# Toward Responsible, Human-Centered AI in EdTech

**Conference Paper** · December 2020

CITATIONS
0

READS
248

**2 authors:**

André Renz
Universität der Künste Berlin
**9** PUBLICATIONS **34** CITATIONS
SEE PROFILE

Swathi Krishnaraja
Universität des Saarlandes
**3** PUBLICATIONS **10** CITATIONS
SEE PROFILE

This paper was presented at ISPIM Connects Global 2020: Celebrating the World of Innovation - Virtual, 6-8 December 2020.
Event Proceedings: LUT Scientific and Expertise Publications: ISBN 978-952-335-566-8

# Toward Responsible, Human-Centered AI in EdTech

André Renz*

University of the Arts Berlin/Weizenbaum Institute for the Networked Society, Hardenbergstraße 32, 10623 Berlin, Germany.
Email: a.renz@udk-berlin.de

Swathi Krishnaraja

University of the Arts Berlin/Weizenbaum Institute for the Networked Society, Hardenbergstraße 32, 10623 Berlin, Germany.
Email: s.swathi-krishnaraja@udk-berlin.de
* Corresponding author

**Abstract:** The increasing relevance of artificial intelligence (AI) applications in various domains has led to high expectations of benefits ranging from precision, efficiency, and optimization to the completion of routine or time-consuming tasks. Particularly in the field of education, there exist a multitude of yet-to-be-answered questions and challenges: How can the privacy of students be maintained? What will be AI's long-term effects on teachers' roles? What are the ethical and social consequences? In this paper, we introduce the concept of human-centered AI, rethinking how an AI system can be developed in line with human values and without posing risks to humanity. As the education market is in the early stages of incorporating AI into educational tools, we believe that this is the right time to create awareness about the use of principles that foster human-centered values and help in building a responsible, ethical, and value-oriented AI.

**Keywords:** education; artificial intelligence; educational technology; innovation; intelligent tutoring systems; human-centered AI; design for value approach

## 1 Introduction

The relevance of artificial intelligence (AI)-supported systems in education, so-called AI in education (AIED), has increased dramatically in recent years, arousing great expectations and offering huge innovation potential for the entire education market (Holmes et al., 2019; EdTechXGlobal, 2016). AI significantly expands traditional practices in education, and new digital solutions are emerging (i.e., intelligent and adaptive educational tools), which are gaining a market share alongside traditional concepts. Basically, such intelligent tools can be categorized into two main areas: narrow AI/weak AI and general AI/strong AI, where the former refers to the intelligent agent that is designed to solve one specific task, whereas the latter refers to an agent that is capable of solving any given problem irrespective of their task or domain. Almost all the available educational tools are a composition of narrow and general AI, whereas a solely built general/strong AI is unlikely to exist in the future (Zawacki-Richter et al., 2019).

Event Proceedings: LUT Scientific and Expertise Publications: ISBN 978-952-335-566-8

As the outcomes of these tools strongly rely on the data behind the task or domain, they impact humans in a number of ways. For example, there are concerns surrounding the usage of private information, such as learner behaviors, abilities, and mental states while performing educational activities (Holmes et al., 2018). Therefore, there is an increased need to address the technological and societal implications associated with the emergence and usage of AIED tools. There is an ongoing discourse on how to operationalize these values during the development of AI systems, rather than applying these rules/guidelines after the deployment. In this paper, we introduce the design-for-values approach—a methodological approach aimed at incorporating moral values as part of the technological design, research, and development (Dignum, 2019). The development of AI systems entails processes, such as identifying societal values, deciding on a moral deliberation approach, and linking values to the formal system requirements and concrete functionalities (Dignum, 2019). The questions that this research endeavors to answer are which societal issues are associated with the digitization of education through AIED tools, and what changes need to be made to these tools to cause humans to accept them as useful and trustworthy.

In this contribution, we focus on how responsible AIED tools can be developed and operationalized. We introduce the aspects of value-centered, human-centered, ethical, and responsible AI in the domain of education, which to our knowledge, have been underexplored. In the following section, we briefly outline the current market developments of AIED. Afterwards, we discuss the current AI applications in the EdTech area. The fourth section introduces the idea of an integrative human-centered AI (HCAI) approach for the development of AI-based EdTech applications. The article concludes with recommendations for action and an outlook on future research extensions.

