Towards Ethical Design Features for Pedagogical Conversational Agents

Conference Paper · August 2022

CITATIONS
2

READS
155

2 authors:

Ricarda Schlimbach
Technische Universität Braunschweig
19 PUBLICATIONS **28** CITATIONS
SEE PROFILE

Bijan Khosrawi-Rad
Technische Universität Braunschweig
13 PUBLICATIONS **12** CITATIONS
SEE PROFILE

Towards Ethical Design Features for Pedagogical Conversational Agents

Completed Research

Ricarda Schlimbach
TU Braunschweig (DE)
r.schlimbach@tu-bs.de

Bijan Khosrawi-Rad
TU Braunschweig (DE)
b.khosrawi-rad@tu-bs.de

Abstract

Pedagogical Conversational Agents (PCAs) offer the potential to increase educational equity worldwide by making learning accessible to all as a service for good, often enabled by artificial intelligence (AI). Yet, there are ethical challenges to the design and use of PCAs that hinder the achievement of individual and social goals. However, in addition to a multitude of directives on the ethical design of information systems, concrete resulting design features for PCAs still fall short in scientific literature. Furthermore, a human-centered ethical discussion that integrates future users' involvement in ethical PCA design is scarce to find. Based on a co-creation process embedded in Design Science Research with a total of 40 students, we derive requirements and concrete features for ethically responsible PCAs and reflect them along with the OECD principles for trustworthy AI. Initial conceptual prototypes visualize exemplary instantiations.

Keywords

Pedagogical Conversational Agent, Design Features, Ethics, Ethical Design.

Introduction

Teaching and learning in schools and other academic institutions is increasingly shifting to the digital space thanks to digitalization. The drive for change in recent years in favor of competency-based, interactive, and digital teaching was further accelerated by forced disruptions during the global pandemic situation (Chi and Wylie 2014; Grogorick and Robra-Bissantz 2021). Digital teaching and learning had to become the norm within a fairly short period of time (Grogorick and Robra-Bissantz 2021). As a result, innovative digital services to support learning became a greater focus of research. One example are Pedagogical Conversational Agents (PCAs), i.e. digital learning facilitators communicating via natural language, which are supposed to support learning as text-based chatbots or speech-based assistants (Gubareva and Lopes 2020; Hobert 2019; McTear et al. 2016). Enabled by rapid advances in AI and natural language processing, they open up the vision of a modern and fair education due to their scalability, 24/7 availability, and potential to tailor their interaction individually to learners and their needs (Maedche et al. 2019). In order to unfold the resulting PCAs' potential for increasing educational fairness, ethical considerations in the design process are gaining relevance (Richards and Dignum 2019). Designing ethically responsible information systems (e.g. PCAs) is scientifically motivated (Rothenberger et al. 2019; Spiekermann et al. 2022; Wambsganss et al. 2021) and reflected in the growing inventory of ethical directives on the ethical use of AI (AlgorithmWatch 2020). However, prescriptive ethical design knowledge for PCAs lacks depth of detail for practical application (Richards and Dignum 2019; Schlimbach et al. 2022) or is not specifically tailored to the educational context (e.g., Feine et al. 2019; Wambsganss et al. 2021). More precisely, current research remains on a highly abstract level and lacks the derivation and discussion of concrete features for the practical design of PCAs from an ethical perspective (Wambsganss et al. 2021). To contribute to this gap, our paper addresses the following research question:

Which design features can be derived from existing ethical directives to foster the ethical design of PCAs?

The possibilities for implementing ethically-aware PCAs are manifold as well as controversial (Schlimbach et al. 2022; Wambsganss et al. 2021). Therefore, we aim to derive corresponding implications for PCAs based on the design features to contribute to expanding the current PCA design knowledge base.

Our article is outlined as follows: In the next chapter, we introduce the research background of ethical PCA design based on the core principles of the Organization for Economic Co-operation and Development (OECD) in dealing with AI and complemented by further supportive literature, before the follow-up chapter introduces the DSR paradigm and the methodological elaboration of design features as a co-creation process with learners. We then derive prescriptive design knowledge and discuss expository instantiations controversially. Our contribution concludes with a summary, by also admitting limitations and hinting at implications for further research avenues.

Research Background

Information Ethics for PCA Design

Information ethics is a philosophical discipline in its own that deals, among other aspects, with moral issues relating to the development and use of AI systems and the ethical view of data-processing technologies and algorithms (Bendel 2021). The particular relevance of the use of AI-based PCAs is demonstrated by the establishment of political committees at the national and international level (e.g., EKKI 2020; EU Commission 2019) to deal with guidelines and recommendations for action on this topic. However, resulting directives are criticized for lacking practical transferability and for too little depth of detail (Wambsganss et al. 2021). At the same time, the high density of guidelines bears the danger of confusion and overregulation (Schlimbach et al. 2022). The AI Global Ethics Inventory lists 173 ethics guidelines in its database in January 2022 (AlgorithmWatch 2020), after the number of guidelines had doubled within two years (EKKI 2020). As an orientation-giving intersection of the plethora of guidelines, the OECD has defined five core "ethical principles for responsible stewardship and trustworthy AI" (OECD 2019a). Since they summarize the input of 50 experts from government, business, society, and academia from 20 nations and are recognized in the scientific literature related to ethical design (Schlimbach et al. 2022; Wambsganss et al. 2021) as well as by the governments of more than 30 countries worldwide (OECD 2019a), we decided to build upon these principles as a foundation for prescriptive design knowledge.

