LUT University

Designing Personality-Adaptive Conversational Agents for Mental Health Care

Ahmad Rangina, Siemon Dominik, Gnewuch Ulrich, Robra-Bissantz Susanne

Designing Personality-Adaptive Conversational Agents for Mental Health Care

Rangina Ahmad[1] · Dominik Siemon[2] · Ulrich Gnewuch[3] · Susanne Robra-Bissantz[1]

Abstract

Millions of people experience mental health issues each year, increasing the necessity for health-related services. One emerging technology with the potential to help address the resulting shortage in health care providers and other barriers to treatment access are conversational agents (CAs). CAs are software-based systems designed to interact with humans through natural language. However, CAs do not live up to their full potential yet because they are unable to capture dynamic human behavior to an adequate extent to provide responses tailored to users' personalities. To address this problem, we conducted a design science research (DSR) project to design personality-adaptive conversational agents (PACAs). Following an iterative and multi-step approach, we derive and formulate six design principles for PACAs for the domain of mental health care. The results of our evaluation with psychologists and psychiatrists suggest that PACAs can be a promising source of mental health support. With our design principles, we contribute to the body of design knowledge for CAs and provide guidance for practitioners who intend to design PACAs. Instantiating the principles may improve interaction with users who seek support for mental health issues.

Keywords Personality-adaptive conversational agent · Chatbot · Virtual assistant · Mental health · Artificial intelligence · Design science research

1 Introduction

The necessity for mental health-related services has increased rapidly (Luxton, 2020). The prevalence of mental health issues (such as anxiety, depression or loneliness) is increasing among people worldwide: One in four people in the world is affected by mental health issues at some point in their lives (WHO, 2017). Access to therapists is limited, resulting in an acute shortage of mental health workers globally (Ta et al., 2020; WHO, 2021). Waiting times to receive treatment, therefore, can profoundly affect a person's quality of life (Luxton, 2014). Further barriers to treatment access include high costs and unequal access to health care resources, but also attitudinal barriers, such as stigma toward professional treatments and skepticism about the effectiveness of treatment (Wasil et al., 2021). Technological advancements, however, have improved access to mental health care services (Jones et al., 2014). Conversational agents (CAs) represent one emerging technology with the potential to help address the barriers that contribute to the unmet health care needs (Luxton, 2020). CAs are software-based systems designed to interact with humans through natural language (Feine et al., 2019). CA is the overarching and general term for software that interacts with users via written or spoken natural language, and includes systems such as chatbots, which provide text-based communication, and virtual

✉ Rangina Ahmad
  rangina.ahmad@tu-braunschweig.de

  Dominik Siemon
  dominik.siemon@lut.fi

  Ulrich Gnewuch
  ulrich.gnewuch@kit.edu

  Susanne Robra-Bissantz
  s.robra-bissantz@tu-braunschweig.de

[1] Chair of Information Management, Institute of Business Information Systems, Technische Universität Braunschweig, Mühlenpfordtstraße 23, 38106 Braunschweig, Germany

[2] Department of Software Engineering, School of Engineering Science, LUT University, Mukkulankatu 19, 15210 Lahti, Finland

[3] Institute of Information Systems and Marketing, Karlsruhe Institute of Technology (KIT), Kaiserstraße 89-93, 76133 Karlsruhe, Germany

assistants, which rely primarily on voice-based interaction (Diederich et al., 2022). Applying Artificial intelligence (AI)-based methods, such as machine learning and natural language processing (NLP), CAs have rapidly become more efficient and enable broad and often cross-topic conversations (Diederich et al., 2022; Nißen et al., 2021). A benefit of CAs is that they are accessible anywhere and available at any time to provide counseling and deliver therapeutic interventions, and they may therefore be a promising source of support helping people cope with mental health issues (Luxton, 2020; Ta et al., 2020). A recent review by Abd-Alrazaq et al. (2021) demonstrates that the usefulness level of mental health care CAs is perceived as high by patients and that they overall have positive perceptions of and opinions about the CAs. Though AI-based CAs are increasingly capable of handling highly complex tasks, such as social support and therapeutic decision-making in a human-like way, currently available CAs, however, do not live up to their full potential yet (Fiske et al., 2019; Graham et al., 2019). Capturing dynamic human behavior to an adequate extent, and providing responses and reactions tailored to users' individual contexts, special interaction dynamics and personalities, still pose a challenge in designing CAs (Grudin & Jacques, 2019; McTear et al., 2016). The field of tailored health communication has long established the need to personalize conversation (Abd-Alrazaq et al., 2021; Smith et al., 2015), since language is a primary tool to understand patients' experiences and express therapeutic interventions (Laranjo et al., 2018). Psychotherapy, in particular, is highly patient-centered in clinical practice, requiring skills such as observing patient behavior and adapting to their individual personalities and needs accordingly (Graham et al., 2019; Laranjo et al., 2018). In their interaction with individuals, contemporary CAs lack the ability to dynamically change their own personality to adapt to the user's personality. Provided a CA has features normally associated with humans, such as the use of natural language or human-like appearance (Nass & Moon, 2000), human beings tend to treat computer systems as social entities and ascribe different personality traits to them (Nass et al., 1994). Thus, as individuals' interactions with computers have a fundamental social nature, users feel more appreciated and comfortable interacting with the machines when they perceive CAs as more human-like (Moon & Nass, 1996; Nass et al., 1993). One way to ensure that a CA is perceived as human-like, is to provide the CA with the ability to dynamically change its personality traits. Since personality preferences differ from person to person (McCrae & Costa Jr, 1997), the personalization of CA conversation is highly important and required (Abd-alrazaq et al., 2019). Based on

an established personality model, and with advances in NLP, we propose the concept of a personality-adaptive conversational agent (PACA). PACAs are able to recognize and express personality by automatically inferring personality traits from users, giving them the ability to adapt to the changing needs and states of users when establishing a personalized interaction with them. As personality differences are manifested in language use, engagement with users can be further enhanced through tailored conversation styles (Kocaballi et al., 2020). While there is a large body of descriptive knowledge on design elements or cues that can be adapted, there is a lack of prescriptive knowledge on ways in which to actually design PACAs. Therefore, we pose the following research question (RQ):

RQ: How to design personality-adaptive conversational agents (PACAs) to improve interaction with users in mental health care?

To answer our RQ, we conduct a research project following a design science research (DSR) approach (Hevner et al., 2004). Our design approach is particularly anchored in two existing kernel theories, namely the five factor model of personality (McCrae & John, 1992) and the 'computers are social actors' paradigm (Nass et al., 1994). On the basis of these theories, we believe that PACAs have the potential to improve interaction with users by personalizing their health care (Luxton, 2014). To the best of our knowledge, there is no study that rigorously derives requirements from both literature issues and user stories to develop design principles, and that further evaluates the preliminary design principles by means of expert interviews, in order to propose an expository instantiation that translates the abstract knowledge, captured in the design principles, into applicable knowledge. Our results indicate that practitioners (e.g., designers, developers, etc.) who instantiate our design principles to create a PACA, receive a promising level of guidance to improve interaction with users in mental health care.

The remainder of this paper is structured as follows: In the section on related work, we give a brief overview of the history and current developments of mental health care CAs. We also provide an overview of current research and highlight the critical aspects in the use of mental health care CAs. In section three, we illustrate the theoretical foundations of our DSR project. Section four presents our research methodology in more detail. Section five contains our derived and evaluated design principles for PACAs, as well as a demonstration of an expository instantiation. Finally, we discuss our results, current limitations, and contributions of our work, and close with a conclusion.

