

RE-EXAMINING THE FUTURE PROSPECTS OF ARTIFICIAL INTELLIGENCE IN EDUCATION IN LIGHT OF THE GDPR AND ChatGPT

Dr. John Y. H. BAI

ORCID: 0000-0001-6688-7837

Center for Open Education Research (COER)
Carl von Ossietzky University of Oldenburg
Oldenburg, GERMANY

Dr. Olaf ZAWACKI-RICHTER

ORCID: 0000-0003-1482-8303

Center for Open Education Research (COER)
Carl von Ossietzky University of Oldenburg
Oldenburg, GERMANY

Dr. Wolfgang MUSKENS

ORCID: 0000-0002-5515-8302

Center for Open Education Research (COER)
Carl von Ossietzky University of Oldenburg
Oldenburg, GERMANY

Received: 08/02/2023 **Accepted:** 16/05/2023

ABSTRACT

Artificial intelligence in education (AIEd) is a fast-growing field of research. In previous work, we described efforts to explore the possible futures of AIEd by identifying key variables and their future prospects. This paper re-examines our discussions on the governance of data and the role of students and teachers by considering the implications of 1) a recent case related to the General Data Protection Regulation (GDPR) and 2) the release of ChatGPT, a generative AI model capable to producing 'human-like' text. These events raise questions for the future of AIEd and the underlying function of assessment, and highlight the importance of active student participation in the integration of AI in education.

Keywords: Artificial intelligence in education, General Data Protection Regulation, ChatGPT, writing and assessment, AI and society.

INTRODUCTION

Artificial intelligence (AI) has been described as a 'general purpose' technology that has the potential to transform societies and economies (Tuomi, 2018). The UNESCO Commission on the Ethics of Scientific Knowledge and Technology (COMEST, 2019) notes that the term "artificial intelligence" does not have one singular definition but is often defined in terms of 'computers that perform tasks usually associated with human intelligence'. COMEST (2019) trace this form back to a 1955 research project, that specified their working assumption that: "the study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it" (McCarthy et al., 2016, p. 12). While our understandings of learning and other features of intelligence continue to evolve with ongoing research, an alternative definition that circumvents the unknowns of human intelligence is provided in Nilsson (2009): "artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment" (p. xiii). The later definition provides a functional

approach and covers a wide range of algorithms and computational systems that can learn to recognize patterns in complex data and generalize those patterns to novel data.

A growing research literature on artificial intelligence in education (AIEd) illustrates the interest in integrating AI technologies in education (see Bozkurt, Karadeniz, Baneres, Guerrero-Roldan, Rodriguez, 2021; Zawacki-Richter, Marin, Bond, & Gouverneur, 2019, for reviews). In their review of primary AIEd research, Zawacki-Richter et al. (2019) noted that very few papers were led by first authors from education departments, and very few discussed ethical considerations. This contrasts with numerous discussion papers that have highlighted the need to address the ethics of AIEd applications (e.g., du Boulay, 2022; Facer & Selwyn, 2021; Tuomi, 2018; Zeide, 2019) and with policy work on the ethics of AI in general (e.g., IEEE, 2019; OECD, 2022; UNESCO, 2021). These discussions raise a number of ethical issues with inequitable applications of AI, such as the risks to data protection and human agency, and the risk of further exacerbating existing inequalities (see Hu, Neupane, Echaiz, Sibal, & Rivera Lam, 2019, for discussion). Within these discussions, an oft-cited mechanism for mitigating risk is to incorporate comprehensive, multi-stakeholder dialogue during the design and implementation of AI systems. Thus, when considering AIEd specifically, the present research project aims to address the need Zawacki-Richter et al. (2019) identified “for educational perspectives on these technological developments” (p. 22).

The initial goal of the present project is to gather the perspectives of educators on the potential futures of AIEd. We first drew on techniques from future management (Fink & Siebe, 2011; Gutschow & Jorgens, 2016, 2019) to identify key variables (strategy elements) that will impact the future of AIEd and their possible outcomes (future options). Combinations of future options provide a structured method for developing strategic scenarios for AIEd applications, while making explicit the underlying assumptions for different macro and meso conditions (see Bai, Zawacki-Richter, & Muskens, 2022, for discussion). These strategic scenarios are developed to serve as the basis for international focus-group discussions and online surveys of faculty members in higher education.

Table 1. Strategy elements and future options identified in Bai, Zawacki-Richter, & Muskens (2022)

Strategy Element	Future Options				
Access to data	Laissez-faire	Heavily regulated	Limited but open	Limited and sporadic	Limited and proprietary
Funding	Primarily market-based	Public-private partnerships		Primarily government-funded	Mixed funding
Role of developers	Off-the-shelf	Subscription-based		Developers as researchers	Open-access
Role of teachers	Replacement	Division of labour		Retraining	Mixed-bag
Scope of applications	Piecemeal			Comprehensive	

In our previous work (Bai et al., 2022), we described the process and reasoning for identifying five strategy elements and their future prospects (Table 1). We also discussed the inherent subjectivity in imagining possible futures, arguing that “much like AI algorithms, human predictions of the future are based on data gathered from the past.” (p. 49). Since the submission of that paper, the ever-changing landscape of AIEd has drawn our attention to two events that prompt us to revisit our previous discussion; namely: 1) a decision by the Garante per la protezione dei dati personali (2021) to fine a university 200,000 Euro for the use of an e-proctoring system that violated sections of the General Data Protection Regulation (GDPR), and 2) the release of ChatGPT (OpenAI, 2022), a language model capable of producing ‘human-like’ text. In light of these two events, the present paper re-examines and expands our previous discussions on the governance of data and the roles of students and teachers.