The OECD Principles for AI

The following OECD principles were adopted in May 2019 (OECD 2019b):

1. "AI should benefit people and the planet by driving inclusive growth, *sustainable development and well-being*.
2. AI systems should be designed in a way that respects the rule of law, *human rights*, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure *a fair and just society*.
3. There should be *transparency and responsible disclosure* around AI systems to ensure that people understand AI-based outcomes and can challenge them.
4. AI systems must function in a *robust, secure and safe way* throughout their life cycles and potential risks should be continually assessed and managed.
5. Organizations and individuals developing, deploying or operating AI systems should be held *accountable* for their proper functioning in line with the above principles."

Since they have a directive scope and reflect a high abstraction level typical for that level of instantiation, their application to specific use cases and ethical design features remains still challenging (Wambsganss et al. 2021).

Design Science Research as Guiding Paradigm

Our methodological approach takes up the design-oriented approach of Design Science Research (DSR) according to Hevner (2007) and aims to counter the lack of design knowledge for ethical PCAs with practical design features (relevance) and to discuss them with supportive scientific literature (Möller et al. 2020) (rigor). Our focus lies on the derivation of prescriptive design knowledge highlighting the relevance cycle. We have therefore taken up the idea of user integration in a co-creation process by deriving design features

for ethically responsible PCAs with 20 students (aged 19-27) of an interdisciplinary academic seminar at a German university and another 20 interviewed students to incorporate their ideas as potential future users (Triviño-Cabrera et al. 2021). The course took place in the winter semester of 2021/22 with bachelor's and master's students from different disciplines (e.g., architecture, biology, teaching, engineering, business informatics). The participants were divided into four heterogeneous groups of five students each to work on the semester-long team project and were given the task of conceptualizing PCA instantiations following ethical directives. The students followed steps 1-9 of the DSR paradigm as illustrated in Figure 1.

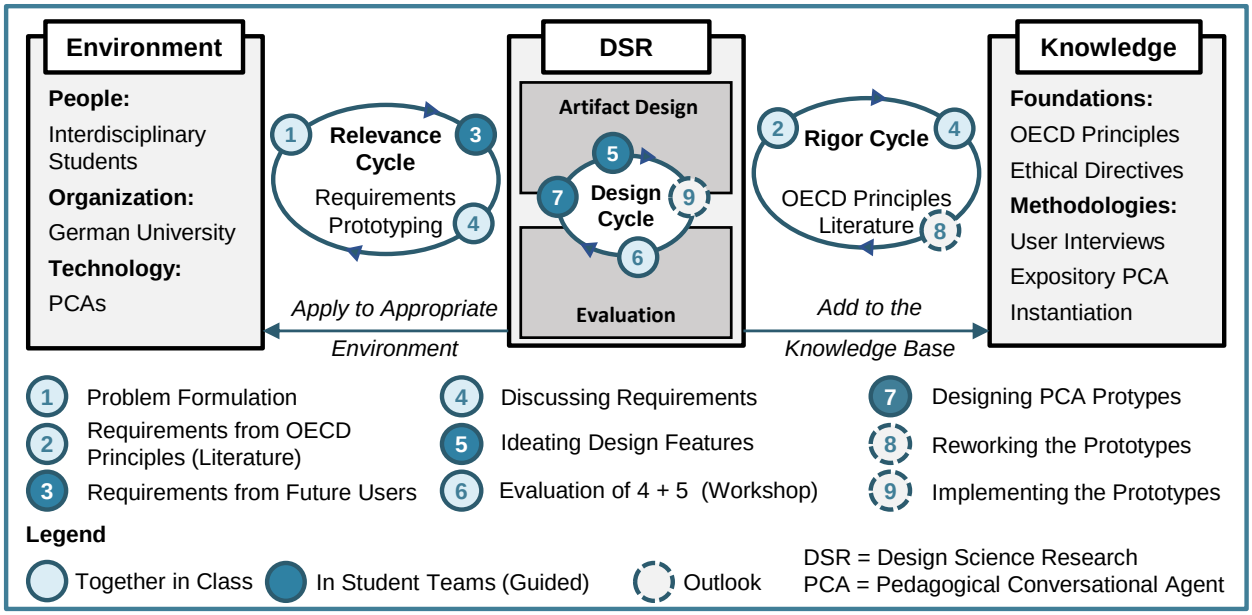

Figure 1. DSR Co-Creation with Students based on Hevner (2007)