2 Related Work

Efforts to develop software-based systems within the health care environment have a long history. In fact, ELIZA – widely considered the first CA in history – first appeared in 1966 in a psychotherapeutic context (Weizenbaum, 1966). Though only developed for demonstration and not commercial purposes, the simple computer program was able to mimic a Rogerian psychotherapist communicating with people in an empathic manner (Weizenbaum, 1966). The interaction resulted in people ascribing human characteristics to it, and some psychiatrists seeing ELIZA's potential computer-based therapy as a "form of psychological treatment" (Kerr, 2003, p. 305). The underlying technology of ELIZA was rather simple: By searching the textual input of its conversation partner for relevant keywords, the machine produced appropriate responses according to rules and directions based on scripts handcrafted by the programmers (Natale, 2018; Peters, 2013). PARRY, another early example of a prototype CA developed in 1979, was designed to simulate and behave like a person with paranoid schizophrenia (Shum et al., 2018). The developers' intention was to find out if other psychiatrists could determine a real paranoid patient from their computer model (Shah et al., 2016). PARRY was a rule-based CA and worked in a similar way as ELIZA, though with better language understanding capabilities (Shum et al., 2018).

Current interest in CAs for mental health care is also nascent, as can be seen in the growing number of online services offered by health care providers (Bendig et al., 2019). More than one-fourth of 15,000 mobile health apps focus on mental health diagnosis or support, according to the World Health Organization (Abd-Alrazaq et al., 2021). Over the last few years, three particularly prominent therapeutic mental health CAs based on AI technologies have emerged: Woebot (Woebot Health, 2021), Wysa (Wysa, 2021) and Tess (X2, 2021). These CAs are publicly available mobile phone applications aimed at helping people to manage symptoms of anxiety and depression by providing counseling services. Woebot is a CA that is built to assess, monitor and respond to users dealing with mental health issues (Woebot Health, 2021). The CA has a responsive way to intervene with in-the-moment help and provide targeted therapies based on cognitive behavioral therapy (Woebot Health, 2021). The AI chatbot Wysa is also based on cognitive behavioral therapy, but employs several other methods, such as behavioral reinforcement and mindfulness, to help clients with depression (D'Alfonso, 2020; Wysa, 2021). According to its developers, Wysa provides 24/7 high-quality mental health support (Wysa, 2021). The mental health chatbot Tess pursues a similar approach by being available 24/7 and delivering conversations in the same way that "a coach or friend would" (X2, 2021). Preliminary studies of the efficacy of all three applications have shown a significant reduction in depression and anxiety levels in the group of participants using the CAs (D'Alfonso, 2020). Though the presented CAs were developed to have specific personalities, they are not able to dynamically change these or to infer user personality in order to be more adaptive towards the users' needs. Therefore, all three CAs represent a rather "one-size-fits-all" solution, by not adequately adapting to the specificities of their users.

Ever since research in the field of human-machine interaction stressed the importance of avoiding "one-size-fits-all" interactions, the customization or personalization of CAs to individual users have become an important research topic (Abd-Alrazaq et al., 2021). Though this topic is still in its infancy (Kocaballi et al., 2019), there has been an increased interest in studies addressing the aspect of personality adaptivity in CAs. Studies from Yorita et al. (2019), Kampman et al. (2019) and Ahmad et al. (2020a, b) show that it is technically feasible to develop a CA that adapts its own personality traits to match the identified traits of the user. Yorita et al. (2019) developed their CA specifically with the purpose of providing emotional support for the user. While the studies by Yorita et al. (2019), Kampman et al. (2019) and Ahmad et al. (2020a, b) primarily focus on text-based human-machine interaction, the work of Völkel et al. presents approaches to adapt a voice assistant's personality to the user, in order to improve interaction experience (Völkel et al., 2020, 2021). Further studies by Ranjbartabar et al. (2018) and Zalake (2020) particularly deal with CAs that, among other factors, aim to adapt to user personality to reduce study stress (Ranjbartabar et al., 2018) or to promote anxiety coping strategies among college students (Zalake, 2020). The research of Wibhowo and Sanjaya (2021) concentrate on CAs for use in clinical psychology and psychotherapy: The authors developed a CA as a warning system to prevent individuals with borderline personality disorder from committing suicide. The previous studies mainly employ empirical methods to describe behaviors in interaction with these systems and consequently generate descriptive knowledge about the use and effectiveness of personality-adaptivity CAs. In addition, these studies all have different focuses, so there is still a research gap in providing prescriptive knowledge about how to design a PACA to improve interactions with users in mental health care.

Due to technological advancements, AI-based CAs have become increasingly capable of handling highly complex tasks with human qualities such as a higher autonomy of decision-making (Brendel et al., 2021) or expressing human-like feelings (Porra et al., 2019). Consequently, the application of CAs can have diverse impacts on individuals – both

positive and negative. The field of mental health care raises ethical considerations by its very nature (D'Alfonso, 2020). Voices have therefore emerged from the research field asking for a reassessment of the potential "dark sides" of AI and the ethical responsibilities of developers and designers (Brendel et al., 2021; Porra et al., 2019). Critics specifically stress the caveats of creating and perfecting human-like CAs with simulated feelings, without considering long-term consequences for human beings, such as deep, emotional attachments (Ahmad et al., 2021a). However, human beings treat computer systems as social entities and ascribe different personality traits to them (Nass et al., 1994); hence, they feel more appreciated and comfortable interacting with the machines when they perceive CAs as more human-like (Moon & Nass, 1996; Nass et al., 1993). To ensure an authentic interaction experience, CAs therefore should be imbued with some degree of human-like features (Gnewuch et al., 2017). While advanced CAs are able to simulate conversations employing therapeutic techniques, it is "not on the near horizon" (D'Alfonso, 2020, p. 113) for CAs to replicate human therapists. In fact, researchers agree that mental health care CAs should be used primarily as support systems, since the interaction experience and relationship that develops between a therapist and a patient is considered a significant factor in the outcome of psychological therapy (D'Alfonso, 2020) and cannot easily be substituted by a machine. The role of CAs in mental health care, rather, is to address individuals in need of treatment who are not receiving any treatment at all due to various barriers (Bendig et al., 2019; Stieger et al., 2018). In this way CAs could provide low-threshold access to mental health care but also bridge the waiting time before approval of psychotherapy (Bendig et al., 2019; Grünzig et al., 2018). Mental health care CAs have the potential to create their own form of interaction experience with users, provided the CA responses to users are tailored to their individual personalities. As a result, a personalized interaction would improve health outcomes among care seekers (Luxton, 2014). Given the widespread interest in mental health care CAs and the lack of prescriptive knowledge on how to design PACAs, it is important to address this research gap for researchers, designers, developers and clinicians.