THE GDPR AND GOVERNANCE OF DATA

In our previous work, we identified access to data as an important strategy element that would impact the trajectory of AIED development (Bai et al., 2022). Furthermore, we contrasted different regulatory approaches ranging from laissez-faire to heavily regulated, arguing that ‘light-touch’ regulatory environments (Cath, Wachter, Mittelstadt, Taddeo, & Floridi, 2018) may allow the collection of more learner data for AIED research and development, while further opening the risks of privacy violations and the misuse of learner data (e.g., Human Rights Watch, 2022; Laird, Grant-Chapman, Venzke, & Quay-de la Vallee, 2022; Russell, Reidenberg, Martin, & Norton, 2018). In contrast, regulatory environments that prioritize privacy and data protection may enforce restrictions that limit some AIED applications.

In regard to the heavily-regulated end of the continuum, our discussion of the EU’s GDPR identified a number of issues that seemed most relevant to AIED. These included the possible designation of some AIED applications as “high-risk” (Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM(2021) 206 final), the right not to be subject to a decision based solely on automated processing (Article 22), and the right to obtain an explanation for automated decisions (Recital 71; but see Edwards & Veale, 2018; Wachter, Mittelstadt, & Floridi, 2018, for critical discussions). However, a recent decision issued by the Italian Data Protection Authority (Garante per la protezione dei dati personali; hereafter “Garante”) highlights gaps in our discussion and identifies a range of articles relevant to future AIED developments in the EU.

In the Garante (2021) decision against the “Luigi Bocconi” University in Milan, a student complained of possible violations of the GDPR arising from the use of an e-proctoring system to identify and monitor students during online exams (see Bincoletto, 2021; European Data Protection Board, 2021; Liguori & Petrucci, 2021; for summaries of the decision in English). Specifically, software provided by Respondus Inc., a USA-based company, would (among other functions): take photos of the students’ identification cards and a panoramic shot of their rooms at the beginning of the exam, and then capture recordings of the students via their webcams and monitor other applications running on the students’ computers during the exam. An AI algorithm then processed the video images to raise ‘flags’ for “suspicious” behaviors (e.g., looking away from the screen, face partially present or absent from video) and assign a “Review Priority” (“low”, “medium”, or “high”) for teachers to review later.

The university noted that the e-proctoring system was implemented during the SARS-CoV-2 emergency and argued that consent from the students served as the legal basis for the processing of students’ biometric data (i.e., facial images of the students). However, the Garante (2021) decision noted that “students’ consent was not freely given... in relation to their unequal position in the university, especially during exams” (Bincoletto, 2021, p. 589). In addition, the decision noted issues with inadequate information given to students regarding the processing and retention of their data, including the transfer of personal data to the USA where Respondus Inc. is based. In total, the decision cited violations of multiple articles in the GDPR including: Article 5 paragraph 1, letters a), c), and e) (lawfulness, fairness, transparency; data minimization; storage limitation); Article 6 (lawful processing); Article 9 (special category data); Article 13 (information); Article 25 (privacy by default and by design); Article 35 (data protection impact assessment); and Articles 44 and 46 (transfers of personal data to a third country; European Data Protection Board, 2021).

This case supports our previous argument on the importance of data governance for the future of AIED but also highlights gaps in our discussion of the GDPR (Bai et al., 2022). Specifically, this case suggests that our previous discussion underestimated the degree of complexity involved in implementing some types of AIED applications in heavily-regulated environments. However, a complete discussion of the legal nuances of the Garante (2021) decision is beyond the scope of the current paper and the ability of its authors. Our summary of the decision is based on English-language sources consisting of a single academic paper (Bincoletto, 2021), materials from webpages (European Data Protection Board, 2021; Liguori & Petrucci, 2021), and a machine-translated version of the original decision. We hope that interested readers, particularly those who are or collaborate with legal scholars fluent in Italian, will continue to explore the implications of the decision for the future of AIED in the EU.

ChatGPT AND RE-ASSESSING ASSESSMENT

Another event that prompted us to revisit our previous discussion is the recent “research preview” release of ChatGPT (OpenAI, 2022). Following from the UNESCO (2015) Qingdao declaration, which noted that: “successful integration of ICT [information and communication technologies] into teaching and learning requires rethinking the role of teachers and reforming their preparation and professional development” (p. 2), our previous work discussed the potential roles of teachers as AI technologies are integrated into education. We argued that “the potential for AI technologies to transform education provides an opportunity to re-examine what it means to teach and to learn” (Bai et al., 2022, p. 63). The release and subsequent availability of ChatGPT presents such an opportunity by raising questions about the underlying function and purpose of written assessments.