First, the lecturer introduced PCAs in the seminar, discussed their opportunities, and formulated the problem of unethical PCA design (1). For this purpose, supportive scientific literature was provided (e.g., Casas-Roma and Conesa 2021; Rothenberger et al. 2019; Wambsganss et al. 2021) and also brought along by the participants themselves (e.g., Atkins et al. 2021; Bailey et al. 2021; McDonnell and Baxter 2019) for controversial discussions in class. In the next step (2), the lecturer presented the five OECD principles for dealing with AI (cf. p.2) to derive ethical requirements complemented by supportive literature. Subsequently, the participants conducted five interviews per team with potential future users to gather further ethical needs for an ethical PCA design (3) and to discuss and supplement them with their own experiences (4). The guideline-based interviews lasted about 30 minutes, included mostly open-ended questions, and addressed the following topics: opportunities and challenges of the use of PCAs from an ethical perspective, necessary control instances for the trustworthy handling of AI in teaching as well as design recommendations to be derived, and ideas for desired functionalities. Each student group determined its own subject focus (e.g., bias prevention) and ideated their findings into concrete design features (5) in a creativity workshop within their group, followed by a discussion and evaluation of the results together with all groups in class (6). The reworked PCA design feature concepts were then exemplarily instantiated as one conceptual PCA prototype (7) per student team. Merging them into one prototype to be actually implemented by a team of professional designers and developers was not in scope of the seminar but is planned for the future (8-9).

Deriving Design Knowledge

Problem Formulation

PCAs might positively revolutionize education, e.g., they could serve as support for under-served professor/learner ratios in academia (Chun Ho et al. 2018; Winkler et al. 2019), positively affect students' engagement (Winkler et al. 2020), or offer individualized support through adaptive technologies (Wambsganss et al. 2020), and potentially provide access to groups previously disadvantaged in education

(Richards and Dignum 2019). Despite these technology-enabled opportunities, however, the problem of lacking prescriptive design knowledge to prevent unethical or discriminatory PCAs remains unsolved. Design biases can play a major role in the development of PCAs, for instance concerning the gender and appearance of the avatar or stereotypical communication behavior (McDonnell and Baxter 2019; Wambsganss et al. 2021). In the form of algorithmic bias, discrimination is systematically reproduced in the algorithms because they learn from the human-supplied data sets in a pattern-based manner, resulting in systematic biases (Casas-Roma and Conesa 2021). The accountability and transparent data usage is a major challenge, especially against the backdrop of rapidly developing AI and accompanying fears of loss of control, explainability of AI algorithms, or increasing user mistrust (Wambsganss et al. 2021). An early, preventive counteraction is important to address the problem of unethical and discriminatory PCAs.

Deriving Design Requirements

DSR steps 2-4 result in the following design requirements (DRs) numbered from DR1.1 to DR5.3 along with the five corresponding OECD design principles:

1. *Sustainable Development & Well-being*: This guideline bundles requirements that should sustainably shape education through PCAs and the well-being of their users. The students address as requirements the sustainable development of technical solutions that are continuously adapted to technical progress (Maedche et al. 2019) and legal conditions (Spiekermann et al. 2022) (DR1.1), the enabling of accessibility (Vu et al. 2016) regardless of the social status of the learners for more educational equity (DR1.2), and the integration of mechanisms that ensure the well-being of learners in the process of usage (Kim et al. 2013) (DR1.3).
2. *Rights & Fairness:* For the second principle, learners call for the intentional non-discriminatory design of the PCA (DR2.1) with measures to overcome algorithmic bias (DR2.2) (Casas-Roma and Conesa 2021; Spiekermann et al. 2022), and the inclusion of democratic rights and values (Wambsganss et al. 2021) (DR2.3).
3. *Transparency & Responsibility:* The requirements include the (self-) obligation to transparently demonstrate data processing (Spiekermann et al. 2022) (DR3.1), compliance with data protection regulations (Alt et al. 2021) (DR3.2), and responsible design of the relationship between the PCA and its human learner without abuse of trust (Richards and Dignum 2019; Zierau et al. 2020) (DR3.3).
4. *Supervised Robustness & Security*: Within the fourth OECD principle, the introduction of control mechanisms (Alt et al. 2021) (DR4.1), measures to protect the privacy of users (Calvaresi et al. 2021) (DR4.2), and the continuous supervision of AI by humans (Rothenberger et al. 2019) (DR4.3) are required.
5. *Accountability:* Requirements in the fifth category include the ability for users to report unethical behavior of the system (DR5.1), regular reporting and response to undesirable functionality (Wambsganss et al. 2021) (DR5.2), as well as accountability in AI promoting social trust through education (Latham and Goltz 2019) and clear legislation in this area (Dignum 2017)(DR5.3).

Deriving Design Features for Ethically Responsible PCAs

In this section, we succinctly present the four prototypes (A-D as illustrated in Figure 2) created in the seminar with their integrated design features; we refer in parentheses to the design requirements to which they respond.

Group A has created a prototype with the help of the software UXPin, which should support students in their time management as well as individually in the learning process. For this purpose, several ethically motivated design features are integrated into the app. Personalized learning suggestions adapt the modules to be completed to the learner's needs (e.g., their schedule or skill level); learning methods (e.g., flashcards, mind maps, exam simulation) are adaptable by the student for diversified learning (DR1.1). To promote well-being and avoid excessive use of the app, learning breaks are integrated into the app as a feature that encourages the user to exercise and take a break from learning, for example, through short games or guided relaxation exercises (DR1.3). Communication behaviors are intended to adapt to the user to enhance comfort and trust, but a blacklist of vulgar expressions and undesirable communication patterns prevents the AI from learning unethical behaviors (DR2.2). A configurable human avatar is intended to serve as a

role model and build trust by allowing characteristics such as body, clothing, language, or gender to be specified by the user (DR2.1). User data should only be stored for specific purposes (e.g., for adaptation purposes based on the user profile) and not be passed on to third parties (DR4.2). Students claim app providers legally liable for data processing (DR5.3). A data protection pop-up as a feature informs the user as soon as (additional) data is stored about him/her and the learner can view this data at any time (DR3.1).