3 Theoretical Background

3.1 Computers Are Social Actors

Extant research has shown that humans treat computers as social actors (Nass et al., 1994). The underlying reason, according to the 'computers are social actors' paradigm, is that humans automatically respond to social cues from computers (e.g., human-like avatars) in ways similar to how they would respond to social cues from another person

(e.g., facial expressions, gestures) (Nass et al., 1994). This behavior can also be observed when users interact with CAs (Schuetzler et al., 2018). The fundamental reason is that CAs engage with users in a uniquely human activity, that is, having a conversation in natural language. According to Fogg (2002), interactive language use is one of the most salient social cues triggering social responses from users. In addition, CAs can display many other social cues, such as human-like avatars, names, vocalization or gestures (Feine et al., 2019). Research has shown that when users respond to these social cues from a CA, they perceive it as more human-like and feel more comfortable interacting with it (Moon & Nass, 1996; Nass et al., 1993). Studies have also shown that incorporating social cues can positively influence various CA-related outcomes, such as trust, enjoyment, and satisfaction (Kocaballi et al., 2019; Liu & Picard, 2005; Schuetzler et al., 2018). And since human beings treat computer systems as social entities and ascribe different personality traits to them (Nass et al., 1994), they perceive CAs as more human-like, particularly when the CAs' expressions are attuned to the users' state (Liu & Picard, 2005). For example, depending on the strength of a CA's language, on the expressed confidence level, as well as on the interaction order, participants attribute an extraverted or introverted personality to the CA (Moon & Nass, 1996; Nass et al., 1995). Current CAs are designed to have a predefined personality (e.g., extraverted) and thus are not able to dynamically change their personality traits based on who they are interacting with. However, research has shown that users' individual personality differences interact with the CA's personality (Al-Natour et al., 2005; Gnewuch et al., 2020). As a result, different users may prefer different CA personalities, indicating that personality-adaptive CAs may be a sensible choice to meet user needs and preferences.

3.2 The Five Factor Model of Personality

Personality is loosely defined as the construct that differentiates individuals from each other, but at the same time makes a human being's behavior, thoughts and feelings (relatively) consistent (Allport, 1961). The dispositional approach considers trait as the key concept of the field of personality. In order to measure an individual's personality, a widely used classification of personality – the five factor model or "Big Five" – has been applied in research (McCrae & John, 1992). Compared to other existing personality models, this multifactorial model was found to be stable across cultures and observers, and provides a taxonomy for the systematic evaluation of individuals (Goldberg, 1993; McCrae & John, 1992). The five fundamental traits that have been identified in this context are openness (to experience), neuroticism (also known as emotional range), conscientiousness, agreeableness, and extraversion (McCrae & John, 1992).

The five factor approach to the taxonomy of traits is based on natural language, and more precisely lexical resources (Bouchet & Sansonnet, 2012). The lexical hypothesis states that most socially relevant and salient personality characteristics have become encoded in the natural language (Boyd & Pennebaker, 2017). Thus, language (verbal and non-verbal) has been found to be a fundamental dimension of personality (Boyd & Pennebaker, 2017). Human language reflects the psychological state and personality based on the frequency with which certain categories of words are used, as well as on the variations in word usage (Boyd & Pennebaker, 2017; Golbeck et al., 2011; Yarkoni, 2010). Psychologists have documented the existence of such cues by discovering correlations between a range of linguistic variables and personality traits across a wide range of linguistic levels (Mairesse et al., 2007). Language use has furthermore been scientifically proven to be unique, relatively reliable over time, and internally consistent, and as Boyd and Pennebaker (2017, p. 63) further state: "Language-based measures of personality can be useful for capturing/modeling lower-level personality processes that are more closely associated with important objective behavioral outcomes than traditional personality measures." In addition to a speaker's semantic content, utterances convey a great deal of information about the speaker, and the more extreme a person's personality trait, the more consistently that trait will be a factor in their behavior (Mairesse & Walker, 2010). For example, extraverts have been found to have a higher rate of speech, to speak more, louder, and more repeatedly, with fewer hesitations and pauses, have higher verbal output, and use less formal language, whereas people who are highly agreeable show a lot of empathy, agree and compliment more, use longer words and many insight words, and make fewer personal attacks on their interlocutor (Mairesse & Walker, 2010). The combination of personality-specific words that people use in everyday life are internally consistent, vary considerably from person to person and is predictive of a wide range of behaviors (Boyd & Pennebaker, 2017; Pennebaker, 2011). Current personality mining services apply AI, specifically NLP technologies, to automatically infer personality traits from an individual's speech or text (Ferrucci, 2012). NLP techniques particularly subserve to the interpretation of massive volumes of natural language elements by recognizing grammatical rules (e.g., syntax, context, usage patterns) of a word, sentence or document (Ferrucci, 2012).

4 Methodology

To generate design knowledge about the design of PACAs, we follow the design science research (DSR) approach proposed by Hevner et al. (2004). The overall goal of DSR projects is to generate rigorously derived and relevant design knowledge. Design knowledge thereby covers all aspects of the relationship between the problem and the solution space (Venable, 2006; Vom Brocke et al., 2020), and aims to create prescriptive knowledge that contributes to both the theory and practice of solving real-world problems (Hevner, 2020). By providing prescriptive solution-oriented design knowledge, DSR therefore represents a complement to descriptive research that explains the nature of a phenomenon or problem. Gregor and Hevner (2013) define descriptive knowledge as omega-knowledge, which includes "descriptions of natural, artificial, and human-related phenomena" and is "composed of observations, classifications, measurements, and the cataloging of these descriptions into accessible forms" (p. A2). In contrast, DSR also creates prescriptive knowledge about artifacts that address the phenomenon or problem (Gregor et al., 2007), which Gregor and Hevner (2013) define as lambda-knowledge. DSR should not only generate original solutions to existing problems, but also demonstrate the beneficial application of such solutions to the problem and show their broad implications (Baskerville & Pries-Heje, 2010; Vom Brocke et al., 2020), thus generating both omega and lambda-knowledge (Hevner, 2020). Prescriptive design knowledge (lambda-knowledge) can be represented in different ways, such as design principles (DPs), design features, or instantiations (Hevner, 2020; Vom Brocke et al., 2020), whereas descriptive knowledge (omega-knowledge) describes observations, measurements or classifications of phenomena and sense-making, such as patterns or natural laws (Gregor & Hevner, 2013). Hence, DSR projects involve an interplay of synergies between descriptive and prescriptive knowledge, as well as the use of existing knowledge and the production of new knowledge (Hevner, 2020).

DPs as a form of formalized knowledge are gaining popularity in the field of information systems (IS) because they allow researchers to capture abstract or meta-knowledge that addresses a class of problems, rather than a single problem (Gregor et al., 2020; Iivari et al., 2021; Purao et al., 2020). To ensure practical relevance and applicability, DPs should imply accessibility, effectiveness, and, most importantly, guidance for action (Iivari et al., 2021). In our research, we follow an iterative and multi-step process that is fundamentally embedded in the three-cycle view of DSR (i.e., relevance cycle, rigor cycle and design cycle) (Hevner, 2007), to generate design knowledge for the design of PACAs for mental health care. The relevance cycle incorporates issues from the problem space in the study and places outcomes from the design cycle in evaluation and practice. The rigor cycle involves the use of kernel theories and scientific methods, as well as research experience and expertise, in the study, while also adding new knowledge created by the researchers to the growing knowledge base. The design cycle, as the core of DSR, includes construction activities for the development

and evaluation of artifacts and design processes (Hevner, 2007). In our DSR project, we generate design knowledge in the form of DPs and an expository instantiation, by involving all three cycles multiple times. We follow the first strategy as proposed by Iivari (2015), in which we construct a meta-artefact (i.e., DPs) as a general solution concept, and then further adapt and follow the methodological components proposed by Möller et al. (2020) for the development of DPs in IS. These authors propose a supportive approach as well as a reflective approach that differ within "the point of artifact design and the logic of generating design principles" (p. 214). The two approaches of Möller et al. (2020) are similar to the two general strategies proposed by Iivari (2015) in that either generalizable design knowledge is created first, which is then concretized, or concrete design knowledge (e.g., design features or instantiations) is created first, which is then abstracted and generalized. However, the approach of Möller et al. (2020) focuses specifically on the construction of DPs and proposes a methodological procedure. We follow the supportive approach, in which DPs are "the provision of design knowledge in advance to support the design of an artifact before the design process takes place" (p. 214). Figure 1 shows our DSR approach based on Hevner (2007), Iivari (2015) and Möller et al. (2020).

