5

Citizenship, Censorship, and Democracy in the Age of Artificial Intelligence

Tetiana Matusevych, Margarida Romero, and Oksana Strutynska

Abstract This chapter delves into the ethical dilemmas that arise from the incorporation of artificial intelligence (AI) into the field of education. It emphasises the importance of media literacy, AI literacy, and critical use of digital technologies in order to combat information conflicts, political manipulation, and AI inequality, among other issues. Potential threats to citizenship, such as AI censorship and disinformation, are examined in this chapter. Discourse is devoted to the dangers of deepfake technology as it pertains to the dissemination of false information and the manipulation of public sentiment; the significance of comprehending AI fundamentals and enforcing ethical standards is underscored.

T. Matusevych (✉) · O. Strutynska
Dragomanov Ukrainian State University, Kyiv, Ukraine
e-mail: t.v.matusevych@udu.edu.ua

O. Strutynska
e-mail: o.v.strutynska@udu.edu.ua

M. Romero
Université Côte d'Azur, Nice, France
e-mail: margarida.romero@univ-cotedazur.fr; margarida.romero@unice.fr

Notwithstanding the potential hazards, this chapter acknowledges the prospective advantages of AI in the field of education, which encompass gamification and adaptive learning paths. The text culminates by emphasising the significance of AI acculturation in enabling individuals to comprehend the ethical intricacies and arrive at well-informed judgements regarding the impact of AI on democracy, education, and citizenship.

Keywords AI fundamentals · AI literacy · Digital competence · Citizenship · AI ethics · AI algorithms

Introduction

The rapid development of digital technologies and their widespread adoption make it necessary to develop citizens' digital competence and to actively engage people in their responsible use (Alexandre et al., 2020). Important components of citizens' digital competence include artificial intelligence (AI) literacy, critical use of digital technologies, and media literacy. These literacies are required due to the threats of information wars, political manipulation, but also digital and AI inequality and divisions (Calvo et al., 2020). These issues highlight the need to engage modern youth in responsible digital citizenship, which is an important component of citizenship education.

One of the conceptual initiatives to overcome these challenges was the creation of the *Digital Citizenship Education program* by the Council of Europe (Committee of Ministers Council Europe, 2019), which aims to provide young people with innovative opportunities to develop the values, attitudes, skills, and knowledge necessary for every citizen to fully participate and fulfil their responsibilities in society. *Digital citizens* are defined by the Council of Europe as people who are able to use digital tools to create, consume, communicate, and interact positively and responsibly with others. Digital citizens understand and respect human rights, embrace diversity, and prioritise lifelong learning as a way of keeping pace with societal changes.

Digital civic education is a holistic approach that aims to develop the basic skills and knowledge needed in today's connected world, as well as

to foster values and attitudes that will ensure their wise and meaningful use. The development, regulation, involvement, and use of AI in education are part of this approach. In particular, it states that AI, like any other tool, offers many opportunities, but also poses significant threats that require the consideration of human rights principles in the early stages of its application. Educators need to be aware of the strengths and weaknesses of AI in education in order to empower their digital civic education practices (Committee of Ministers Council Europe, 2019). In this regard, Artificial Intelligence for Social Good (AI4SG) is also an important concept for civic education, which is becoming increasingly popular in professional circles (Hager et al., 2019). Projects aimed at using AI for the social good range from applications to help the hungry, to combating natural disasters, game-theoretic models aimed at poaching prevention, online HIV education for homeless youth, prevention of gender-based violence, and psychological support for students. (Floridi et al., 2020).

But despite the fact that new initiatives are emerging every day, researchers note that there is still only a limited understanding of what constitutes AI 'for the social good'. The lack of a clear understanding of what makes AI socially useful in theory, or what can be described as AI4SG in practice, and how to replicate its initial successes from a policy perspective creates at least two major obstacles for developers: avoidable mistakes and missed opportunities. AI software is shaped by human values, which, if not carefully selected, can lead to 'good AI goes bad' scenarios (Floridi et al., 2020). Thus, the issues of the value component and ethical principles of AI use are becoming increasingly relevant.