Group B created the prototype "Felix - Your learning companion" (cf. screenshot B) and pushes the focus topics bias (DR2.1;2.2), privacy (DR4.2), and social justice (DR1.1). First, educational equity is to be increased by financing Felix externally (e.g., advertisements, funding, government grants) so that access to education is made easier for as many social classes as possible and, in perspective, also for developing countries via PCAs (DR1.2). The risk of addiction through excessive use of app applications should be prevented by push messages that regularly encourage users to take breaks and offline relaxation phases (DR1.3). The group suggests integrating gamified elements, such as a leaderboard to motivate competitively in class, but also notes that points received should be tailored to individual performance potential in the spirit of inclusion (DR1.2). The fox as an avatar was chosen allegorically because of its traits attributed in fables to promote learning and was deliberately chosen as an avatar in contrast to a human embodiment so as not to discriminate against any humanoid trait expression (DR2.1). The fox symbolizes various states (e.g., it becomes fat when "fed" by solving mock exams or curls symbolically during loading processes) and is intended to introduce a humorous element to make learning fun (DR1.3). Fairness measures (DR1.1) fed into the technology are intended to prevent algorithmic bias as a feature (DR2.2). Data protection is to be regularly reviewed by a supervisory body with data protection officers (DR4.3) and adapted to current legislation (DR3.2). A decentralized software architecture (DR4.2), transparency self-commitment (DR3.1), data encryption (DR4.2), crowd-based monitoring of the code (4.3), and anonymity options for the user like pseudonymization and privacy settings (DR4.2) support responsible data handling (DR3.3). In addition, an interface for reporting discriminatory behavior (DR5.1) is available as a feature. A heterogeneous and interdisciplinary council of experts (e.g., from IT, education, science) (DR4.3) should decide on incoming reports (5.2) and develop sustainable measures (DR1.1) for the practical implementation of ethics-sensitive PCAs and also publish them (DR5.3).

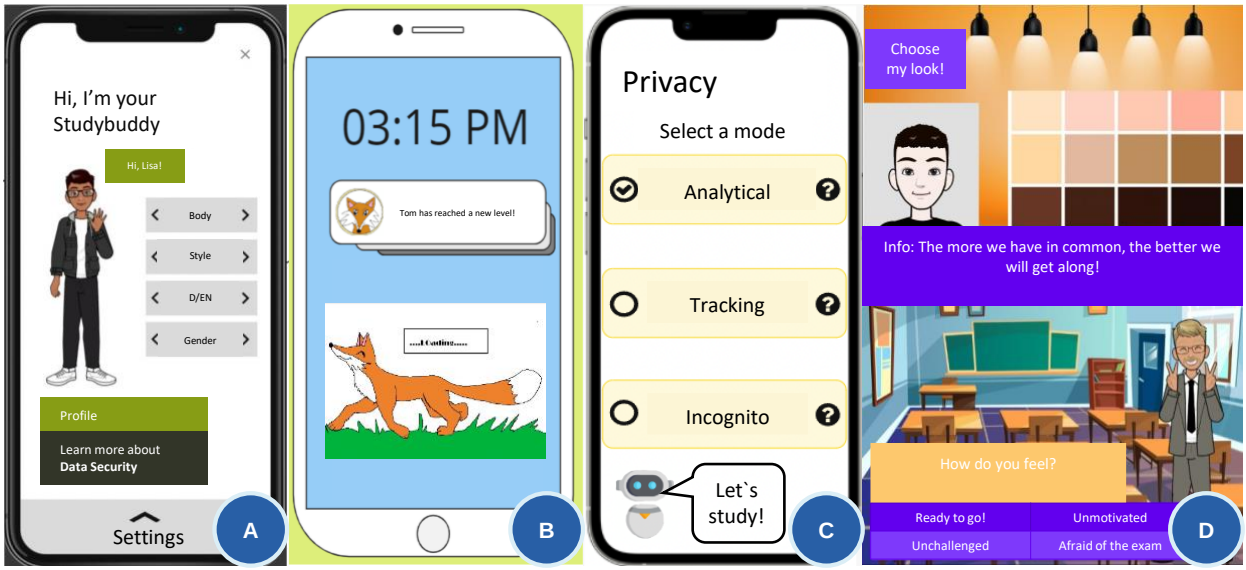

Figure 2. Fragments of Prototypes A and B

Group C emphasizes the importance of handling personal data in AI-supported PCAs but also emphasizes the opportunity to revolutionize individualized learning, independent of space and time, by ethically responsible use of future technologies (e.g., cloud computing, deep learning, voice assistance systems) (DR1.1). The group chose a robot as an avatar design to transparently visualize at all times that they interact with a machine (DR2.1) and thus counteract perceived social manipulation (due to what is perceived as human communication). Chat functions with other human learners are intended to prevent social isolation (DR1.3). The selection of a preferred privacy mode (analytical/tracking/incognito as shown in screenshot C) gives the user freedom of choice (DR4.2). Fixed usage periods are intended to responsibly limit the

consumption of the application (DR1.3). The extent of the individualization options depends on the users' willingness to share user data (DR3.3), so they can decide for themselves how much data will be shared with the AI.