## 2 Market development of AIED

The implementation of AI technologies holds high potential for innovation in a number of different fields. In the education sector, providers are entering the market in increasing numbers with the promise of offering intelligent learning solutions through data-based and AI-driven applications. According to Renz and Hilbig (2020), such systems are currently the highest level of EdTech applications, which enable the individualization of learning paths. Although AI-based EdTech applications offer high innovation potential for the business models of providers and users/learners, there are hardly any EdTech companies (at least in the German market) already working with such technology. Most of the analyzed companies are in the range of low-data and data-enhanced business models.

In a theoretical analysis, Renz et al. (2020a) also show that the innovation potential for the use of AI-based elements in education already exists, but they often only have a subjunctive character and can barely claim practical evidence. A worldwide survey of stakeholders from the education sector shows that already 20% of the surveyed EdTech companies have invested in and implemented AI technologies and another 21% are currently testing AI technologies for their businesses (Global Executive Panel, 2019). In addition to the emerging innovation dynamic of EdTech companies, the COVID-19 pandemic can be currently seen as a tipping point for faster market development (Renz et al., 2020b; Renz, 2020). In a market analysis of two AI-driven EdTech applications—language learning platforms (LLP) and learning management systems (LMS)—Renz et al. (2020b) demonstrate that the COVID-19 pandemic has already caused a market shift from low-data business models to data-enhanced business models. The authors assume that the significant

increase in the use of EdTech applications during the crisis will also lead to more data-driven EdTech applications entering the market. The increasing number of users of corresponding EdTech applications is leading to the generation of more data relating to learning behavior. Such data provide a basis for the further development of AI-based learning systems.

Additionally, there are adaptive learning solutions on the market that work on the principles of rules. Holmes et al. (2019) note that science, technology, engineering and mathematics (STEM) subjects in particular play an important role in the development of such AIEDs. Among the most common AIED applications are intelligent tutoring systems (ITS), which allow individualized learning paths with step-by-step tutorials (Alkhatlan & Kalita, 2018). One reason why STEM subjects are especially low-hanging fruits for ITS applications is the existence of clearly defined rules and a well-structured approach (Holmes et al., 2019). Hence, the discourse shows that EdTech companies are preparing for the development and use of AI technologies and will further accelerate the existing innovation dynamics in the education market in the coming years.

## 3 Current AI applications in education

Ahmad et al. (2020) present a bibliometric analysis of AI applications in education. The authors divide the field of AI applications in education into ITS, evaluation, personalized learning, recommender systems, student performance, sentiment analysis, retention and dropout, and classroom monitoring. Another overview of current AI applications in education can be found in Holmes et al. (2019). The authors classify four main types of application areas for AIED: ITS, dialogue-based tutoring systems (DBTS), explorative learning environments (ELE), and automatic writing assessment (AWE). The following chart summarizes some of the most popular EdTech providers according to the classification of Holmes et al. (2019).

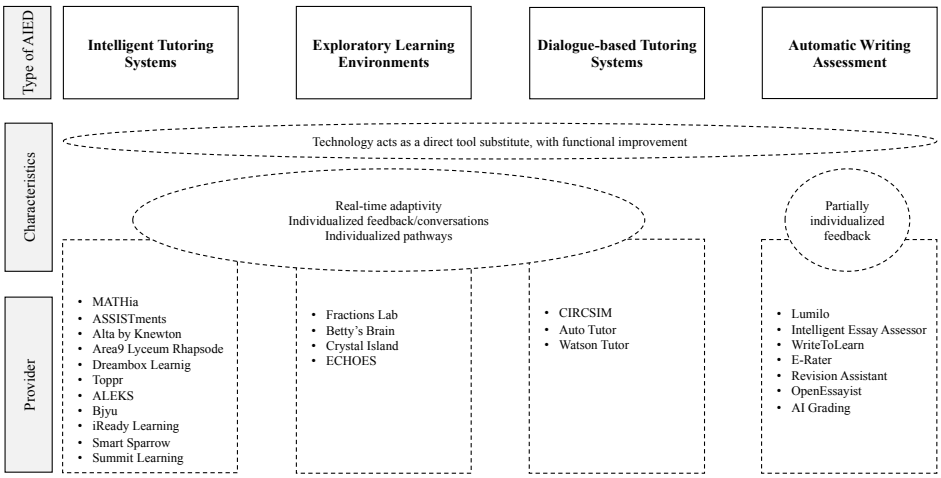


**Figure 1** Overview of current AIED applications (Holmes et al., 2019)

Whether an intelligent learning system operates on the basis of data on individual learning behavior or whether it is rule-based is not always transparent to the user. Nevertheless, it can be expected that in the coming years, more and more EdTech applications will be