In general, the ethical issues of AI development and application are the focus of many researchers and international institutions, which has led to the creation of numerous initiatives, committees, and institutes of AI ethics. The Montreal AI Ethics Institute (MAIEI) published a guideline including the norms for responsible AI research and use (Dilhac et al., 2018). The Montreal Declaration for a Responsible Development of Artificial Intelligence is founded upon a set of ethical principles centred on seven fundamental values: well-being, autonomy, justice, privacy, knowledge, democracy, and responsibility. As a result of the analysis, researchers identified 84 published sets of ethical principles for AI that

coincide in five areas: transparency, fairness and honesty, harmlessness, responsibility, and confidentiality (Jobin et al., 2019).

Current Problems in the Integration of AI into Education

Despite the extensive debate around the ethical use of AI, there are a number of problems that have not yet been solved. The main ones are as follows:

1. *The polymorphic nature of the ethics of using AI in education.* AI ethics raises a number of complex issues centred on data (e.g., consent and data privacy) and how this data is analysed (e.g., transparency and trust). It is clear, however, that the ethics of AI in education cannot be reduced to issues of data and computational approaches alone. Research into the ethics of data and computing for AI in education is necessary but not sufficient. The ethics of AI in education should also take into account the ethics of education (Holmes et al., 2022).

2. *Potential threats to fundamental rights and democracy.* The results produced by AI depend on how it is designed and the data it utilises. Both the design and the data may be intentionally or unintentionally biassed. For example, some important aspects of the problem may not be programmed into the algorithm or may be programmed to reflect and repeat structural biases (Borenstein & Howard, 2021). In addition, the use of numbers to represent complex social realities may carry risks of apparent simplicity (Holmes et al., 2022). Another significant threat to human rights is data bias, which can affect the training of large language models (LLM) and result in biassed models For example, if an educational institution has student performance data biassed towards a certain ethnic, gender, or socioeconomic group, the AI system can learn to favour students from that particular group (Manyika & Silberg, 2019; Roselli et al., 2019). A further example of gender bias, this time in neural network algorithms for image generation, is provided in the research by Nikolić and Jovičić (2023). When working with the visual generative AI

services *DALL-E 2* and *Stable Diffusion*, the researchers observed the types of images the neural networks produced, particularly in relation to the representation of women in STEM. Specifically, when using the prompts 'engineer', 'scientist', 'mathematician', or 'IT expert', between 75 and 100% of the AI-generated images featured men, reinforcing stereotypes about male-dominated STEM professions: both as occupations that primarily attract men and as professions where men are prominent compared to women (Nikolić & Jovičić, 2023).

3. *AI colonialism in education.* AI colonialism can be understood in relation to the 'control, domination, and manipulation of human values' (Faruque, 2022, para. 25). AI colonialism is developed under an industrial perspective in which the business applications of AI are moved forward mainly for commercial objectives instead of humanistic ones. In 2020, in spite of the coronavirus pandemic, venture capital investment in AI startups reached US$75 billion for the year, of which about US$2 billion was invested in AI in education companies, predominantly based in the US. It is these companies that sell their interpretations around the world, creating what is called the colonialism of AI in education. This problem makes addressing cultural diversity one of the most challenging topics in AI in education (Blanchard, 2015). We can also consider the use of the English language as a form of colonialism in technology, given that it is the default language of consumption as well as the academic language in which AI specifications, frameworks, and ethical guidelines are produced and debated. Given that language not only conveys a symbolic representation of society, but also a cultural perspective of a certain context, we should acknowledge that having a default language reduces cultural perspectives and creates an accessibility challenge for those lacking the requisite language skills.

4. *Lack of a universal approach to regulating the ethical issues of using AI in education.* Unlike healthcare, where there are long-established ethical principles and codes of conduct for the treatment of humans, education (outside of university research) does not have the same universal approach or a commonly accepted model for ethics committees (Holmes et al., 2022). As such, when it comes to the

use of AI in education, most discussions treat students as data subjects rather than human beings. The learner activity is quantified in a way that reduces the representation of the learner into a quantitative model, mostly based on certain learning analytics. As a result, commercial players and schools may involve children in AI-driven systems without any ethical or other risk assessment (Holmes et al., 2022). In Europe, the AI Act (Pagallo et al., 2022) aims to advance the regulation of AI systems with the goal of making them 'safe, transparent, traceable, non-discriminatory, and environmentally friendly' (EU AI Act, 2023). Meanwhile, other countries are developing their own initiatives independently, without joint cooperation on a universal approach to AI regulation.