Our DSR approach is divided into two main phases, addressing the problem space and the solution space (including the evaluation of the solution), which in turn consist of three sub-steps, respectively. To generate and evaluate a solution in the form of DPs for the identified problems, we use a mapping diagram. Mapping diagrams help to visualize the connection and derivation logic between DPs and meta-requirements, as well as between the problem space and the solution space (Möller et al., 2020). In this way, connections between and derivations of the individual aspects become clearer. The mapping diagram of the derivation and construction of the MRs and DPs can be found in the Appendix (see Fig. 4). An interplay of synergies between the rigor cycle, relevance cycle, and design cycle is reflected in each sub-step.

Step 1: Problem Space

An important prerequisite for effective and practical design knowledge is a good understanding and description of the underlying problem space. To conceptualize the problem space, Maedche et al. (2019) propose considering and analyzing four key concepts of the problem space; stakeholders, needs, goals, and requirements. In our approach, the process of understanding and defining the problem space consists of three steps addressing the needs of the stakeholders and issues from the application domain, which we subsequently captured in requirements. To conceptualize the problem space, we followed three steps that build upon each other:

a) Identifying issues from the application domain (relevance cycle), literature and kernel theories (rigor cycle)
b) Deriving user stories (relevance cycle) through a qualitative survey (rigor cycle)
c) Deriving meta-requirements (design cycle) through user stories, literature issues, application domain aspects, and kernel theories

We approached the problem space by reviewing extant literature and identifying current problems in mental health

Fig. 1 Research approach based on Iivari (2015), Möller et al. (2020) and Hevner (2007)
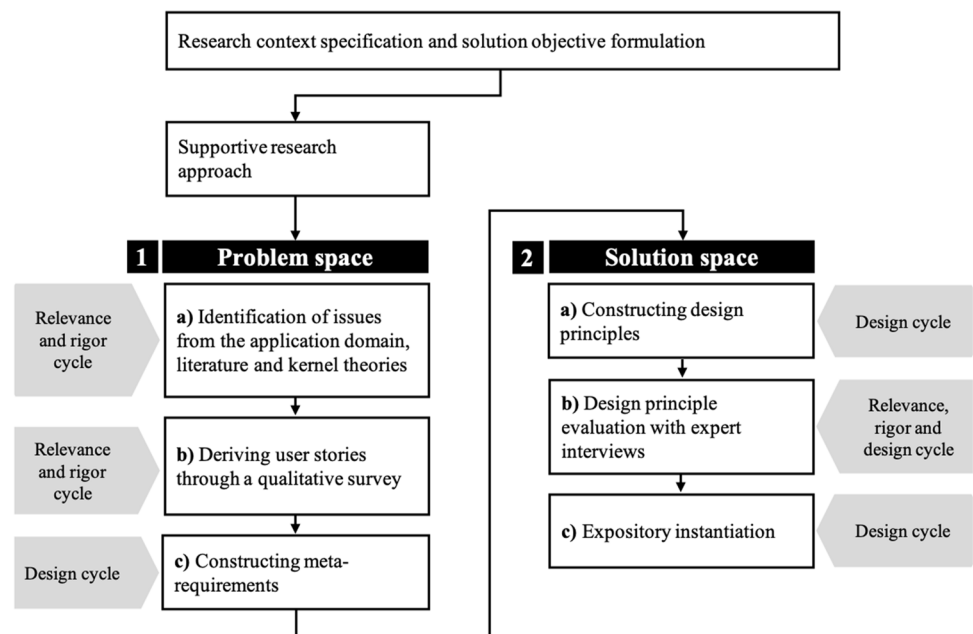
care. In a next step, and based on our kernel theories, we conducted a user survey to capture the perspective of potential users and to derive meta-requirements together with the literature issues. We followed an explorative approach and conducted a qualitative study (Babbie, 2020) using an open questionnaire with the aim to capture comprehensive user stories about PACAs in the context of mental health care. Our open questionnaire consisted first of an extensive explanation of the functionality and nature of a PACA to make sure participants from all backgrounds understood the concept. After that, we used an open questionnaire to capture user stories about the design of PACAs in the context of mental health care. These open questions asked for general requirements (e.g., mental health support, safety and privacy) and design-related requirements (e.g., behavior, communication style) (see Table 1 for an excerpt from the survey with sample responses). To help participants visualize what a conversation between a patient and a PACA might look like, we created a predefined chat record. For our simulated dialogue, we used the conversational design tool Botsociety, which allows prototyping and visualizing CAs. The conversation was provided in the form of a video. Figure 2 shows a mock-up of the simulated interaction. The video with the entire conversation can be viewed here: https://youtu.be/-sfSNJwCCI0

The survey was distributed via our private network and the crowdsourcing platform Mechanical Turk (mTurk) and was carried out in December 2020. A total of 60 respondents participated in the study, producing more than 6865 words of qualitative data answering the open questions, which took roughly between 25 and 40 min to complete. Table 1 shows an excerpt from the survey with four open questions and example responses that were used for qualitative content analysis.

Participants (32 male, 28 female) were between 23 and 71 years old, with an average age of 36 years. In order to

analyze the data, we followed a qualitative content analysis by coding the answers of the participants, which consisted mainly of inductive category forming (Mayring, 2014). The authors conducted the coding process independently, and whenever the results differed, we discussed the points of disagreement until we reached a consensus. In the last sub-step, we constructed meta-requirements from the captured user stories, the issues from the literature and the application domain.

Step 2: Solution Space and Evaluation

Our proposed solution to the problem space identified in step 1 is a PACA for mental health care, for which we present design knowledge in the form of DPs. We constructed our DPs based on the meta-requirements, and then evaluated them with experts from the application domain. For the systematic derivation and evaluation of our solution space, we followed the following steps:

a) Constructing design principles (design cycle).
b) Evaluating design principles (design cycle) with expert (relevance cycle) interviews (rigor cycle).
c) Designing an expository instantiation (design cycle).