developed on the basis of AIED. Therefore, it is essential to establish appropriate regulations that consider the responsible and sustainable development of such applications. In the following, we present human-centered AI as a possible approach that promises responsible implementation of AI in education and educational products.

## 4 Human-centered AI in education

As AI is gradually being adopted in education for the purpose of teaching and learning, there is a persistent debate on the educational value of technology (Luckin & Cukurova, 2019). As a result of this uncertainty, the progress of AI technology and learning analytics (LA) in education lags far behind when compared to other domains, such as healthcare and finance. The fear of AI completely rejecting the role of teachers is a major concern for both teachers and educational institutions (Popenici & Kerr, 2017). It is merely due to this fact that many research projects, such as the European Humane AI project[1], Stanford Institute for Human-Centered Artificial Intelligence[2], and other research institutes, such as MIT and UC Berkeley (Xu, 2019), took initiatives to work toward understanding the human aspects of AI and develop a more responsible AI that enhances the capabilities of humans rather than replacing them. Although there is no concrete framing of the definition of HCAI, the general understanding of HCAI is a design thinking approach that puts humans at the center of AI development, rather than thinking of AI automation as a removal of human agency and control. Furthermore, HCAI "is designed with a clear purpose for human benefit while being transparent about who has control over the data and algorithms" (Schmidt, 2020, p. 2). Schneiderman (2020) reframes AI as a replacement of algorithms and AI systems with humans at the center, thereby calling HCAI a second Copernican revolution.

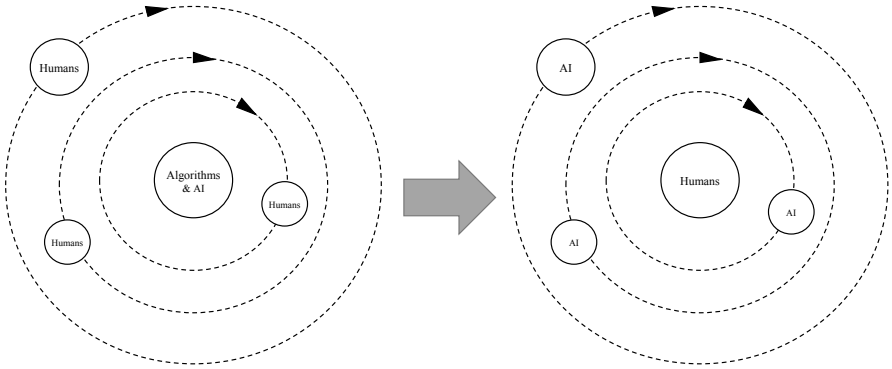


**Figure 2** The Transformative Power of HCAI (Schneiderman, 2020)

In this paper, we propose that working toward human-centered approaches in developing educational technologies will help both providers and consumers of EdTech solutions. Currently, AI systems in ITS enhance the technology by offering personalized lessons to students that are based on their learning patterns, their knowledge and interest in a particular area, etc., but the HCAI approach enhances human capabilities by allowing the teachers to build their own computerized lessons by using the insights gathered from the AI tutoring system (Weitekamp, 2020). To the best of our knowledge, there are no cases

[1] Humane AI | Human-Centered Artificial Intelligence
[2] Stanford HAI: Home

in the education sector of explicit consideration given to an HCAI approach when developing AIED. In general, it remains quite unclear which areas already use HCAI approaches. Nevertheless, there are some cases where the approaches to HCAI development in AI applications in different areas have changed to achieve better user experiences. One example of the implementation of HCAI approaches can be found in the healthcare sector. With the help of AI, possible tumors can be identified in X-rays. This enables radiologists to quickly focus on areas suggested by the AI and provide treatments for patients. This is an example of AI working for a radiologist and not necessarily making choices/decisions by itself (Dembrower, 2020). Another example of HCAI use can be found in the area of customer management. Some companies employ chatbots and digital agents to automate and streamline responses, which leads to a less-than-ideal customer experience. The HCAI approach allows the AI system to work for the call center agent to identify the right information and speed up the answering process by also providing better customer experiences (Forbes Insights, 2020).