5. *Challenges of 'ethics washing'*. A large number of commercial actors in the tech sector publish ethics guidelines to 'wash away' concerns regarding their policies. This increasing instrumentalization of ethical guidelines by technology companies is called 'ethics washing' and refers to a situation where ethics is used by companies as an acceptable facade to justify deregulation, self-regulation, or market-driven governance, and is increasingly identified with the self-serving use and pretence of ethical behaviour (Bietti, 2020; van Maanen, 2022). For AI in education, given that children are the primary users of these commercial AI technologies, it is important to develop and implement robust ethical guidelines and avoid any 'ethics washing' (OECD, 2021).

6. *Lack of systematic application of ethical principles for the use of AI*. Although universities usually have robust research ethics procedures, most university or commercial AI research companies do not oversee AI ethics. This may be partly due to the fact that, in the early days of AI, research using human data was considered minimally risky (Holmes et al., 2022). The way AI is integrated into educational academic activities is regulated at different levels, including in schools or research labs, university and school departments, and government bodies such as the Ministries of Education of individual countries, all of whom may have different policies related to the ethical principles of AI. We should also consider that the end-user may not be able to evaluate the ethical principles of AI technologies,

meaning that AI creators have a responsibility to design and deploy AI technologies that recognise the varied ethical principles found in different national and professional domains.

7. *Threats of excessive and unjustified use of AI.* Overuse of AI can be problematic. Examples include investing in AI applications that turn out to be ineffective or applying AI to tasks for which it is not suited, such as explaining complex societal problems (Holmes et al., 2022). Automation of certain human activities related to education necessitates a high degree of sensitivity in determining what is appropriate, rather than merely what is possible. For example, it is important to continue to learn foreign languages even if automatic translations are available. Not only because human-to-human interaction is more empathetic, but also because there is a potential cognitive decline if we are not engaging in effortful cognitive activities such as learning, speaking, or writing in foreign languages. Finally, when evaluating the applications of AI, it is important to consider the energy consumption associated with its use versus alternative information search processes that may be less energy-intensive (Yang et al., 2021).

8. *Challenges of accountability and responsibility.* For educational institutions, the question is not only whether AI can be used in children's education, but also how accountability and responsibility should be determined when educators decide to adopt or reject any systemic recommendation (Holmes et al., 2022).

9. *Challenges of conflict of interest or 'AI loyalty'.* The concept of conflict of interest, or 'AI loyalty' (Aguirre et al., 2021) in educational institutions is largely absent in the current body of literature. For whom do AI systems work? Is it the students, schools, the education system, commercial players (e.g., AI edtech companies), or politicians? The question is not necessarily about the ethics of the technology itself, but rather about the ethics of the decision makers leading the companies behind AI's development, implementation, and use (Holmes et al., 2022). Understanding AI loyalty means clearly defining ownership and any conflicts of interest. To increase the transparency and credibility of AI applications, system developers and controllers should be required to clearly align the loyalty

of their AI systems and governance structures with the interests of learners and other stakeholders of their systems (Holmes et al., 2022). In their daily work, educators should be acculturated to the fundamentals of AI in order to decide which of its uses are relevant for the teaching or learning process, but also to be able to decline the use of certain AI technologies that are not relevant to their teaching practices.

10. *Decreased social connection and overreliance on technology.* There is a risk that an increase in time spent using AI systems will lead to a decrease in the interactions between students, teachers, and classmates. Kids might also begin substituting other human interactions (e.g., conversions with families and friends) with conversational AI systems amplifying and exacerbating the public health crises of loneliness, isolation, and disconnection (Bailey, 2023).

11. *AI threats to citizenship.* Widespread threats, such as AI censorship and AI misinformation, can lead to the manipulation of public opinion, contribute to the incitement of conflicts on various grounds (e.g., racial, religious, etc.), and contribute to the worsening of existing inequalities and stereotypes (e.g., gender inequalities). Filgueiras (2022) highlights the risks of AI in the context of authoritarianism in developing countries, which can amplify surveillance mechanisms put in place to control citizens considered a threat to the current establishment.