Our DPs are formulated based on the approach proposed by Gregor et al. (2020, p. 2), who defined the anatomy of design principles so that DPs are "understandable and useful in real-world design contexts." The authors point out the importance of including the concerned actors when formulating the DPs to complement the anatomy, with the aim that DPs should be "prescriptive statements that show how to do something to achieve a goal" (Gregor et al., 2020, p. 2). The anatomy of a DP consists of the aim, implementer, and user; context; mechanism; and rationale, and is presented in the form of heuristic rules. The

Table 1 Excerpt from the survey with sample responses

Question	Example response
Do you think the concept of a PACA - with the computer system having a personality and being able to adapt to the user's personality - is useful/ helpful in mental health therapy?	*"Yeah I think it makes you feel less like you're just sharing your emotions with an inanimate object, it makes you feel like there is some more meaning to sharing it with something that can at least pretend to care. It makes it feel more worthwhile to have it respond with something other than just a planned response."*
Please comment briefly, why you think a specific trait is important in your opinion.	*"Being talkative will help in communicating feelings and affection is a trait that people with mental problems struggle to receive and comprehend."*
In your opinion, in which of the roles should a mental healthcare PACA slip into? Please explain briefly why.	*"Looking at mental states means an emotional state of mind. A friend or companion are the closest to a patient and are therefore necessary."*
What are, in your opinion, reasons that speak against communicating with a PACA? What concerns would you have in your interaction with the PACA?	*"I suppose it raises questions or suspicions about who could be reading or observing your conversations with the PACA. How secure and confidential is your information in the short and long term. It's a question of trust."*

Fig. 2 Mock-Up of the conversation between Raffi (PACA) and Jules (User)

evaluation of constructed artifacts such as DPs is an essential step in the design cycle to generate rigorous design knowledge. To evaluate our constructed DPs, we chose a qualitative approach in the form of expert interviews. To ensure relevance to the application domain, we selected experts, such as psychologists and psychiatrists, and had them evaluate the constructed DPs from a psychotherapist's perspective. The criteria for selecting experts were a minimum of at least two years of professional experience in the care or therapy of mentally ill people, as well as a completed apprenticeship or studies in medicine, psychology, or another related field. The experts were selected and approached from our personal network. The work of EX1 (a psychologist) and EX6 (a social worker and therapist) particularly focus on youth care, while EX2 to EX5 are all psychiatrists who trained as psychotherapists and work with patients of all ages. EX2 deals specifically with geriatric psychiatry. The six experts were interviewed between March and April 2021. Each interview took between 50 and 80 min. Table 2 shows an overview of the interview panel, including the experts' education and professional background.

The interviews were conducted with the support of a semi-structured interview guideline (Mayring, 2014). Interviews that follow a guideline are based on pre-defined questions that help orient the interviewer and thus ensure that all critical aspects of the scope of the interview are included. In addition, queries and deviations by the experts are possible (Mayring, 2014). The interview guide started with general questions about the profession of the experts, their professional experience, and their experience in dealing with CAs. This was followed by a detailed explanation of the concept of a PACA and follow-up questions by the interviewer to ensure that the concept was understood. Thereupon, all DPs were examined individually and evaluated by the experts, who were asked to recapitulate each DP in their own words, to again check whether they understood each DP. The experts were then asked to comment on each DP in terms of its relevance and to add possible missing aspects or highlight particularly relevant aspects. Subsequently, the audio recordings were transcribed and coded using MaxQDA (version 2020) qualitative data analysis software. The results were incorporated in the further development of the DPs. Despite the anatomy of a DP and different frameworks for

Table 2 Expert panel for the DP evaluation

ID	Gender	Age	Education	Professional Background
EX1	Female	31	Studied psychology and has five years of working experience.	Works in the in-patient youth welfare in a therapeutic residential school with patients with mental disorders.
EX2	Male	38	Studied human medicine, specialized in neurology and psychiatry. Has nine years of working experience.	Works in a private clinic for psychiatry and psychotherapy (geriatric psychiatry) as a senior physician.
EX3	Female	29	Studied human medicine, completed an apprenticeship in psychotherapy. Has four years of experience.	Works as a medical specialist in psychiatry and psychotherapy in an outpatient and inpatient facility.
EX4	Male	28	Studied human medicine, completed an apprenticeship in psychiatry. Has two years of experience.	Works in a private clinic for psychiatry and psychotherapy.
EX5	Female	32	Studied human medicine, specialized in psychiatry. Has six years of experience.	Works in a private clinic for psychiatry and psychotherapy as a senior physician.
EX6	Female	35	Studied social work, has completed an apprenticeship as a therapist. Has eight years of working experience.	Works in residential youth care in a therapeutic residential home with patients with mental disorders.

the accumulation of design knowledge (Rothe et al., 2020), the reusability and use of DPs for practitioners is often challenging. Iivari et al. (2021) address this problem and specifically propose strengthening the actability and guidance for the implementation of DPs to maintain the practical relevance of DSR. Since DPs represent formalized design knowledge, we design an expository instantiation (Gregor et al., 2007) that translates the abstract knowledge captured in DPs into applicable knowledge. The next section presents the result of the two steps and their sub-steps.

5 Designing PACAs for Mental Health Care

5.1 Deriving Design Principles for PACAs

Within our problem space, we first identified several issues specific to our application domain (AD), as well as current literature issues (LI) related to CAs. The issues relating to the AD consequently establish the relevance of our research due to societal needs (i.e., unmet mental health care). The LIs reveal current issues and related theories of CAs regarding, for example, their interaction capabilities, which are captured in the literature. In contrast to the issues of the AD, the LIs originate from the knowledge base and show us, for example, why CAs have limited conversational ability. Both thus contribute to the relevance and rigor of our DSR project.

The World Health Organization reports that one in four people in the world is affected by mental health issues at some point in their lives (WHO, 2017). Particularly in times of humanitarian crisis, the necessity for health-related services increases rapidly (Luxton, 2020;Torous et al., 2020; WHO, 2021) **(AD1)**. As a result of this increase, there is a serious shortage of mental health workers globally (9 per 100,000 population), which in turn contributes to unmet health care needs (Luxton, 2020; Prakash & Das, 2020)

(AD2). Thus, there is a high need for offering IT-based mental health services to surpass the availability of healthcare workers and ease the burden on them. In addition to the high prevalence of mental health issues, there is a strong social stigma attached to mental illness (WHO, 2017). Therefore, patients with mental health issues are considered particularly vulnerable. Studies also show that the personality of a user plays a crucial role in the adoption of emerging technology that raises concerns about data security and privacy (Junglas et al., 2008). Since the field of mental health handles highly sensitive data, paying attention to the user's personality is important. Hence, if patient safety is not addressed appropriately, a lack of privacy mechanisms, and thus a loss of trust, and could cause harm to people who exhibit sensitive mental health conditions (Luxton, 2020) **(AD3)**. Although CAs are considered an emerging technology with the potential to help address these issues, the adoption of CAs in mental health care has been slower than expected (Graham et al., 2020). Psychotherapy is a highly patient-centered clinical practice. This means that a successful conversation is particularly dependent on each patient's individual dynamic behavior and the therapist's ability to adapt to the patient's specific personality in order to form a therapeutic relationship (Laranjo et al., 2018; Luxton, 2014). However, contemporary CAs do not capture user personality and individual dynamic human behavior to an adequate extent (Ahmad et al., 2021b; Yorita et al., 2019) **(LI1)**. As a result, the CA's knowledge of the patient's personality and behavior is restricted, and their ability to effectively adapt and consequently to develop rapport with the patient is limited. The needs and preferences of users while interacting can be fundamentally different. However, many CAs are focused on a "one size fits all"-approach instead of pursuing personalized communication. Contemporary CAs therefore insufficiently tailor responses to patients' individual contexts, special interaction dynamics, and personality (Grudin & Jacques, 2019; McTear et al., 2016) **(LI2)**. This in turn results in

decreased communication satisfaction, user engagement, and dialogue quality (Kocaballi et al., 2019). Another literature issue concerns current CAs' limited ability to hold longer conversations or answer more complex questions (Chakrabarti & Luger, 2015; Clark et al., 2019). Although many CAs are increasingly capable of handling highly complex tasks with their natural language understanding capabilities, CAs are not sophisticated enough to recreate the richness of a conversation with a human therapist (Gaffney et al., 2019) **(LI3)**. As a consequence, CAs fail to sufficiently engage the patients and to respond with suitable advice (Kocaballi et al., 2019). A further identified literature issue is a CA's degree of anthropomorphism: The perceived level of humanness can either have a positive or a negative impact on users (Prakash & Das, 2020) **(LI4)**. Hence, the degree of anthropomorphism of a CA can affect a patient's judgments and attitudes (Kim et al., 2019), which can lead to the uncanny valley phenomenon - a feeling of uncanniness when CAs become too human-like (Kim et al., 2019). Table 3 sums up all ADs and LIs.