ChatGPT is a recent addition to a group of Generative Pre-trained Transformer (GPT) models developed by OpenAI. ChatGPT’s precursor GPT-3 is a 175-billion parameter model trained on large amounts of text data scraped from the internet (Brown et al., 2020). Like ChatGPT, GPT-3 was “designed to generate sequences of words, code or other data, starting from a source input” or “prompt” (Floridi & Chiriatti, 2020, p. 684). However, one of the key developments since GPT-3.5 seems to be that ChatGPT was fine-tuned through human feedback (see Lambert & von Werra, 2022) to produce text in a ‘conversational way’ (OpenAI, 2022). In practice, the interface allows users to input sequences of prompts and obtain outputs in a familiar chatbot format. However, ChatGPT has attracted much attention for its ‘human-like’ output across a wide range of different prompts, demonstrating impressive ‘few-shot’ and ‘zero-shot’ performance without the need to fine-tune the model for specific tasks (Brown et al., 2020; cf. Sanh et al., 2022, for demonstration of improved zero-shot performance after multi-task training).

The release of ChatGPT drew substantial media coverage and boosted the estimated value of OpenAI (Rudolph et al., 2023), which operates as a “capped profit” company (Brockman et al., 2019). While some of the media coverage has been aptly critiqued as “alarmist and sensationalist reporting” (Rudolph et al., 2023, p. 13), a growing number of academic articles have also started to explore the capabilities and implications of ChatGPT for education (e.g., Gilson et al., 2022; Susnjak, 2022; Tate, Doroudi, Ritchie, & Xu, 2023; Ventayen, 2023). A common concern identified in these discussions is ChatGPT’s capacity to generate ‘human-like’ essays that can evade plagiarism detectors (e.g., Ventayen, 2023; Yeadon et al., 2022) and its potential use for academic misconduct during online exams (Susnjak, 2022). Indeed, recent research suggests that GPT-3.5 can generate 300-word essays comparable to second year Physics students (Yeadon et al., 2022) and ChatGPT can generate correct answers for 42-64% of questions drawn from question banks used to prepare for the United States Medical Licensing Examination (Gilson et al., 2022; see also Bommarito & Katz, 2022, for related work on GPT-3.5 and multiple-choice questions provided by the USA National Conference of Bar Examiners).

Technological developments that challenge the integrity of written assessments are not unprecedented. For example, within the literature on automated essay scoring (AES; see Bai, Zawacki-Richter, Bozkurt et al., 2022, for review), Perelman and colleagues’ work on the Basic Automatic BS Essay Language (BABEL) Generator demonstrates how synthetic text can be generated to contain features often associated with high-scoring essays (see Perelman, 2020, for summary). However, while BABEL-generated text is deliberately nonsensical, ChatGPT can generate text that is difficult for humans to distinguish from that written by other humans (see Ippolito, Duckworth, Callison-Burch, & Eck, 2020, for related research with GPT-2).

The potential for AI-generated text to evade both human and automated systems for detecting plagiarism raises questions not only about the feasibility of using written work for assessment, but also about the purpose and function of essays (see Rudolph et al., 2023; Tate et al., 2023, Yeadon et al., 2022, for discussions). As Rudolph et al. (2023) note, in a potential future where students can submit AI-generated essays that are then automatically screened and scored by other AI tools, “a first AI circumvents a second AI and is assessed by a third AI. All that the humans do is press a couple of keys, and nobody learns anything” (p. 13). This perspective is an apt reminder that essays are not only assigned for the purpose of assessment, but also serve an important role in learning. Indeed, writing itself can be conceptualized as an opportunity to exercise and refine critical thinking skills (see Flower & Hayes, 1981; Kurfiss, 1988; for related discussions). How the practice of writing and its relation to critical thinking will change in the future depend on how tools like

ChatGPT are applied and what affordances these technologies offer for students and teachers (see Kasneci et al., 2023; Rudolph et al., 2023, for discussions).

AN EXAMPLE ESSAY GENERATED BY ChatGPT

We ran a brief probe of ChatGPT's capabilities with the prompt "Write a 300-word essay on the ethics of AI with 3 references in APA format". The output was a vague but passable essay, logically structured with three body paragraphs between an introduction and a conclusion (Appendix A). The content identified some relevant ethical concerns, including the potential for malicious uses of AI (see e.g., McGuffie & Newhouse, 2020), the risk of entrenching existing inequalities (see e.g., Hu et al., 2019; IEEE, 2019), and the implications of AI on human labor (see e.g., Frey & Osborne, 2013; but also, Coelli & Borland, 2019). However, the essay lacked in-text citations and all three references generated by ChatGPT contained inaccuracies. Other authors have previously observed that ChatGPT-generated references tend to be inaccurate or fictitious (e.g., Tate et al., 2023). Likewise, we found that "The ethics of artificial intelligence", a highly cited chapter first published by Bostrom and Yudkowsky in 2014 was misattributed to "Green, T. R.", and "The ethics of algorithms" does not seem to exist but was nonetheless attributed to "Sunstein, C. R.". These author names may refer to Timothy R. Green and Cass R. Sunstein, both prolific and highly cited academics in the fields of agriculture and law, respectively.

The misattribution of authors and the 'stitched-together' composition of the references highlight an important facet of language models: that they do not (always) faithfully reproduce strings in the training data (cf. Carlini et al., 2021). Instead, language models predict sequences of words probabilistically based on patterns in the training data (Bender et al., 2021; Floridi & Chiriatti, 2020). Thus, it is possible that strings (e.g., authors, titles, and publishers) that appear more frequently in the training data are more likely to be represented in ChatGPT's output (see Wei, Garrette, Linzen, & Pavlick, 2021, for related work). However, such a conclusion remains speculative and requires further analysis of ChatGPT-generated references.