12. *AI censorship.* Due to the rapid development of AI, the threat of AI censorship has been added to internet censorship and the deletion of uncomfortable content practices (i.e., for political elites who want to influence public opinion). Censorship of political content on Chinese and Russian social media (e.g., active deletion of messages posted by individuals) is already having a corresponding impact on public opinion in these countries (Bamman et al., 2012; Ermoshina et al., 2022; Yang, 2016). Furthermore, AI censorship can dramatically affect the objectivity of people's perceptions of information. For example, motivated groups can adjust the algorithms of AI systems in such a way that inconvenient facts are hidden from the general public. The research on setting censorship parameters in neural networks and AI services by Ermoshina (2023) is an example of this

type of manipulation. Another example of this type of manipulation can be found in the Russian neural network Kandinsky 2.1, which returns images of flowers when given prompts of 'war in Ukraine', 'Ukrainian Flag', or even just the word 'Ukrainian' whether entered in Russian or English. This censorship is also present in the visual generative AI services of other countries: Chinese *ERNIE-ViLG*, American *DALL-E 2*, *Stable Diffusion*, *Midjourney* (Ermoshina, 2023).

13. *AI misinformation.* One of the biggest AI threats is disinformation and the potential of AI systems to generate massive amounts of propaganda. Dishonest groups can use these manipulations to incite hatred on a variety of grounds (e.g., racial, religious, etc.), to hurt people's feelings, and to arouse anger and other unpleasant emotions. An example of such disinformation can be found in the recent blog post exposing AI-generated fake images of violence in Gaza and Israel (Gault, 2023).

    One strategy for influencing public opinion and spreading misinformation or panic among populations is the creation and distribution of fake videos using deepfake technology. Deepfake is an AI-based method of generating fabricated images and videos by combining and superimposing images or videos onto other images or videos out of context (Sharma & Kaur, 2022). Often, this technology is used to discredit a person or for purposes of revenge. Deepfake videos regarding the war in Ukraine are examples of attempts to sow panic among the Ukrainian population with one example showing Ukrainian President Volodymyr Zelensky issuing a fabricated statement calling for the end of resistance to Russian aggressions (Rayon, 2022). Another recent example of such manipulations is a deepfake of the Commander-in-Chief of the Armed Forces of Ukraine, Valeriy Zaluzhny, calling for a coup d'état against the President of Ukraine (Espreso, 2023). Other examples of deepfake manipulations include the creation of nude images of Spanish schoolgirls highlighting an alarming trend among younger users. In this particular instance, a group of mothers, refusing to tolerate such exploitation, targeted not only the company responsible for creating the images, but also appealed to educational authorities to

highlight the severe impact such images can have on the affected teenagers. The use of deepfakes poses a significant challenge to the epistemic trust undermining the reliability and importance of social communication (Twomey et al., 2023).

Discussion

We have addressed in this chapter the different threats of AI in relation to citizenship, democracy, and censorship. Part of the challenges arising in the ethical use of AI in education require a better understanding of AI fundamentals. The understanding of data (i.e., collection and management), but also the way algorithms work, is important for ensuring that teachers and learners can develop their work from a critical thinking perspective.

AI, like any other tool, offers many opportunities, but also poses many threats that require the consideration of human rights principles at the earliest stages of its implementation. Educators should be aware of the strengths and weaknesses of AI in education to empower their digital, civic education practices. In particular, AI services and tools can be used to design adaptive learning paths, recommend resources, and offer scaffolding and other forms of assistance (e.g., to assign different levels of complexity, interaction, and differentiation). AI can also support the gamification of learning through the creation of engaging and interactive scenarios, challenges, and simulations that promote problem solving, critical thinking, creativity, collaboration, and digital literacy or citizenship. Furthermore, AI-based 'chatbots' can be developed for teachers and parents to support both the disciplinary and transversal aspects of education.

While AI can pose risks to citizenship when lacking an acculturation and regulation of its use, AI in education also provides advantages for both educators and learners by allowing them to avoid routine work and focus on creative tasks (Romero et al., 2021; Septiani et al., 2023). AI, through machine and deep learning, can enrich education and profoundly affect the interactions between teachers, students, and

citizens at large. In this way, AI in education can promote free expression and independent and critical thinking through learning opportunities (Committee of Ministers Council Europe, 2019; Richardson & Milovidov, 2019).

According to Frąckiewicz (2023), the main areas of AI that can contribute to the development of education for responsible citizenship include: (i) developing the global dimension of responsible citizenship through the promotion of intercultural understanding; (ii) facilitating access to information and education; (iii) informed citizenship; (iv) democratisation of education, making it more accessible to students; and (v) the development of digital literacy skills. By developing digital literacy, AI can help students become responsible consumers of information and active participants in online discussions (Frąckiewicz, 2023).