As we already mentioned, one of the ways in which AI is being implemented in the field of education is through ITS. ITS aims at providing personalized instruction and feedback to students, often through AI technology and without a human teacher. This suggests that, traditionally, AI algorithms and systems have been developed with the notion of harnessing the efficiency of machine automation (by measuring algorithmic performance, confidence interval, etc.) and celebrating what the AI systems could do. In contrast, we propose to build EdTech solutions considering HCAI's design thinking approach, which centers human values by measuring human performance while remaining amenable to human feedback and agency and celebrating the new powers that people have (Schneiderman, 2020). A recent work by Weitekamp et al. (2020) posits newly developed methods involving AI technologies that allow a teacher to teach an AI system (precisely ITS) that, in turn, teaches students. In this method, a human teacher demonstrates to the computer how to solve specific problems, such as multi-column addition. If the computer provides the wrong solution to the problem, this indicates to the human teacher possible trouble spots for students. This authoring process helps teachers understand the trouble spots for students because the machine learning system also often stumbles at the same places that students do. As we are currently losing direction of whether to enhance machine capabilities or human capabilities, it is the right time to rethink how AI systems are being developed to satisfy educational purposes. Thus, HCAI is essential to ensure that AI solutions responsibly prioritize human values and human dignity.

Thus far, limited attention has been given to the role of human values during the development of AI technology. Therefore, various scientists are currently trying to develop design approaches that focus on human values (Auernhammer, 2020). Each of the different design approaches provides a valuable perspective when designing for people. One possible approach is value sensitive design (VSD)—a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner—that provides different perspectives on societal, diversity, interaction, and human needs in the design of computer systems, such as AI. The VSD approach provides an opportunity to research and examine the impact of AI on people through a specific lens.

Another solution for mitigating the described challenges is by following a design for values methodological approach. This approach aims at making moral values part of technological design and development (Dignum, 2019). Often, values are interpreted as high-level abstract concepts that are hard to operationalize into concrete functionalities. A design-for-values approach resolves this by proposing a way that puts human rights, human dignity,

Event Proceedings: LUT Scientific and Expertise Publications: ISBN 978-952-335-566-8

and human freedom at the center of AI design. Using this design-for-values approach to methodological design assists in building HCAI that helps in identifying societal values, deciding on a moral deliberation approach (through algorithms, user control, regulation), and thereby linking these values to formal system requirements and concrete functionalities (Dignum, 2019).

Xu (2019) proposes an extended HCAI framework (see Figure 3) that includes three main components:

(1) *Ethically aligned design*, which creates AI solutions that avoid discrimination, maintain fairness and justice, and do not replace humans.
(2) *Technology that fully reflects human intelligence*, which further enhances AI technology to reflect the depth characterized by human intelligence (more like human intelligence).
(3) *Human factors design* to ensure that AI solutions are explainable, comprehensible, useful, and usable.

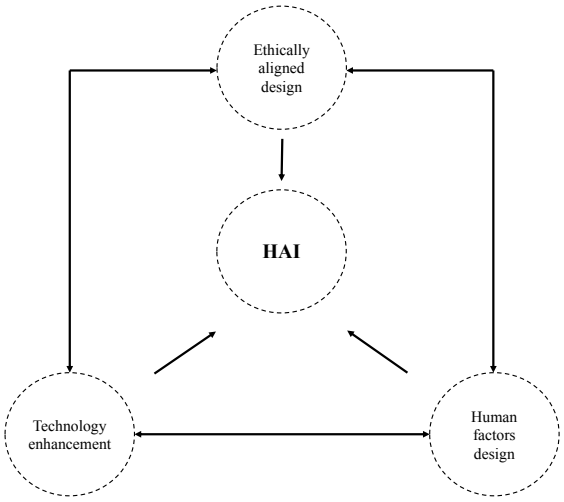


**Figure 3** An extended HCAI framework (Xu, 2019).

Following this, a wide range of initiatives to establish ethical guidelines and frameworks comes into practice each day, acting as a quick solution to address these ethical, societal, and legal problems and altogether build a socially responsible AI. Some examples follow. AlgorithmWatch[3] is a non-profit organization that helps in pointing out ethical conflicts and currently holds 150 ethical guidelines for making algorithmic decision-making processes effective and inclusive. AI4People[4] Ethical framework offers a series of recommendations for the development and adoption of AI and is especially tailored to the European context. In their meta-analysis of ethical frameworks, Floridi and Cowls (2019) find that almost all of the guidelines collectively agree on the same set of principles and themes, but what seems to be missing is a concrete methodology for the application/mapping of these principles in practice.