Our underlying kernel theories – the five factor model and the 'computers are social actors' paradigm – are reflected in the user stories (US) derived from our qualitative study (see 1.b) of Fig. 1 and Table 1 for example responses). When the 60 participants recruited using mTurk and from our private network were asked to describe their ideal PACA for mental health care, they ascribed the virtual therapist with similar social cues and characteristics as they would have a human therapist. We aggregated the most common and repetitive USs into three categories: support, safety, and behavior. The category *support* sums up the ways in which the PACA could actively support the patient by improving its mental health service. The USs include around the clock availability **(US1)**, an easy conversation while interacting **(US2, US3)**, memorized conversations **(US4)**, a personalized interaction experience **(US5)**, competence **(US6)**, support and helpfulness for mental health therapy **(US7, US8)** and the ability to refer patients to human therapists **(US9)**. The category *safety* contains USs in which many participants stated that they want the PACA to be trustworthy **(US10)**, so they can build a relationship with the PACA **(US11)**. A

basic requirement for this would be that the patient's data should be secure and not misused by the PACA **(US12)**. The participants further indicated that there should be an option where the PACA-patient conversation is monitored by a human therapist **(US13)** so that the patient does not become dependent on the PACA and/or gets desocialized **(US14)**. A PACA should also be able to notice warning signals **(US15)**. The third category, *behavior*, refers to the characteristics that an ideal mental health care PACA should have, according to our participants. Although many of the participants agreed that a PACA should be and act rather human-like **(US16)**, as well as be able to communicate via voice and facial expressions **(US17)**, their preferences concerning the PACA's communication style and personality traits varied strongly between participants. **US18-US28** in Table Table 4 sums up the aggravated PACA characteristics that were mentioned at least twice by various participants.

Based on our understanding of the problem space (i.e., AD, LI, US), we derived seven meta-requirements (MR) that ultimately led to six DPs for PACAs. After our expert evaluation, we adjusted the wording and refined the principles following the structure of Gregor et al. (2020). We elaborate on the MRs, as well as on the derived DPs, in the following section.

Mental health issues are increasing among people worldwide (AD1) and there is a dramatic shortage of mental health professionals (AD2), which is why users asked for an easily accessible and always available option (US1) to support them with their mental health issues (US3). To address this lack of human support for people in need, PACAs should always be available for their users, to put them at ease at any time **(MR1)**. However, even when the PACA is available 24/7, not all users may actively seek help when they need it the most (AD3), because mental illness is still often deemphasized or stigmatized. Users also stated that they want a supportive and motivating PACA (US7) that supports them proactively (US24). To prevent users suffering in silence, PACAs should therefore take the initiative to reach out to their users on a regular basis **(MR2)**. Based on these two MRs, we thus propose:

Table 3 Derived issues from application domain & literature

#	Application Domain (AD) and Literature Issues (LI)
AD1	Humanitarian crisis: Increased necessity for health-related services (Torous et al., 2020; WHO, 2021)
AD2	Shortage in health care providers and therapists (Luxton, 2020; 2020; WHO, 2021)
AD3	Patient safety: sensitive data and privacy concerns (Luxton, 2020, WHO, 2017)
LI1	Inability to adapt dynamically to a user's personality (Ahmad et al., 2021b; Yorita et al., 2019)
LI2	No tailored responses to a user's preferred communication style (Kocaballi et al., 2019)
LI3	Limited conversational ability (Chakrabarti & Luger, 2015; Clark et al., 2019)
LI4	Lacking adaptability to a user's preferred degree of CA humanness (Kim et al., 2019)

Table 4 Derived user stories from qualitative study

Category	#	User Story (US)
Support	US1	I want the PACA to be accessible and available for me 24/7.
	US2	I want the PACA to be easy to talk to like a friend/pen-pal/therapist.
	US3	I want the PACA to be able put me at ease.
	US4	I want the PACA to memorize our conversations.
	US5	I want a personalized interaction experience with the PACA.
	US6	I want the PACA to be competent.
	US7	I want the PACA to be supportive and motivate me.
	US8	I want the PACA to be helpful for my mental health therapy.
	US9	I want the PACA to be able to refer me to a human therapist.
Safety	US10	I want the PACA to be trustworthy.
	US11	I want to be able to build a relationship with the PACA.
	US12	I want my data to be secure and not misused.
	US13	I want the PACA to be able to be monitored by my therapist.
	US14	I don't want to get desocialized and/or become dependent on the PACA.
	US15	I want the PACA to notice warning signals.
Behavior	US16	I want the PACA to be and act human-like.
	US17	I want the PACA to be able to have a voice and have facial expressions.
	US18	I (don't) want the PACA to use emojis.
	US19	I want the PACA to have a rich vocabulary.
	US20	I (don't) want the PACA to be formal.
	US21	I want the PACA to be calm, patient and receptive.
	US22	I want the PACA to be kind and polite.
	US23	I want the PACA to be humorous, witty and curious.
	US24	I want the PACA to be proactive and interactive.
	US25	I want the PACA to be confident and firm.
	US26	I want the PACA to be empathetic, sensitive and caring
	US27	I want the PACA to be open minded and attentive.
	US28	I want the PACA to be flexible and adaptive.

DP1: Principle of Proactive Support

For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support patients independently of therapist, location and time, ensure that the PACA is accessible 24/7 and proactively checks in on the user on a regular basis, so they can receive support at any time.

Another MR we derived from the problem space is a PACA's communication competence, particularly when it comes to developing and building a long-term relationship with a patient to improve well-being. Currently, CAs have limited conversational abilities (LI3), which is why users expressed their wishes that PACAs should be competent communication partners (US6) that can memorize past conversations (US4), have a rich vocabulary (US19), and are overall helpful for a patient's mental health therapy (US8) by finding supportive and motivating words for the patient (US7) (**MR3**). We therefore propose:

DP2: Principle of Competence

For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support users independently of therapist, location and time, provide the PACA with a domain-specific knowledge base and incorporate therapeutic techniques, so the user feels understood and perceives the PACA as competent.