The potential of AI for improving or harming citizenship and education will depend on the way citizens and governments decide to use and regulate it. An acculturation to the fundamentals of AI is required for each citizen to move beyond the role of mere 'consumer' of technology while also developing a critical, yet creative, perspective on its impact related to citizenship, well-being, education, and democracy.

References

Aguirre, A., Reiner, P. B., Surden, H., & Dempsey, G. (2021). AI loyalty by design: A framework for governance of AI. In J. B. Bullock & others (Eds.), *The Oxford handbook of AI governance*. Available at SSRN Scholarly Paper 3930338. https://papers.ssrn.com/abstract=3930338

Alexandre, F., de Barretin, R., Becker, J., Comte, M.-H., Courbin, M., Cruchon, S., Lagarrigue, A., Masse, B., De Quatrebarbes, S., Stein, J., Terosier, C., & Vieville, T. (2020). *Understanding intelligently artificial intelligence: A citizens' open formation.* International Workshop on Education in Artificial Intelligence K-12 (EDUAI-20).

Bailey, J. (2023, Серпень 8). AI in education. *Education Next, 23*(4), 28–35. https://www.educationnext.org/a-i-in-education-leap-into-new-era-machine-intelligence-carries-risks-challenges-promises/

Bamman, D., O'Connor, B., & Smith, N. (2012). Censorship and deletion practices in Chinese social media. *First Monday*. https://doi.org/10.5210/fm.v17i3.3943

Bietti, E. (2020). From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 210–219). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372860

Blanchard, E. G. (2015). Socio-cultural imbalances in AIED research: Investigations, implications and opportunities. *International Journal of Artificial Intelligence in Education, 25*(2), 204–228. https://doi.org/10.1007/s40593-014-0027-7

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

Calvo, D., Cano-Orón, L., & Abengozar, A. E. (2020). Materials and assessment of literacy level for the recognition of social bots in political misinformation contexts. *ICONO 14, Revista de comunicación y tecnologías emergentes, 18*(2), 111–136.

Committee of Ministers Council Europe. (2019). *Recommendation CM/Rec(2019)10 of the Committee of Ministers to member States on developing and promoting digital citizenship education.* https://rm.coe.int/090000168098de08

Dilhac, M. A., Christophe, A., & Lehoux, P. (2018). *La Déclaration de Montréal IA responsable.* https://declarationmontreal-iaresponsable.com/la-declaration/

Ermoshina, K. (2023, Червень 15). *Украина, протесты и сексизм. Ксения Ермошина—О цензуре в нейросетях* (Ukraine, protests and sexism. Ksenia Yermoshina—On censorship in neural networks). Теплица социальных технологий. https://te-st.org/2023/06/15/ai-censorship/

Ermoshina, K., Loveluck, B., & Musiani, F. (2022). A market of black boxes: The political economy of Internet surveillance and censorship in Russia. *Journal of Information Technology & Politics, 19*(1), 18–33. https://doi.org/10.1080/19331681.2021.1905972

Espreso. (2023, Листопад 7). *Мережею шириться діпфейк із нібито Залужним і закликом до збройного повстання* (Dipfake with allegedly Zaluzhnyi and a call for armed rebellion spreads online) [News website]. Espreso. https://espreso.tv/merezheyu-shiritsya-dipfeyk-iz-nibito-zaluzhnim-i-zaklikom-do-zbroynogo-povstannya

European Parliament. (2023, August 6). *EU AI Act: First regulation on artificial intelligence*. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Faruque, M. U. (2022). AI versus human consciousness: A future with machines as our masters? *Renovatio: The Journal of Zaytuna College*. https://renovatio.zaytuna.edu/article/ai-versus-human-consciousness

Filgueiras, F. (2022). The politics of AI: Democracy and authoritarianism in developing countries. *Journal of Information Technology & Politics, 19*(4), 449–464. https://doi.org/10.1080/19331681.2021.2016543

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to design AI for social good: Seven essential factors. *Science and Engineering Ethics, 26*(3), 1771–1796. https://doi.org/10.1007/s11948-020-00213-5

Frąckiewicz, M. (2023, May 4). *The role of AI in fostering global citizenship education*. TS2 SPACE. https://ts2.space/en/the-role-of-ai-in-fostering-global-citizenship-education/