From an educator's perspective, learning to choose appropriate references and integrating ideas from different sources are important skills particularly in higher education. The ability to identify who stated what, how they stated it, and on what basis they stated it are all important components for engaging critically with research literature and tracking the development of different ideas. Thus, more work is needed to address whether these processes can or should be 'offloaded' to tools like ChatGPT and to explore the potential consequences of such offloading (see Risko & Gilbert, 2016).

From a broader perspective, amongst papers that do not focus solely on technical capabilities and potential applications, a common concern raised about technological development is what effect these technologies may have on people (Bender et al., 2021; Prabhakaran, Mitchell, Gebru, & Gabriel, 2022; Tuomi, 2018). In the case of language models, Floridi and Chiriatti (2020) note that their availability "represents the arrival of a new age in which we can now mass produce good and cheap semantic artefacts" (p. 690). These "cheap semantic artefacts" may be used dishonestly (Susnjak, 2022) or maliciously (McGuffie & Newhouse, 2020) to change the behavior of their human readers, or they may be used as learning opportunities to engage in critical thinking (Rudolph et al., 2023). Regardless of how they are used, the value of semantic artefacts depends the value assigned to them by humans. As Bender et al. (2021) argue, "the human tendency to attribute meaning to text" (p. 618), combined with the ability of language models to reproduce biases in their training data, means that uncritical reading of synthetic text can further perpetuate inaccurate and harmful worldviews (see McGuffie & Newhouse, 2020, for a salient example). Therefore, it is vital that education sectors prepare students with the skills to engage critically with a world filled with both human- and AI-generated media.

THE ACTIVE ROLE OF STUDENTS IN AIEd

The discussions of how students may use or "misuse" generative-AI tools highlight the importance of student participation in the development of AIEd. In our previous work, we argued that "learners and teachers, the stakeholders who stand to be most affected by AIEd integration, need to be active participants in multi-

stakeholder discussions about how AI is integrated in education.” (Bai et al., 2022, p. 62). However, due to our initial goal of gathering educator perspectives, we did not discuss the role of students as an equally important strategy element (cf. Table 1). Indeed, multiple participants in preliminary focus-group discussions have highlighted this gap by raising the issue of student agency. Thus, this final section addresses the active role of students for the future of AIED.

Research on the motivations of students engaging in plagiarism and academic misconduct is not new (see e.g., Bennett, 2005; Parnter, 2022). Indeed, Eaton, Stoesz, and Seeland (2022) have traced the industry of commercial contract cheating back to the 1930s. However, the availability of tools like ChatGPT makes the means of producing essays widely accessible and lowers the costs of time, money, and effort involved. As Yeadon et al. (2022) note “these tools essentially represent a wide spread democratization of paid essay writing services” (p. 9). Therefore, it seems pertinent to consider the value proposition for students tasked with written assignments. Specifically, what purpose does and should essay writing fulfill? If written assignments serve primarily as an assessment tool, students focused on extrinsic goals (see Krou, Fong, & Hoff, 2021) may perceive essay writing simply as an obstacle to obtaining good grades. In contrast, if essays serve as opportunities for students to exercise critical thinking and receive meaningful feedback, students may perceive writing to be an important skill to practice. Ultimately, if we want to understand how students will use ChatGPT and other AI tools, we need to engage students in participatory research (Prabhakaran et al., 2022) and allow them space to share their perspectives. Nevertheless, it is up to institutions and educators to make the value of writing explicit and to engage students in collaborative relationships (see Macfarlane & Tomlinson, 2017, for a critical discussion).

The release of ChatGPT has led to very different responses across different institutions (Rudolph et al., 2023) with some calling for further development in AI tools capable of detecting synthetic text (Susnjak, 2022). However, such an adversarial approach may result in a ‘cat-and-mouse game’ between students and institutions, in which learning is relegated below the goal of countering the latest tool used by the other side (see Perelman, 2020). The legal challenge that resulted in the *Garante* (2021) decision provides just one example of such an outcome. AlgorithmWatch (2020) report another case in which, despite student protests, the Danish Ministry of Education implemented an exam monitoring software for online high-school exams. In response, a student gained access to the underlying code and shared details for bypassing the system (Schou, 2019). While acknowledging the importance of ensuring assessment integrity, these challenges to institutional decisions demonstrate students critically engaging with their world (hooks, 2010) and highlight the importance of consulting with students before implementing AIED tools.

Institutional decisions regarding AI tools also have implications for the relationship between students and the teachers tasked with assessing their learning. Adversarial and ‘policing’ approaches could degrade trust and disrupt the practices of teaching and learning. In contrast, Rudolph et al. (2023) makes a compelling case for “an approach that builds trusting relationships with our students in a student-centric pedagogy and assessments for and as learning rather than solely assessments of learning” (p. 13). While the most suitable approach is likely to differ across different contexts and cultures, the cases of generative AI and digital-proctoring tools considered above highlight the importance of actively involving students in the integration of AI in education.