The fourth group D designed their prototypical concept in a virtual learning environment from the perspective of a high-ability student (IQ > 129) and focuses on the inclusion of minorities (DR1.2). In doing so, on the visual level, they propose user customization of a human avatar whose gender, skin color, hair, eye color and shape, body shape, and clothing can be selected (see screenshot D) along with other body-related elements (e.g., glasses, hearing aid, wheelchair) to represent a wide range, especially of underrepresented characteristics (DR2.1). Integrated queries are designed to infer the emotional state of the learner (e.g., anxiety about the test, motivation to learn, frustration due to being under-challenged) and to adapt the gestures of the humanoid teacher (DR2.1) as well as the learning content and cognitive level according to the information provided (DR1.3). The group has deliberately decided in favor of adaptable features by the user (DR2.3) because these can be implemented much more sparingly so that responsible handling of user data is easier to implement (DR3.1). These do not have to be fed into adaptive technologies (DR1.1), but the user carries out all settings on his/her own responsibility (DR3.3). This is intended to leave no room for algorithmic bias (DR2.2).

Reflection on the Results

The multifaceted and sometimes contradictory implementation of the ethical directives into concrete design features by the student groups shows that the directives offer great interpretative scope for application. Thus, it is important to discuss design knowledge at the level of design features to provide impulses for the practical design of ethically responsible PCAs. The following Figure 3 summarizes design features for ethically responsible PCAs derived in the co-creation process with students as future users.

OECD Principle	Design Requirements	Proposed Design Features
1 **Sustainability & Well-Being**	1.1 Sustainable development 1.2 Accessibility for everyone 1.3 Mechanisms for well-being	• Open source code • Externally financed, openly accessible • Push-up notifications for breaks; usage time control, integrated jokes
2 **Human Rights & Fairness**	2.1 Discrimination avoidance 2.2 No algorithmic bias 2.3 Democratic values & rights	• Non-human avatar or adaptable humanoid embodiment • Controlling fairness measures • Value-sensitive communication
3 **Transparency & Responsibility**	3.1 Transparent data management 3.2 Conformity with GDPR 3.3 Responsible PCA design	• Explanatory info buttons; right of inspection of stored user data • Check for consistency with data laws • Liability of the service supplier
4 **Robustness & Security**	4.1 Technical control mechanisms 4.2 Data security 4.3 Permanent AI supervision	• Data protection protocols, encryption • Privacy settings (e.g., incognito mode) • Controlling expert committee; permanent veto option over AI
5 **Accountability**	5.1 Auditability 5.2 Reporting 5.3 AI education and legislation	• Retraced logging and tracking • Interface to report discrimination • Comprehensive info on AI and its scope, functions and legislation

Figure 3. Overview of the Derived Design Knowledge

The most diverging feature among the four groups is the avatar design with the common goal of promoting diversity and preventing discrimination: While groups A and D propose human avatars that are adaptable to the learner in a variety of trait expressions, group C chose a robotic avatar to emphasize the AI's non-human communication, and group B chose a fox as a symbolic fable character with attributed traits such as

cleverness and curiosity. However, the teams' reasoning differed seriously, as group A recommends making the avatar more mature, older, and different from the learner for making diversity visible, whereas group D explicitly calls for similarity as an identification figure. These contrasting perspectives are also reflected in the scientific literature. The computers are social actors (CASA) theory states that people show social reactions towards computers although they know that they are machines (Nass and Moon 2000). So-called *social cues* further support this behavior, for example through human-like avatars or the integration of humorous elements (Feine et al. 2019). The so-called *persona effect* revealed in studies with students that the presence of a lifelike character in an interactive learning environment can have a strong positive effect on students' perception of their learning experience (Lester et al. 1997). Human-like avatars can also positively influence the trust and credibility perceived towards the PCA, but only if the humanoid design happens to a balanced degree (Feine et al. 2019). In case of a mental model mismatch between user expectation and the actual PCA design, this can also harm the user experience (Luger and Sellen 2016). For this reason, some authors in the literature also tend to advocate for animal-like designs, for example in the care context (Moyle et al. 2019).

Three of the four groups also addressed - primarily based on their own experiences - the need to prevent excessive consumption of digital app applications by providing appropriate features (e.g., limited usage time or notification of breaks and relaxation), thus raising the aspect of addiction potential and uncontrolled use of digital services. This aspect underlines the value of a user-centered co-creation process because the students brought up the fear of uncontrolled consumption and thus bring in the need for recreational breaks away from digital environments (Mirbabaie et al. 2020) instead of maximizing PCA usage.