The fourth MR specifically addresses patient safety (AD3). PACAs should protect patient data and always respect the users' privacy, in ways similar to how human therapists build therapeutic relationships (US11) by being a trusted source of mental health support (US10) and by remaining strictly confidential (US12). Other safety requirements for a PACA are that it should support its users but not make them dependent on it (US14). Besides, a PACA should recognize critical trigger words (e.g., suicidal thoughts) and involve a human therapist (US9, US13). PACAs also need to analyze their patients' data to learn more about their individual needs, preferences and personalities, to give tailored responses (LI2). In this

regard, the PACA should be highly transparent towards its users (**MR4**). Hence, we propose:

DP3: Principle of Transparency
For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support patients independently of therapist, location and time, ensure that the PACA is transparent in communicating patient safety and privacy issues, so that users trust the PACA with their health-related concerns and feel safe when sharing sensitive data.

One important aspect that users mentioned frequently and couched differently, is the PACA's social role (US2). Different people in a patient's life can provide social support, including therapists, friends, and family members. The responses in our study showed that different people preferred different types of social roles for the PACA (US2): While some participants indicated that their ideal PACA should, for example, take on the role of a witty and humorous friend (US23) who interacts in a more informal manner (US20), others described their PACA as a virtual therapist who should be more confident and firm (US25), and more formal in appearance (US20). Therefore, PACAs should be able to take on a social role (e.g., friend, therapist) based on the user's individual needs and preferences (**MR5**). We consequently propose:

DP4: Principle of Social Role
For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support users independently of therapist, location and time, allow users to choose between different social roles so the PACA can take on the user's preferred social role and adapt to their needs.

Today, CAs can largely be customized only by user input. Customization is also mostly limited to simple and external aspects of a CA (LI4). However, the perceived level of humanness, which can be expressed through verbal and non-verbal language, can have either a positive or negative influence on users (LI4). A large number of our participants, though, stated that they wish to have a personalized interaction with their PACA (US5) and want their PACA to be and act human-like (US16). They further expressed their wish to interact with a PACA that can communicate non-verbally via voice and/or facial expressions (US17). Therefore, adapting the degree of adaptable anthropomorphism of a PACA, as well as further options of communication via voice and/or facial/body expressions, is required **(MR6)**. We therefore propose:

DP5: Principle of Anthropomorphism
For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support users independently of therapist, location and time, allow the users to choose what type of PACA they want to interact with (chatbot, voice assistant, embodied conversational agent), so they can determine the PACA's degree of anthropomorphism based on individual needs.

Successful psychotherapy depends on the therapist's ability to adapt to the patient's specific personality to form a therapeutic relationship (LI1–2). However, current CAs follow a "one size fits all" approach instead of pursuing personalized communication. This is in line with our study participants' requirements, that is, a personalized interaction experience with the PACA (US5). In addition, users may have vastly different preferences for communication with a PACA. This observation aligns with the responses from our user stories (US18-US28). For example, while some participants preferred texting with an extraverted and witty PACA (US23), others indicated they would rather trust an introverted PACA that communicates in a calm and soothing voice (US21). In summary, our participants stated that they not only wish a personalized interaction (US5) but also a flexible and adaptive PACA (US28) Therefore, the personality of a PACA should be aligned with the users' preferences (**MR7**). We propose the following principle:

DP6: Principle of Personality Adaptivity
For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support users independently of therapist, location and time, imbue the PACA with language cues specific to different personality dimensions, to enable the PACA to adapt to the user's preferred communication style and increase interaction quality.

5.2 Evaluating Design Principles for PACAs

DP1: Principle of Proactive Support was rated as an important principle by all experts; both constant availability and regular, proactive check-ups were named as highly useful features for a PACA. EX5 and EX6 pointed out that around-the-clock availability and more regular check-ins are advantages that PACAs might have over human mental health professionals. As to whether the PACA should always act in a proactive manner, the experts almost unanimously stated that the majority of patients would most probably appreciate such a function. Specifically, patients with certain mental health conditions, such as anxiety disorders or depressions, are often not able to actively seek help, which is why a regular check-in by the PACA would be helpful. However,

EX4 also noted that for serious psychiatric illnesses, such as schizophrenia or paranoia, a proactive PACA might be too intrusive, giving the patient the feeling of being observed and monitored in a way which would be not beneficial for their mental health state. EX3 and EX5 further indicated that the patient, or even their treating therapist, should be able to select the intervals at which the PACA should become proactive, as too much or too little intervention can influence the quality of the support. We therefore slightly adjusted the description for DP1 and now mention that the PACA should check in on the user on a *mutually agreed* regular basis (see Table 5).

The experts considered *DP2: Principle of Competence* as equally important as DP1, stating that a PACA that is used in a mental health care context most definitely needs to have domain-specific knowledge and should be able to use therapeutic techniques in conversation with the patient. EX3, EX4 and EX5 remarked that the inclusion of 'therapeutic techniques' does not necessarily mean incorporating techniques that are specific to certain schools of psychotherapy (e.g., cognitive behavioral therapy, psychodynamic psychotherapy) in the PACA, but rather that the PACA should able to understand the patient's individual concerns and then dynamically respond with suitable advice. However, four of the experts were slightly skeptical whether PACAs could be designed so that they are sophisticated enough to recreate the richness of a conversation with a human therapist. While EX2 and EX4 highlighted the importance of "reading between the lines" in order to react to warning signals, EX3 indicated that the PACA's way of communication should also be perceived as "authentic" by showing understanding, otherwise patients could easily get disappointed by the PACA's lack of perceived competence. The experts agreed with the formulation of DP2, and therefore we did not change the description.

All six experts considered *DP3: Principle of Transparency* a fundamental prerequisite for ensuring patient safety. According to EX1 and EX6, a PACA should be under a duty of professional secrecy equivalent to that of a healthcare professional. EX2 further stated that, in order to build trust with the patient, a PACA should be absolutely transparent about what happens with the patient's data, because, as EX4 further remarked, a PACA would otherwise "risk losing their patients' trust." EX5 pointed out that, although it is crucial to inform the patient about the medical confidentiality at the beginning of the very first session, it does not necessarily have to be an integral part of every session. However, a PACA should always be able to address the privacy terms when asked by the user. EX4 suggested reminding the user at regular intervals that writing with the PACA is a "safe space." Moreover, all experts highlighted the importance of building a steady therapeutic relationship, as rapport and trust can only be built over a longer period. To achieve this,

a PACA must be considered a safe space for the users. The wording of DP3 was accepted by the experts, hence nothing was changed in the description.

The preliminary version of *DP4: Principle of Social Role* was not formulated clearly enough, since the majority of experts needed some explanation. It was not intuitively clear what we meant by *social role*, so we added specific examples ("e.g., friend, therapist etc.") to the DP4 description. Once we mentioned the examples, the meaning was clear and the experts did not need further explanations. The principle was rated by the experts as valuable, however not as important as the previous DPs. EX3 and EX1 mentioned the importance of specific roles that people reflect for patients and approved the idea of having a PACA that reflects a specific role. EX4 and EX6 argued that, from a therapeutic perspective, it might not always be effective for the patients' therapy progress if the PACA continuously takes on the role of a friend who never "counters or addresses unpleasant topics." EX2 further indicated that specific roles or genders can be associated with fear or aggression. To avoid this, EX5 suggested that the PACA should be able to change its social role situationally, even within one session. EX2 further assumed that a patient who can choose between different social roles for their PACA would be more likely to use the service than if the option did not exist. After receiving this feedback, we elaborated on DP4 by adding that the user can also switch between certain social roles to promote therapy progress (see Table 5).