[3] https://algorithmwatch.org/en/
[4] https://www.eismd.eu/ai4people/

Overall, we suggest that AI systems should not only keep humans in the loop but also provide higher levels of human control. As the system becomes more human-centered and under the control of humans, it leads to an increase in human performance and achieves higher levels of self-efficacy, mastery, creativity, and responsibility. For example, the StudySmarter learning app[5] provides tailored learning plans while offering students a level of autonomy to create their own flashcards and summaries and to finalize their lesson plans.

## 5 Recommendations for action and further research

Tremendous development of EdTech applications is taking place. However, the use of AI within these tools is scarce. In our opinion, one reason for the restrained development might be the fact that society is, in principle, rather skeptical about the use or development of AI systems due to the dystopian framing of the concept of AI. Avanade's (2017) research on HCAI shows that 88% of the global business and IT decision-makers say they do not know how to use AI, and 79% say that internal resistance limits their implementation of it. Due to this gap in innovation and development, we propose that EdTech providers should consider developing more HCAI-based approaches unfolding the potential of AI in education. This allows us to clearly envision the benefits that AI systems have to offer. We propose that this is the right time to consider value-conscious design principles for more human-centered and responsible AI that takes the social, legal, and moral values into consideration prior to and during the development of the product. In the current market, AIED systems work in two ways. They are either (1) *rule-based*, where the system is given a set of rules and it keeps applying these rules to problems to find an answer, or (2) *learning-based*, where the system observes, finds patterns, and makes predictions independently. The German EdTech sector is currently in the first phase, where rule-based AIED systems are developed with pre-defined rules.

However, as the shift is taking place toward developing modern AI systems that are learning-based, concerns have emerged regarding AI taking over human control, algorithmic violations caused by bad data, socioeconomic inequality, and privacy violations. Consequently, the AI community is shifting its focus to emphasize HCAI, due to the mixed opinions on AI (from different actors involved in education–government, private entities, parents, leaders of institutions) and its capabilities, and to rethink how to develop an AI system that complies with human values and does not pose risks to humanity. The shift is happening slowly and it is not that noticeable because HCAI has the same capabilities as AI, with the only difference being that instead of replacing human workers, HCAI aims at augmenting human workers and enhancing business outcomes.

Educating stakeholders about the potential and utopian capabilities of HCAI will help in attaining the goal of student and teacher success and reduce the anxiety or fear of AI systems. Interestingly, we are finding an increasing number of initiatives at the public level (e.g., Elements of AI in Finland[6] and AI Campus in Germany[7]) that provide information about the opportunities/potential and challenges/risks of AI and raise awareness of the topic. In addition, data literacy initiatives are increasingly emerging to improve the general understanding of how data can be used. Such initiatives and educational projects indicate that society should develop an awareness of AI technologies. For further research, it would

[5] https://www.studysmarter.de/
[6] https://www.elementsofai.de
[7] https://ki-campus.org

Event Proceedings: LUT Scientific and Expertise Publications: ISBN 978-952-335-566-8

be exciting to analyze concrete cases of how HCAI approaches influence the development of AIED in the educational market and application readiness.

## References and Notes


Ahmad, K., Iqbal, W., El-Hassan, A., Qadir, J., Bendaddou, D., Ayyash, M., Al-Fuquaha, A. 2020. Artificial intelligence in education: A panoramic review. Available at: https://www.researchgate.net/publication/342323576_Artificial_Intelligence_in_Education_A_Panoramic_Review.

Alkhatlan, A., Kalita, J. 2018. Intelligent tutoring systems: A comprehensive historical survey with recent developments. Available at: https://arxiv.org/pdf/1812.09628.pdf.