Furthermore, it is striking that the proposed design features are almost exclusively adaptable by the learner and hardly exploit the possibilities of AI for adaptivity (Atkins et al. 2021). This also reflects the still low users' acceptance of this future technology (Latham and Goltz 2019) and is additionally underlined by the high perceived risks in data protection and the demand for more educative information about AI and its relevance in learning (Lameras and Arnab 2022) as informative design features. This again shows that not only technologically possible features should be developed for the user, but also that design features must be conceived *with* the users to integrate their fit with real user needs at an early stage and thus promote the later acceptance of the application (Khosrawi-Rad et al. 2022). However, users might not (fully) understand the potential of new technologies or might not know what is best for them, so design decisions should not exclusively rely on future users. Since PCAs are also artifacts designed by humans and for humans, a decisive aspect of their programming is to design them mindful of human resources and with human well-being in mind. Only a diverse team of designers and developers is capable of representing sound expertise and a variety of user needs.

Limitations

Our contribution offers concrete suggestions for the ethically responsible design of PCAs with a particular focus on practical relevance from the user perspective. Nevertheless, our design knowledge findings are limited and require further research. We focused strongly on the needs, requirements, and design ideas of students as future users, which on the one hand raise a new and important perspective; on the other hand, they also bring in bias, since the design features reflect the opinions of a total of 40 students in Germany, who are not necessarily representative for potential future user groups of PCAs (e.g., in other learning scenarios or other cultures) and have only limited expertise (e.g., in terms of technical feasibility). In addition, individual features (e.g., avatar design) have met with controversy even within this rather heterogeneous user group, which is why this manifests an even more in-depth discussion with additional potential users. Furthermore, the DSR process was designed in favor of practical relevance and elaboration of concrete concepts but should be more solidly grounded in the rigor cycle (Hevner 2007). This includes the systematic analysis of PCA design directives from an ethical viewpoint, as well as the iterative implementation and evaluation of design knowledge into mature IT artifacts and their scientific discussion. Besides, the OECD Principles we based upon, only apply to countries with democratic values and limit therefore the transferability of our findings. Nevertheless, we consider our initial results to be valuable in order to focus more on the needs and ideas of future users in a scientifically embedded approach in the sense of human-centered service design.

Conclusion

PCAs hold the potential for inclusive and empowering services in education (Richards and Dignum 2019). At the individual level, they can offer their users tailored learning support, relieve the burden on the organizational level, and potentially provide more global educational equity through their digital accessibility and constant availability (Maedche et al. 2019; Vu et al. 2016). However, those chances require their ethically responsible design with human well-being in mind. Despite numerous abstracted guidelines, there is a lack of concrete design features for the practical development of ethically designed PCAs (Wambsganss et al. 2021). Since PCAs involve innovative technologies, such as AI, they should focus on the needs-oriented derivation of design knowledge with future users (Dignum 2017) on the one hand and non-discriminatory control mechanisms on the other hand. In the context of an academic seminar, we derived various design features for ethical PCAs with a total of 40 students under the paradigm of DSR and implemented them exemplarily in four conceptual prototypes. Their classification and reflection along the OECD principles for responsible and trustworthy AI demonstrate the complexity of their interpretation and at the same time underlines the limits of user-driven co-creation processes.

Consequently, future research should address an appropriate level for design directives to allow certain design freedom on the one side, but also to provide clear orientation and unambiguous answers to controversially discussed aspects (e.g., humanoid PCA design) on the other side. In this respect, design knowledge should focus more on the transfer and discussion of concrete features. Furthermore, future research is still needed to discover an optimal balance between a needs-oriented co-creation with users while at the same time taking expert knowledge into account. Otherwise, there is a risk of not discovering latent needs or hidden (technological) potentials in the first place, or of developing solutions that are detached from the actual problem. Our contribution aims to initiate a debate on those aspects and calls for in-depth research to evaluate, detail, and anchor our initial results for a more human-centered future of learning in general and the ethically responsible derivation of design features for PCAs in particular.

Acknowledgments

This contribution results from the project StuBu (grant number 21INVI06), which is funded by the German Federal Ministry of Education and Research (BMBF). We would like to thank the TU Braunschweig students of the seminar *Service Learning: Digital Transformation* (winter term 2021/22) for their dedicated participation and inspiring ideas for ethical design features.

REFERENCES

AlgorithmWatch. 2020. "AI Ethics Guidelines Global Inventory," *AI Ethics Guidelines Global Inventory*. (https://inventory.algorithmwatch.org, accessed February 14, 2022).

Alt, R., Göldi, A., Österle, H., Portmann, E., and Spiekermann, S. 2021. "Life Engineering: Towards a New Discipline," *Business & Information Systems Engineering* (63:2), pp. 191–205. (https://doi.org/10.1007/s12599-020-00680-x).

Atkins, S., Badrie, I., and Otterloo, S. 2021. "Applying Ethical AI Frameworks in Practice: Evaluating Conversational AI Chatbot Solutions," *Computers and Society Research Journal*.

Bailey, J., Patel, B., and Gurari, D. 2021. "A Perspective on Building Ethical Datasets for Children's Conversational Agents," *Frontiers in Artificial Intelligence* (4). (https://doi.org/10.3389/frai.2021.637532).

Bendel, O. 2021. "Soziale Roboter in Der Moral," in *Soziale Roboter*, Springer, pp. 149–167.