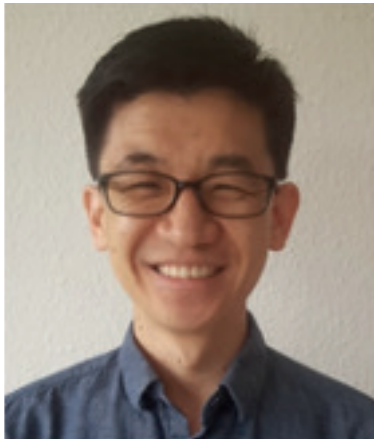
CONCLUSION

In the present project, we use future-planning methods to explore the opportunities and risks of integrating AI in education (see Bai et al., 2022). However, such an approach relies on ongoing engagement with a diverse range of perspectives (Prabhakaran et al., 2022). We concluded our previous work by acknowledging that “the visions of possible futures presented here are neither complete nor objective” (p. 63). While the present paper re-examines and extends our discussion in light of the *Garante* (2021) decision and the release of ChatGPT, our original conclusion remains true. Our visions of possible futures remain incomplete and the collaborative activities involved in shaping the future of AIED require ongoing, multi-stakeholder dialogue between legal scholars (e.g., Bincoletto, 2021), computer scientists and AI ethicists (e.g., Bender et al., 2021; Prabhakaran et al., 2022), educators (Zawacki-Richter et al., 2019), and students.

Authors' Note: This article has been presented in the 5th International Open & Distance Learning Conference- IODL 2022.

Acknowledgements: This project is funded by the Volkswagen Foundation and the Ministry of Science and Culture in Lower Saxony (2N3743). We thank the team of international researchers at the Center for Open Education Research (COER) for insightful discussions that shaped the elements in Table 1; particular thanks to Victoria I. Marin, Cengiz Hakan Aydin, Aras Bozkurt, Insung Jung, Junhong Xiao, and Patricia Slagter van Tryon for their active collaboration. We also thank Sreekari Vogeti and Laura N. Peters for their valuable comments on early drafts of this paper.

BIODATA and CONTACT ADDRESSES of AUTHORS



Dr. John Y. H. BAI is a Postdoctoral Researcher in the Faculty of Education and Social Sciences at the Carl von Ossietzky University of Oldenburg. He completed his PhD in Experimental Psychology at the University of Auckland, New Zealand, where he worked as a Research Fellow and Lecturer. In collaboration with the international team at the Center for Open Education Research (COER), he is currently working on the project “Prospects for the Future of Learning: Artificial Intelligence Applications in Higher Education”, funded by the Volkswagen Foundation and The Lower Saxony Ministry of Science and Culture. This international project aims to collect perspectives of higher education faculty to contribute to the interdisciplinary and multi-stakeholder discussion on AI and society. His research interests include artificial intelligence, research design, and the interaction between individuals and their environments.

John Y. H. BAI
Center for Open Education Research (COER), Faculty of Education and Social Sciences
Address: Carl von Ossietzky Universität Oldenburg, 26129, Oldenburg, Germany
Phone: +49-(0)441-798-0
E-mail: john.yihao.bai@uni-oldenburg.de



Dr. Olaf ZAWACKI-RICHTER is a professor of educational technology at the University of Oldenburg in Germany. He is the Dean of the Faculty of Education and Social Sciences and Director of the Center for Open Education Research (COER). Olaf has over 25 years of professional experience in the field of open, distance, and digital education. He has also served as a consultant and advisor, including work for the United Nations' International Labor Organization, the Office of Technology Assessment at the German Bundestag, and the German Science and Humanities Council (Wissenschaftsrat). Dr. Zawacki-Richter has authored over 150 journal articles and edited several books, including “Online Distance Education – Towards a Research Agenda”, “Systematic Reviews in Educational Research”, “Open and Distance Education in Australia, Europe and the Americas: National Perspectives in a Digital

Age” (Vol. 1), and “Open and Distance Education in Asia, Africa and the Middle East: National Perspectives in a Digital Age” (Vol. 2) – all published open access. He is an Associate Editor of the “Online Learning Journal” (OLJ) and a member of the editorial board of the “International Review of Research in Open and Distance Learning” (IRRODL), “Open Learning”, the “Turkish Online Journal of Distance Education” (Turkiye), and the “The Journal for Higher Education Development” (Austria). In 2022 Olaf published together with Professor Insung Jung the open-access Handbook of Open, Distance, and Digital Education (ODDE) in Springer's Major Reference Series.

Olaf ZAWACKI-RICHTER

Center for Open Education Research (COER), Faculty of Education and Social Sciences

Address: Carl von Ossietzky Universität Oldenburg, 26129, Oldenburg, Germany

Phone: +49-(0)441-798-0

E-mail: olaf.zawacki.richter@uni-oldenburg.de



Dr. Wolfgang MUSKENS has worked as a researcher at the Carl von Ossietzky University of Oldenburg, Faculty of Educational and Social Sciences since 2001. Wolfgang has worked in the field of Recognition of Prior Learning for more than 15 years, and has been involved in many projects on APL (Accreditation of Prior Learning), RPL, and PLAR (Prior Learning Assessment and Recognition). He contributed to the establishment of Germany's first PLAR-service at the University of Oldenburg in 2017. His academic background is a Doctorate in personality psychology and a habilitation in the field of Lifelong Learning. He studied at the Universities of Dusseldorf and Koblenz-Landau, where he earned his degree in 2001. With regards to methods, he is interested in structure-giving techniques like Latent Class Analysis and Factor Analysis.

