspect of human-centered values and fairness. The youth implied that using AI takes different sets of learning outcomes compared to not using AI to generate artwork. Therefore, for the youth, it is important to disclose when and how AI was drawn on to generate creative works. This understanding of fairness also relates to transparency and explainability because the AI tool could include features for disclosing its use.

AI-generative arts app to augment students' drawings

Using AI-generative art to augment students' drawings also made it possible for youth to discuss OECD AI principles in relation to their personal projects. Youths engaged with the principle of transparency and explainability when comparing and contrasting the original youth drawing to the AI-altered image. For instance, Emma, one of the youths, said: "*Its [AI is] not really transparent. I think AI just works randomly*". The statement suggests that the AI tool did not transparently present the layers of edits it performed, and the image sources it drew on to arrive at the altered image, thus, understanding how the AI arrived at the outcome was not possible.

Youth further engaged with the OECD AI principle of human-centered values and fairness while altering photographs of themselves with the AI-generative arts tool. Emma, who had short hair, loaded a photograph of herself into the app to generate variations of herself, similar to the way the student drawing in the scenario was generated. The AI quickly generated a version of herself with added male features. She then prompted the tool to show a female character, and the AI added long hair and a fringe to her photo. Emma commented on this later: *"AI understands things in a narrow way. If, for example, a boy gets turned into a girl, there will be mean people laughing at him."* With this statement, she implies that the tool can have negative consequences related to how the AI interprets people, misrepresenting those using the tool. Emma noticed that the AI tool fails to understand rich representations of gender, which uncovers opportunities to explore how AI is trained and how this may impact AI outcomes.

Discussion

The findings suggest that the AI-generative arts scenarios supported youth in engaging with the OECD AI principles. Including arts-based approaches to engage in ethics that can be a rich approach for young people to critically evaluate the use of AI in education based on their personal design-based and production-based experiences. To further foster engagement with the principles, the findings suggest that the facilitation could focus on explicitly eliciting conversations about artistic style, for example, by asking who owns an artistic style and considering how AI might choose different styles, and identifying differences between AI-generated art and original artworks. This could broaden participants' perceptions of AI-generated artworks and spark further conversations about the OECD AI principles of human-centered values and fairness, and transparency and explainability. The youths' discussions related to transparency and explainability suggest that activities for tracing AI decision-making processes, including the kind of artist styles used to alter original artworks, could deepen

student conversations about this principle. To further unpack transparency and explainability in relation to the use of AI-generative arts to augment student drawing, the scenarios could include a time-lapse video of the creation of a digital artwork, similar to the videos some artists create to demonstrate the creation of a particular filter or image manipulation using digital graphic design or video editing tools. Moreover, youth's experiences with how AI represents them and others can foster conversations related to human-centered values and fairness. Diving into how AI generates content through analyzing outcomes can foster conversations about data sets and bias within AI systems. Activities that implement student representations of themselves and others could further foster discussions about human-centered values and fairness. The way AI outcomes represent people can bring about discussions of data sets, biases, and their implications for educational settings. Furthermore, we observed overlaps between the AI ethics principles within the youth interactions and conversations. Getting youth to articulate these intersections would be an important next step in better understanding how learning about and with AI ethics principles can take place in productive ways. Overall, engaging with AI-generated arts through the production of personally meaningful projects supported youth in engaging with AI ethics principles based on first-hand experiences. Notably, these activities also served as a window into conversations related to art education, such as art styles, art influences, authorship, and participation.

References

Antle, A. N., Murai, Y., Kitson, A., Candau, Y., Dao-Kroeker, Z., & Adibi, A. (2022). "There are a LOT of moral issues with biowearables" . . . Teaching design ethics through a critical making biowearable workshop. In *Interaction Design and Children*. https://doi.org/10.1145/3501712.3529717

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, 1(1), 61–65. https://doi.org/10.1007/s43681-020-00002-7

Doroudi, S. (2022). The intertwined histories of artificial intelligence and education. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-022-00313-2

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T. T., Shum, S. B., Santos, O. C., Rodrigo, M. M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2021). 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

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Morales-Navarro, L., Kafai, Y. B., De Castro, F. B., Payne, W. D., DesPortes, K., DiPaola, D., … Vakil, S. (2023). Making sense of machine learning: Integrating youth's conceptual, creative, and critical understandings of AI. *arXiv (Cornell University)*.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers & Education: Artificial Intelligence*, *2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. T. (2022). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, *28*(4), 4221–4241. https://doi.org/10.1007/s10639-022-11316-w

Organization for Economic Co-operation and Development (2021). *OECD Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Papert, S. (1993). The children's machine: Rethinking school in the age of the computer. Basic Books.

Roschelle, J., Lester, J., & Fusco, J. (2020). *AI and the future of learning: Expert panel report*. Digital Promise. https://circls.org/reports/ai-report

Schank, R. C. (1980). How much intelligence is there in artificial intelligence? *Intelligence*, *4*(1), 1–14. https://doi.org/10.1016/0160-2896(80)90002-1

Tseng, T., Murai, Y., Freed, N., Gelosi, D., Ta, T. D., & Kawahara, Y. (2021). PlushPal: Storytelling with interactive plush toys and machine learning. *Interaction Design and Children*. https://doi.org/10.1145/3459990.3460694

Acknowledgments

This work was supported by the Institute for Ethics in Artificial Intelligence and the Institute for Advanced Study (TUM-IAS) at the Technical University of Munich.