Calvaresi, D., Calbimonte, J.-P., Siboni, E., Eggenschwiler, S., Manzo, G., Hilfiker, R., and Schumacher, M. 2021. "EREBOTS: Privacy-Compliant Agent-Based Platform for Multi-Scenario Personalized Health-Assistant Chatbots," *Electronics (Switzerland)* (10:6), pp. 1–30. (https://doi.org/10.3390/electronics10060666).

Casas-Roma, J., and Conesa, J. 2021. "Towards the Design of Ethically-Aware Pedagogical Conversational Agents," in *Advances on P2P, Parallel, Grid, Cloud and Internet Computing*, Lecture Notes in Networks and Systems, L. Barolli, M. Takizawa, T. Yoshihisa, F. Amato, and M. Ikeda (eds.), Cham: Springer International Publishing, pp. 188–198. (https://doi.org/10.1007/978-3-030-61105-7_19).

Chi, M. T. H., and Wylie, R. 2014. "The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes," *Educational Psychologist* (49:4), pp. 219–243. (https://doi.org/10.1080/00461520.2014.965823).

Chun Ho, C., Lee, H. L., Lo, W. K., and Lui, K. F. A. 2018. "Developing a Chatbot for College Student Programme Advisement," in *2018 International Symposium on Educational Technology (ISET)*, pp. 52–56. (https://doi.org/10.1109/ISET.2018.00021, accessed April 25, 2022).

Dignum, V. 2017. "Responsible Artificial Intelligence: Designing AI for Human Values," *ICT Digital Conference*, Daffodil International University, pp. 1–8.

EKKI. 2020. "Unterrichtung Der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und Wirtschaftliche, Soziale und Ökologische Potenziale, Bericht Der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung Und Wirtschaftliche, Soziale Und Ökologische Potenziale." (https://dserver.bundestag.de/btd/19/237/1923700.pdf, accessed April 25, 2022).

EU Commission. 2019. *Europäische Kommission, Generaldirektion Kommunikationsnetze, Inhalte Und Technologien, (2019). Ethik-Leitlinien Für Eine Vertrauenswürdige KI*, EU Publications Office. (https://data.europa.eu/doi/10.2759/856513).

Feine, J., Gnewuch, U., Morana, S., and Maedche, A. 2019. "A Taxonomy of Social Cues for Conversational Agents," *International Journal of Human-Computer Studies* (132), pp. 138–161. (https://doi.org/10.1016/j.ijhcs.2019.07.009).

Grogorick, L., and Robra-Bissantz, S. 2021. "Digitales Lernen und Lehren: Führt Corona zu einer zeitgemäßen Bildung?" *HMD Praxis der Wirtschaftsinformatik* (58:6), pp. 1296–1312. (https://doi.org/10.1365/s40702-021-00806-z).

Gubareva, R., and Lopes, R. P. 2020. *Virtual Assistants for Learning: A Systematic Literature Review*, in (Vol. 1), presented at the CSEDU 2020 - Proceedings of the 12th International Conference on Computer Supported Education, pp. 97–103.

Hevner, A. 2007. "A Three Cycle View of Design Science Research," *Scandinavian Journal of Information Systems* (19:2), pp. 87–92.

Hobert, S. 2019. "How Are You, Chatbot? Evaluating Chatbots in Educational Settings - Results of a Literature Review," in *Lect. Notes Informatics (LNI), Proc. - Series Ges. Inform. (GI)* (Vol. P-297), Pinkwart N. and Konert J. (eds.), Gesellschaft fur Informatik (GI), pp. 259–270. (https://doi.org/10.18420/delfi2019_289).

Khosrawi-Rad, B., Schlimbach, R., and Robra-Bissantz, S. 2022. "Gestaltung Virtueller Lern-Companions durch einen Co-Creation Prozess," *Wirtschaftsinformatik 2022 Proceedings*.

Kim, K. J., Park, E., and Sundar, S. S. 2013. "Caregiving Role in Human–Robot Interaction: A Study of the Mediating Effects of Perceived Benefit and Social Presence," *Computers in Human Behavior* (29), pp. 1799–1806. (https://doi.org/10.1016/j.chb.2013.02.009).

Lameras, P., and Arnab, S. 2022. "Power to the Teachers: An Exploratory Review on Artificial Intelligence in Education," *Information* (13:1), Multidisciplinary Digital Publishing Institute, p. 14. (https://doi.org/10.3390/info13010014).

Latham, A., and Goltz, S. 2019. "A Survey of the General Public's Views on the Ethics of Using AI in Education," in *Artificial Intelligence in Education*, Lecture Notes in Computer Science, S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, and R. Luckin (eds.), Cham: Springer International Publishing, pp. 194–206. (https://doi.org/10.1007/978-3-030-23204-7_17).

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., and Bhogal, R. S. 1997. "The Persona Effect: Affective Impact of Animated Pedagogical Agents," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, New York, NY, USA: Association for Computing Machinery, March 27, pp. 359–366. (https://doi.org/10.1145/258549.258797).

Luger, E., and Sellen, A. 2016. *"Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents*, Proceedings of the 2016 CHI conference on human factors in computing systems, pp. 5286–5297.

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., and Söllner, M. 2019. "AI-Based Digital Assistants," *BISE* (61:4). (https://doi.org/10.1007/s12599-019-00600-8).