Avanade. 2017. Will today's security stand up to tomorrow's threats?. Available at: https://www.avanade.com/-/media/asset/point-of-view/security-services-strategy-pov.pdf?la=nl-nl&ver=1&hash=15584ADD972BC0929F355B97822478EC.

Auernhammer, J. 2020. Human-centered AI: The role of Human-centered Design Research in the development of AI. DOI: 10.21606/drs.2020.282.

Dembrower, K., Wåhlin, E., Liu, Y., Salim, M., Smith, K., Lindholm, P., Eklund, M., Strand, F. 2020. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: A retrospective simulation study. DOI: https://doi.org/10.1016/S2589-7500(20)30185-0.

Dignum, V. 2019. Responsible artificial intelligence: How to develop and use AI in a responsible way. In book series Artificial Intelligence: Foundations, Theory, and Algorithms (AIFTA). DOI: https://doi.org/10.1007/978-3-030-30371-6.

EdTechXGlobal 2016. EdTechXGlobal report 2016 – Global EdTech industry report: A map for the future of education and work. Available at: http://ecosystem.edtechxeurope.com/2016-edtech-report.

Floridi, L., Cowls, J. 2019. A Unified Framework of Five Principles for AI in Society. DOI: 10.1162/99608f92.8cd550d1.

Forbes Insights. 2020. How AI is revamping the call center. Available at: https://www.forbes.com/sites/insights-ibmai/2020/06/25/how-ai-is-revamping-the-call-center/.

Global Executive Panel. 2019. Adoption of AI in education is accelerating. Massive potential but hurdles remain. Available at: https://www.holoniq.com/notes/ai-potential-adoption-and-barriers-in-global-education/.

Holmes, W., Anastopoulou, S., Schaumburg, H., Mavrikis, M. 2018. Technology-enhanced Personalized Learning: Untangling the Evidence. Robert Bosch Stiftung GmbH, Stuttgart.

Holmes, W., Bialik, M., Fadel, C. 2019. Artificial Intelligence in Education – Promises and Implications for Teaching and Learning. Independently published, Boston.

Luckin, R., Cukurova, R. 2019. Designing educational technologies in the age of AI: A learning sciences-driven approach. In British Journal of Educational Technology, Vol. 50. DOI: 10.1111/bjet.12861.

Popenici, S.A.D., Kerr, S. 2017. Exploring the impact of artificial intelligence on teaching and learning in higher education. In Research and Practice in Technology Enhanced Learning, Vol. 12(22). DOI: https://doi.org/10.1186/s41039-017-0062-8.

Renz, A. 2020. Die Corona-Krise als Chance für die Digitalisierung der Bildung. Available at: https://www.equeo.de/corona-krise-als-chance-fuer-die-digitalisierung-der-bildung/.

Renz, A., Hilbig, R. 2020. Prerequisites for artificial intelligence in further education: Identification of drivers, barriers, and business models of educational technology companies. In International Journal of Educational Technology in Higher Education, Vol. 17(14).

Renz, A., Krishnaraja, S., Gronau, E. 2020a. Demystification of artificial intelligence in education – how much AI is really in the educational technology? In International Journal of Learning Analytics and Artificial Intelligence in Education, Vol. 2.

Renz, A., Krishnaraja, S., Schildhauer, T. 2020b. A new dynamic for EdTech in the age of pandemics, presented at: ISPIM Innovation Conference – Virtual, June 2020.

Schmidt, A. 2020. Interactive human-centered artificial intelligence: A definition and research challenges. In International Conference on Advanced Visual Interfaces (AVI '20), September 28–October 2, 2020, Salerno, Italy. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3399715.3400873.

Schneiderman, B. 2020. Human-centered artificial intelligence: Three fresh ideas. AIS Transactions on Human-Computer Interaction, Vol. 12(3), 109–124. DOI: 10.17705/1thci.00131. Available at http://aisel.aisnet.org/thci/vol12/iss3/1.

Weitekamp, D., Harpstead, E., and Koedinger, K. 2020. An interaction design for machine teaching to develop AI tutors. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–11. DOI: https://doi.org/10.1145/3313831.3376226.

Xu, W. 2019. Toward human-centered AI: A perspective from human–computer interaction. In Interactions, Vol. 26(4), 42–46. DOI: https://doi.org/10.1145/3328485.

Zawacki-Richter, O., Marín, 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 Education Technology in Higher Education, Vol. 16(39).