Wolfgang MUSKENS

Center for Open Education Research (COER), Faculty of Education and Social Sciences

Address: Carl von Ossietzky Universität Oldenburg, 26129, Oldenburg, Germany

Phone: +49-(0)441-798-0

E-mail: wolfgang.mueskens@uni-oldenburg.de

REFERENCES

- AlgorithmWatch. (2020). *Automating society report 2020*. <https://automatingsociety.algorithmwatch.org>
- Bai, J. Y. H., Zawacki-Richter, O., Bozkurt, A., Lee, K., Fanguy, M., Cefa Sari, B., & Marin, V. I. (2022). Automated essay scoring (AES) systems: Opportunities and challenges for open and distance education. In *Proceedings of The Tenth Pan-Commonwealth Forum on Open Learning (PCF10)*. <https://doi.org/10.56059/pcf10.8339>
- Bai, J. Y. H., Zawacki-Richter, O., & Muskens, W. (2022). Developing strategic scenarios for artificial intelligence applications in higher education. In *5th International Open and Distance Learning Conference Proceedings Book* (pp. 47-70). Eskisehir, Turkey. <https://iodl.anadolu.edu.tr/proceedings-iodl-2022>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). <https://doi.org/10.1145/3442188.3445922>
- Bennett, R. (2005). Factors associated with student plagiarism in a post-1992 university. *Assessment & Evaluation in Higher Education*, 30, 137-162. <https://doi.org/10.1080/0260293042000264244>
- Bincoletto, G. (2021). E-proctoring during students' exams: emergency remote teaching at stake. *European Data Protection Law Review*, 7, 586-591. <https://doi.org/10.21552/edpl/2021/4/15>
- Bommarito II, M., & Katz, D. M. (2022). *GPT takes the Bar Exam*. arXiv preprint. <https://doi.org/10.48550/arXiv.2212.14402>
- Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In K. Frankish & W. Ramsey (Eds.), *The Cambridge Handbook of Artificial Intelligence* (pp. 316-334). Cambridge University Press. <https://doi.org/10.1017/CBO9781139046855.020>

- Bozkurt, A., Karadeniz, A., Baneres, D., Guerrero-Roldan, & A.E., Rodriguez, M. E. (2021). Artificial intelligence and reflections from educational landscape: A review of AI studies in half a century. *Sustainability*, 13, Article 800. <https://doi.org/10.3390/su13020800>
- Brockman, G., Sutskever, I., & OpenAI. (2019, March 11). *OpenAI LP*. OpenAI Blog. <https://openai.com/blog/openai-lp/>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J, Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901. <https://doi.org/10.48550/arXiv.2005.14165>
- Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., ... & Raffel, C. (2021, August 11-13). Extracting Training Data from Large Language Models. In *The Proceedings of the 30th USENIX Security Symposium* (pp. 2633- 2650). <https://www.usenix.org/system/files/sec21-carlini-extracting.pdf>
- Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the ‘good society’: The US, EU, and UK approach. *Science and Engineering Ethics*, 24, 505-528. <https://doi.org/10.1007/s11948-017-9901-7>
- Coelli, M. B., & Borland, J. (2019). Behind the headline number: *Why not to rely on Frey and Osborne’s predictions of potential job loss from automation*. Melbourne Institute Applied Economic & Social Research. https://melbourneinstitute.unimelb.edu.au/__data/assets/pdf_file/0005/3197111/wp2019n10.pdf
- Commission on the Ethics of Scientific Knowledge and Technology (2019). *Preliminary study on the ethics of artificial intelligence*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000367823>
- du Boulay, B. (2022). Artificial Intelligence in Education and Ethics. In O. Zawacki-Richter & I. Jung (Eds.), *Handbook of Open, Distance and Digital Education*. Springer Nature Singapore. https://doi.org/10.1007/978-981-19-0351-9_6-1
- Eaton, S.E., Stoesz, B.M., Seeland, J. (2022). Listening to Ghosts: A qualitative study of narratives from contract cheating writers from the 1930s onwards. In S. E. Eaton, G. J. Curtis, B. M. Stoesz, J. Clare, K. Rundle, & J. Seeland (Eds.), *Contract Cheating in Higher Education: Global Perspectives on Theory, Practice, and Policy* (pp. 251-269). Springer International Publishing. https://doi.org/10.1007/978-3-031-12680-2_18
- Edwards, L., & Veale, M. (2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke Law & Technology Review*, 16, 18-84. <https://doi.org/10.2139/ssrn.2972855>
- European Data Protection Board (2021, September 16). *Italian SA finds monitoring system for online university exams to be in breach of privacy, fines university*. https://edpb.europa.eu/news/news/2021/italian-sa-finds-monitoring-system-online-university-exams-be-breach-privacy-fines_en
- Facer, K., & Selwyn, N. (2021). *Digital technology and the futures of education: Towards ‘Non-Stupid’ optimism*. Paper commissioned for the UNESCO Futures of Education report. <https://unesdoc.unesco.org/ark:/48223/pf0000377071.locale=en>
- Fink, A., & Siebe, A. (2011). *Handbuch Zukunftsmanagement: Werkzeuge der strategischen Planung und Fruherkennung*. Campus Verlag GmbH.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30, 681-694. <https://doi.org/10.1007/s11023-020-09548-1>
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32, 365-387.
- Frey, C. B., & Osborne, M. (2013). *The future of employment*. Oxford Martin Programme on Technology and Employment. <https://www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf>