Gault, M. (2023, November 7). *Adobe Is selling AI-generated images of violence in Gaza and Israel*. Vice. https://www.vice.com/en/article/3akj3k/adobe-is-selling-fake-ai-generated-images-of-violence-in-gaza-and-israel

Hager, G. D., Drobnis, A., Fang, F., Ghani, R., Greenwald, A., Lyons, T., Parkes, D. C., Schultz, J., Saria, S., Smith, S. F., & Tambe, M. (2019). *Artificial intelligence for social good* (arXiv:1901.05406). arXiv. https://doi.org/10.48550/arXiv.1901.05406

Holmes, W., Persson, J., Chounta, I.-A., Wasson, B., & Dimitrova, V. (2022). *Artificial intelligence and education: A critical view through the lens of human rights, democracy and the rule of law*. Council of Europe.

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

Manyika, J., & Silberg, J. (2019, June). *Notes from the AI frontier: Tackling bias in AI (and in humans)*. Multidisciplinary symposium on ethics in AI hosted by DeepMind Ethics and Society. https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/MGI-Tackling-bias-in-AI-June-2019.pdf

Nikolić, K., & Jovičić, J. (2023, April 3). Reproducing inequality: How AI image generators show biases against women in STEM. UNDP. https://www.undp.org/serbia/blog/reproducing-inequality-how-ai-image-generators-show-biases-against-women-stem

OECD. (2021). *OECD digital education outlook 2021: Pushing the frontiers with artificial intelligence, blockchain and robots*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/education/oecd-digital-education-outlook-2021_589b283f-en

Pagallo, U., Ciani Sciolla, J., & Durante, M. (2022). The environmental challenges of AI in EU law: Lessons learned from the Artificial Intelligence Act (AIA) with its drawbacks. *Transforming Government: People, Process and Policy, 16*(3), 359–376. https://doi.org/10.1108/TG-07-2021-0121

Rayon. (2022, Березень 17). *Компанія Мета видалила діпфейк із Зеленським* (Meta deletes Zelenskyy's diplomatic account). Rayon. https://rayon.in.ua/news/498190-kompaniya-meta-vidalila-dipfeyk-iz-zelenskim

Richardson, J., & Milovidov, E. (2019). *Digital citizenship education handbook: Being online, well-being online, rights online*. Council of Europe.

Romero, M., Heiser, L., & Galindo, L. (2021). *L'intelligence artificielle en éducation, formation/acculturation et modélisation*. https://hal.science/hal-03195064

Roselli, D., Matthews, J., & Talagala, N. (2019). *Managing bias in AI*. Paper Presented at the Companion Proceedings of The 2019 World Wide Web Conference, 539–544. https://doi.org/10.1145/3308560.3317590

Septiani, D. P., Kostakos, P., & Romero, M. (2023). Analysis of creative engagement in AI tools in education based on the #PPai6 framework. In B Z. Kubincová, F. Caruso, T. Kim, M. Ivanova, L. Lancia, & M. A. Pellegrino (Eds.). *Methodologies and intelligent systems for technology enhanced learning, workshops—13th international conference* (Vol. 769, pp. 48–58). Springer. https://doi.org/10.1007/978-3-031-42134-1_5

Sharma, M., & Kaur, M. (2022). A review of deepfake technology: An emerging AI threat. In B G. Ranganathan, X. Fernando, F. Shi, & Y. El Allioui (Eds.), *Soft computing for security applications* (pp. 605–619). Springer. https://doi.org/10.1007/978-981-16-5301-8_44

Twomey, J., Ching, D., Aylett, M. P., Quayle, M., Linehan, C., & Murphy, G. (2023). Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. *PLoS ONE, 18*(10), e0291668. https://doi.org/10.1371/journal.pone.0291668

van Maanen, G. (2022). AI ethics, ethics washing, and the need to politicize data ethics. *Digital Society, 1*(2). https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=27314669&v=2.1&it=r&id=GALE%7CA712345544&sid=googleScholar&linkaccess=abs

Yang, F. (2016). Rethinking China's Internet censorship: The practice of recording and the politics of visibility. *New Media & Society, 18*(7), 1364–1381. https://doi.org/10.1177/1461444814555951

Yang, S. J. H., Ogata, H., Matsui, T., & Chen, N.-S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence, 2*, 100008. https://doi.org/10.1016/j.caeai.2021.100008