McDonnell, M., and Baxter, D. 2019. "Chatbots and Gender Stereotyping," *Interacting with Computers* (31:2), pp. 116–121. (https://doi.org/10.1093/iwc/iwz007).

McTear, M., Callejas, Z., and Griol, D. 2016. *The Conversational Interface: Talking to Smart Devices*, Springer International Publishing. (https://doi.org/10.1007/978-3-319-32967-3).

Mirbabaie, M., Marx, J., Braun, L.-M., and Stieglitz, S. 2020. *Digital Detox – Mitigating Digital Overuse in Times of Remote Work and Social Isolation*, Proceedings of the International Conference on Information Systems, Wellington, New Zealand.

Möller, F., Guggenberger, T. M., and Otto, B. 2020. "Towards a Method for Design Principle Development in Information Systems," Proceedings of the International Conference on Design Science Research in Information

Systems and Technology, pp. 208–220. (https://doi.org/10.1007/978-3-030-64823-7_20).
Moyle, W., Bramble, M., Jones, C. J., and Murfield, J. E. 2019. "'She Had a Smile on Her Face as Wide as the Great Australian Bite': A Qualitative Examination of Family Perceptions of a Therapeutic Robot and a Plush Toy," *The Gerontologist* (59:1), pp. 177–185. (https://doi.org/10.1093/geront/gnx180).
Nass, C., and Moon, Y. 2000. "Machines and Mindlessness: Social Responses to Computers," *Journal of Social Issues* (56:1), pp. 81–103. (https://doi.org/10.1111/0022-4537.00153).
OECD. 2019a. "The OECD Artificial Intelligence (AI) Principles." (https://oecd.ai/en/ai-principles, accessed February 11, 2022).
OECD. 2019b. "What Are the OECD Principles on AI?," *OECD Observer*, June 2. (https://www.oecd-ilibrary.org/economics/what-are-the-oecd-principles-on-ai_6ff2a1c4-en, accessed April 21, 2022).
Richards, D., and Dignum, V. 2019. "Supporting and Challenging Learners through Pedagogical Agents: Addressing Ethical Issues through Designing for Values," *British Journal of Educational Technology* (50). (https://doi.org/10.1111/bjet.12863).
Rothenberger, L., Fabian, B., and Arunov, E. 2019. *Relevance of Ethical Guidelines for Artificial Intelligence – A Survey and Evaluation*, Proceedings of the European Conference on Information Systems, Stockholm & Uppsala, Sweden.
Schlimbach, R., Khosrawi-Rad, B., and Robra-Bissantz, S. 2022. "Quo Vadis: Auf dem Weg zu Ethik-Guidelines für den Einsatz KI-baisierter Lern-Companions in der Lehre?," *HMD Praxis der Wirtschaftsinformatik* (59:2).
Spiekermann, S., Krasnova, H., Hinz, O., Baumann, A., Benlian, A., Gimpel, H., Heimbach, I., Köster, A., Maedche, A., Niehaves, B., Risius, M., and Trenz, M. 2022. "Values and Ethics in Information Systems: A State-of-the-Art Analysis and Avenues for Future Research," *BISE* (64:2), pp.247-264. (https://doi.org/10.1007/s12599-021-00734-8).
Triviño-Cabrera, L., Chaves-Guerrero, E. I., and Alejo-Lozano, L. 2021. "The Figure of the Teacher-Prosumer for the Development of an Innovative, Sustainable, and Committed Education in Times of COVID-19," *Sustainability* (13:3), Multidisciplinary Digital Publishing Institute, p. 1128. (https://doi.org/10.3390/su13031128).
Vu, P., Fredrickson, S., and Meyer, R. 2016. "Help at 3:00 AM! Providing 24/7 Timely Support to Online Students via a Virtual Assistant," *Online Journal of Distance Learning Administration* (19:1), State University of West Georgia.
Wambsganss, T., Höch, A., Zierau, N., and Söllner, M. 2021. "Ethical Design of Conversational Agents: Towards Principles for a Value-Sensitive Design," *Wirtschaftsinformatik 2021 Proceedings*.
Wambsganss, T., Söllner, M., and Leimeister, J. 2020. "Design and Evaluation of an Adaptive Dialog-Based Tutoring System for Argumentation Skills," in *ICIS Proceedings*.
Winkler, R., Hobert, S., Salovaara, A., Söllner, M., and Leimeister, J. M. 2020. "Sara, the Lecturer: Improving Learning in Online Education with a Scaffolding-Based Conversational Agent," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: Association for Computing Machinery, pp. 1–14.
Winkler, R., Söllner, M., Neuweiler, M. L., Conti Rossini, F., and Leimeister, J. M. 2019. "Alexa, Can You Help Us Solve This Problem?: How Conversations With Smart Personal Assistant Tutors Increase Task Group Outcomes," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow Scotland Uk: ACM, May 2, pp. 1–6. (https://doi.org/10.1145/3290607.3313090).
Zierau, N., Engel, C., Söllner, M., and Leimeister, J. M. 2020. "Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda," in *Proceedings of the 15th International Conference on Wirtschaftsinformatik*, Potsdam, Germany, pp. 99–114. (https://doi.org/10.30844/wi_2020_a7-zierau).