- Garante per la protezione dei dati personali. (2021). *Ordinanza ingiunzione nei confronti di Università Commerciale “Luigi Bocconi” di Milano - 16 settembre 2021 [9703988]*. <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9703988>
- General Data Protection and Regulation. (Regulation 2016/679). European Parliament, Council of the European Union. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- Gilson, A., Safranek, C., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2022). *How well does ChatGPT do when taking the medical licensing exams? The implications of large language models for medical education and knowledge assessment*. medRxiv preprint. <https://doi.org/10.1101/2022.12.23.22283901>
- Gutschow, K., & Jorgens, J. (2016). *Einführung von Verfahren zur Validierung nichtformalen und informellen Lernens – Anforderungen und Handlungsoptionen Zwischenbericht*. Federal Institute for Vocational Education and Training.
- Gutschow, K., & Jorgens, J. (2019). *Einführung von Verfahren zur Validierung nichtformalen und informellen Lernens – Anforderungen und Handlungsoptionen*. Federal Institute for Vocational Education and Training.
- hooks, b. (2010). *Teaching critical thinking: Practical wisdom*. Routledge. <https://doi.org/10.4324/9780203869192>
- Hu, X., Neupane, B., Echaiz, L. F., Sibal, P., & Rivera Lam, M. (2019). *Steering AI and advanced ICTs for knowledge societies: A Rights, Openness, Access, and Multi-stakeholder Perspective*. UNESCO Publishing. <https://unesdoc.unesco.org/ark:/48223/pf0000372132>
- Human Rights Watch. (2022). *“How dare they peep into my private life?” Children’s rights violations by governments that endorsed online learning during the Covid-19 pandemic*. <https://www.hrw.org/report/2022/05/25/how-dare-they-peep-my-private-life/childrens-rights-violations-governments>
- IEEE. (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems* (1st Ed.). The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. <https://standards.ieee.org/content/ieeestandards/en/industry-connections/ec/autonomous-systems.html>
- Ippolito, D., Duckworth, D., Callison-Burch, C., & Eck, D. (2019). *Automatic detection of generated text is easiest when humans are fooled*. arXiv preprint. <https://doi.org/10.48550/arXiv.1911.00650>
- Kasneji, E., Sessler, K., Kuchemann, S., Bannert, M., Dementieva, D., Fishcher, F., ... & Kasneji, G. (2023). *ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education*. edArXiv preprint. <https://doi.org/10.35542/osf.io/5er8f>
- Krou, M. R., Fong, C. J., & Hoff, M. A. (2021). Achievement motivation and academic dishonesty: A meta-analytic investigation. *Educational Psychology Review*, 33, 427-458. <https://doi.org/10.1007/s10648-020-09557-7>
- Kurfiss, J. G. (1988). *Critical Thinking: Theory, Research, Practice, and Possibilities*. ASHE-ERIC Higher Education Report No. 2. George Washington University. <https://files.eric.ed.gov/fulltext/ED304041.pdf>
- Laird, E., Grant-Chapman, H., Venzke, C., & Quay-de la Vallee, H. (2022). *Hidden harms: The misleading promise of monitoring students online*. The Center for Democracy & Technology. <https://cdt.org/wp-content/uploads/2022/08/Hidden-Harms-The-Misleading-Promiseof-Monitoring-Students-Online-Research-Report-Final-Accessible.pdf>
- Lambert, N. & von Werra, L. (2022, December 9). *Illustrating Reinforcement Learning from Human Feedback (RLHF)*. Hugging Face Blog. <https://huggingface.co/blog/rlhf>
- Liguori, L. & Petrucci, L. (2021, October, 21). *The Italian Data Protection Authority fines Italian university Bocconi for unlawful processing of students’ personal data*. Portolano Cavallo. <https://portolano.it/en/newsletter/portolano-cavallo-inform-digital-ip/the-italian-data-protection-authority-fines-italian-university-bocconi-for-unlawful-processing-of-students-personal-data>

- Macfarlane, B., & Tomlinson, M. (2017). Critiques of student engagement. *Higher Education Policy*, 30, 5-21. <https://doi.org/10.1057/s41307-016-0027-3>
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*, 27, 12-14.
- McGuffie, K., & Newhouse, A. (2020). *The radicalization risks of GPT-3 and advanced neural language models*. arXiv preprint. <https://doi.org/10.48550/arXiv.2009.06807>
- Nilsson, N. J. (2009). *The quest for artificial intelligence*. Cambridge University Press.
- OECD. (2022). *Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- OpenAI (2022, November 30). *ChatGPT: Optimizing language models for dialogue*. OpenAI Blog. <https://openai.com/blog/chatgpt/>
- Parther, C. (2022). The rise of contract cheating in graduate education. In S. E. Eaton, G. J. Curtis, B. M. Stoesz, J. Clare, K. Rundle, & J. Seeland (Eds.), *Contract Cheating in Higher Education: Global Perspectives on Theory, Practice, and Policy* (pp. 251-269). Springer International Publishing. https://doi.org/10.1007/978-3-031-12680-2_17
- Perelman, L. (2020). The BABEL generator and e-Rater: 21st Century writing constructs and automated essay scoring (AES). *Journal of Writing Assessment*, 13. <https://escholarship.org/uc/item/263565cq>
- Prabhakaran, V., Mitchell, M., Gebru, T., & Gabriel, I. (2022). *A human rights-based approach to responsible AI*. arXiv preprint. <https://doi.org/10.48550/arXiv.2210.02667>
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20, 676-688. <https://doi.org/10.1016/j.tics.2016.07.002>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?. *Journal of Applied Learning and Teaching*, 6. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Russell, N. C., Reidenberg, J. R., Martin, E., & Norton, T. B. (2018). *Transparency and the marketplace for student data*. Center on Law and Information Policy, Fordham Law School. <https://ir.lawnet.fordham.edu/clip/4>
- Sanh, V., Webson, A., Raffel, C., Bach, S. H., Sutawika, L., Alyafeai, Z., ... & Rush, A. M. (2021). *Multitask prompted training enables zero-shot task generalization*. arXiv preprint. <https://doi.org/10.48550/arXiv.2110.08207>
- Schou, C. (2019, March 9). *The nadir of surveillance (Den Digitale Provevagt)*. <https://secret.club/2019/03/07/exam-surveillance.html>
- Susnjak, T. (2022). *ChatGPT: The end of online exam integrity?*. arXiv preprint. <https://doi.org/10.48550/arXiv.2212.09292>
- Tate, T., Doroudi, S., Ritchie, D., & Xu, Y. (2023). *Educational research and AI-generated writing: Confronting the coming tsunami*. EdArXiv preprint. <https://doi.org/10.35542/osf.io/4mec3>
- Tuomi, I. (2018). *The impact of artificial intelligence on learning, teaching, and education*. European Commission, Joint Research Centre. <https://doi.org/10.2760/12297>
- UNESCO (2015) *Qingdao Declaration ICT- and Post-2015 Education. Seize digital opportunities. Lead education transformation*. United Nations Educational, Scientific and Cultural Organization. https://en.unesco.org/icted/sites/default/files/2019-04/37_qingdao_declaration.pdf
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- Ventayen, R. J. M. (2023). *OpenAI ChatGPT Generated Results: Similarity Index of Artificial Intelligence-Based Contents*. <http://doi.org/10.2139/ssrn.4332664>

- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, 7, 76-99. <https://doi.org/10.1093/idpl/ix005>
- Wei, J., Garrette, D., Linzen, T., & Pavlick, E. (2021). *Frequency effects on syntactic rule learning in transformers*. arXiv preprint. <https://doi.org/10.48550/arXiv.2109.07020>
- Yeadon, W., Inyang, O. O., Mizouri, A., Peach, A., & Testrow, C. (2022). *The death of the short-form physics essay in the coming ai revolution*. arXiv preprint. <https://doi.org/10.48550/arXiv.2212.11661>
- Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 1-27. <https://doi.org/10.1186/s41239-019-0171-0>
- Zeide, E. (2019). Artificial intelligence in higher education: Applications, promise and perils, and ethical questions. *Educause Review*, 54, 31-39. <https://er.educause.edu/articles/2019/8/artificial-intelligence-in-higher-education-applications-promise-and-perils-and-ethical-questions>

APPENDIX

ChatGPT output following the prompt “Write a 300-word essay on the ethics of AI with 3 references in APA format” (italics added):

The Ethics of AI is a growing area of concern as artificial intelligence continues to develop and become more prevalent in our society. AI has the potential to change the way we live, work, and interact with the world in profound ways, but it also raises a number of ethical questions about its impact on our lives and the way it is used.

One of the main concerns about AI is that it has the potential to be used for malicious purposes, such as cyber attacks, hacking, and mass surveillance. As AI becomes more advanced, it will become easier for individuals or organizations to use it to carry out harmful activities without detection. This is why it is important to ensure that AI is developed and used in an ethical and responsible manner, with strong privacy and security measures in place to prevent abuse.

Another area of concern is that AI has the potential to be biased and perpetuate existing inequalities in society. For example, if an AI system is trained on biased data, it may make biased decisions that discriminate against certain groups of people. It is crucial to ensure that AI is developed in a way that is transparent and accountable, and that it is tested and validated to avoid perpetuating existing biases.

Lastly, there are also concerns about the potential job losses and other economic impacts that may result from the widespread adoption of AI. While AI has the potential to greatly increase productivity and efficiency, it may also lead to widespread job losses as machines replace human workers. This is why it is important to think about the ethical implications of AI and ensure that its benefits are distributed fairly.

In conclusion, the ethics of AI is a complex and important area of discussion. To ensure that AI is developed and used in an ethical and responsible manner, it is crucial to consider its potential impacts on privacy, security, and the distribution of its benefits. By considering these ethical questions and working to develop AI in an ethical and responsible way, we can ensure that its benefits are realized while minimizing its potential harm.

References:

Green, T. R. (2018). The ethics of artificial intelligence. Cambridge University Press.

Zhang, H., & Chen, Y. (2019). Artificial intelligence: A comprehensive review and ethical considerations. Journal of medical systems, 43(12), 592.

Sunstein, C. R. (2020). The ethics of algorithms. Cambridge University Press.