The experts rated *DP5: Principle of Anthropomorphism* as of similar importance as DP4, stating that it promotes better adaptability to the individual user. From a psychotherapeutic perspective, the principle can be specifically beneficial when patients "cannot talk about their issues, but rather prefer to write" (EX2), or when patients "have difficulty reading emotions from text, non-verbal language can help" (EX4). EX6 further stated that some patients need to feel a PACA's social presence, for example in the form of an embodied CA, to open up and feel comfortable. EX3, however, doubted the efficacy of this, as she was rather critical towards CAs that are too anthropomorphized, stating that it could lead to negative dependencies. She therefore indicated that DP5 must be viewed and designed with caution. EX1 remarked that the PACA's level of humanness might affect how patients perceive the PACA's competence. EX5 suggested that familiar voices can be helpful in crisis mode and thus can be considered a useful feature for the PACA. As DP5 was comprehensible and the experts agreed on the wording of the description, nothing was modified.

Concerning *DP6: Principle of Personality Adaptivity*, all experts pointed out that capturing a patient's dynamic behavior and individual personality during therapy is an essential step towards forming a trustful therapeutic relationship. EX2 and EX3 explained that building rapport with patients

usually takes several hours of therapy until the patients are able to slowly open up. In this context, EX4, EX2 and EX1 highlighted the importance of language, and stressed that the style in which they communicate with their patients, specifically, plays a crucial role in the quality of their interaction. All six experts stated that, in the context of mental health care, the most valuable feature with which to imbue a PACA is the ability to capture patient personality, just as a human therapist would do. EX3 stated that, although therapists have their own unique personality and therapeutic techniques, they very much act (to some degree) like a PACA by adapting to their patients' personalities. Therefore, DP6 was rated as particularly important. However, EX4 and EX5 remarked that, for therapeutic purposes, the PACA should also be able to change its personality and communication style to a "provocative style" (EX4) to "break the patient through their reserve from time to time" and to "not become a people pleaser" (EX5). EX4 added that a PACA should be able to dynamically change its personality, if the goal is to achieve therapeutic progress. To make DP6 more comprehensible, we changed "preferred communication style" to "personality" (see Table 5). Table 5 summarizes the revised and final six DPs.

In general, the experts agreed that the six DPs cover all important criteria for designing a PACA for mental health

care. They saw potential in designing PACAs to address the issues from the ADs; however, they also strongly emphasized that experts in psychotherapy should be involved in the design of mental health care PACAs. They further indicated that a user's experienced level of mental health issue(s) can play an important role in whether a PACA can be effective or not. The psychiatrists stated in particular that realistically, a PACA alone would not be able to help users with severe mental health issues, but could rather be a useful tool for both the patient and their treating therapist. Though they were familiar with the terms *chatbot* and *digital assistant*, the experts admitted that visualizing a PACA with therapist-like capabilities would only be possible if they could "experience [it] first-hand, to see how it really works" (EX2).

5.3 Expository Instantiation

To show the applicability of our DPs and provide guidance for the implementation of a PACA, we developed an expository instantiation (Gregor et al., 2007; Iivari et al., 2021). Design knowledge, especially DPs, tends to be highly abstract and consequently cannot always be implemented easily and directly. Therefore, we transformed our defined DPs into an expository instantiation that can assist in "representing the design knowledge both as an expository

Table 5 Evaluated design principles for PACAs

#	Title	Design Principles (DPs)
DP1	Principle of Proactive Support	For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support patients independently of therapist, location and time, ensure that the PACA is accessible 24/7 and proactively checks in on the user on a mutually agreed regular basis, so they can receive support at any time.
DP2	Principle of Competence	For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support users independently of therapist, location and time, provide the PACA with a domain-specific knowledge base and incorporate therapeutic techniques, so the user feels understood and perceives the PACA as competent.
DP3	Principle of Transparency	For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support patients independently of therapist, location and time, ensure that the PACA is transparent in communicating patient safety and privacy issues, so that users trust the PACA with their health-related concerns and feel safe when sharing sensitive data.
DP4	Principle of Social Role	For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support users independently of therapist, location and time, allow users to choose between different social roles (e.g., friend, therapist etc.) so the PACA can dynamically take on the user's preferred social role, but can also switch between social roles that promote the user's therapy progress.
DP5	Principle of Anthropomorphism	For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support users independently of therapist, location and time, allow the users to choose the type of PACA they want to interact with (chatbot, voice assistant, embodied conversational agent), so they can determine the PACA's degree of anthropomorphism based on individual needs.
DP6	Principle of Personality Adaptivity	For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support users independently of therapist, location and time, imbue the PACA with language cues specific to different personality dimensions to enable the PACA to adapt to the user's preferred personality and increase interaction quality.

device and for purposes of testing" (Gregor et al., 2007, p. 322). For this purpose, we modeled a PACA starting from the proposed DPs, so that system engineers and software developers have the abstract design knowledge of the DPs available in a transferable form. We opted for a graphical modeling for the specification, design, documentation and visualization of software components and interfaces to transfer our DPs in applicable form. In our expository instantiation, we decided to model the functionality of a PACA in a technology-independent representation. Since technologies change very quickly, especially in the field of AI (e.g., chatbot services), the model can serve as a blueprint for future solutions. Figure 3 shows the model diagram of a PACA for mental health support.

Interaction with the PACA takes place via a conversational user interface (CUI), which ensures continuous availability for the user and the possibility to proactively contact the user through push notifications, for example (DP1). Examples of typical CUIs include messaging applications, such as WhatsApp or Facebook Messenger, or a standalone service implemented on a website or in a mobile application. The CUI is also responsible for the representation of the PACA and its anthropomorphic features (DP5). To be able to display all cues desired by the user, the CUI should be able to provide both text and speech-based output and should be able to represent an embodied PACA (i.e., body language). Therefore, integration in a website or as a mobile application that allows a rich representation of a PACA is preferable. The bot service is responsible for the logical interaction with the user, and therefore is able to process natural language. In addition, the bot service can adapt its conversational design, which determines the style of communication and

personality adaptivity (DP6). The conversational design is usually created by a conversational designer, who should co-develop it with a therapist in the case of a PACA. The conversational design is guided by the application logic, which in turn draws on the database and the results of the personality mining service. This functionality could be realized with services such as Google Dialogflow, IBM Watson Assistant, Amazon Lex, or Wit.AI.

The application is the core of the PACA and contains the general logic of the system. It accesses information from the knowledge base (database) and the information from the personality insight service to decide how to interact with the user. The application specifies not only the social role (DP4), but also the type of communication (DP6) and the degree of the anthropomorphism (DP5), which is the core of its adaptability. Together with the knowledge base and the identified personality of the user, it represents the competence of the PACA (DP2), which is reflected in the conversational design of the bot service and the CUI. The application should be implemented as platform-independently as possible and be able to communicate with the database, with the personality mining service, and with the bot service, which is why technologies or programming languages such as JavaScript, Python, Scala, Ruby, or Go are suitable. The personality mining service is responsible for analyzing the user's messages and determining the personality traits of the user. It accesses the data that the user enters via the CUI, and analyzes it. The results are then passed on to the application. One approach is to analyze specific word usage and build statistical models that determine the personality traits of the user. The foundation of these procedures was laid by Pennebaker and Francis (1996) with the Linguistic Inquiry
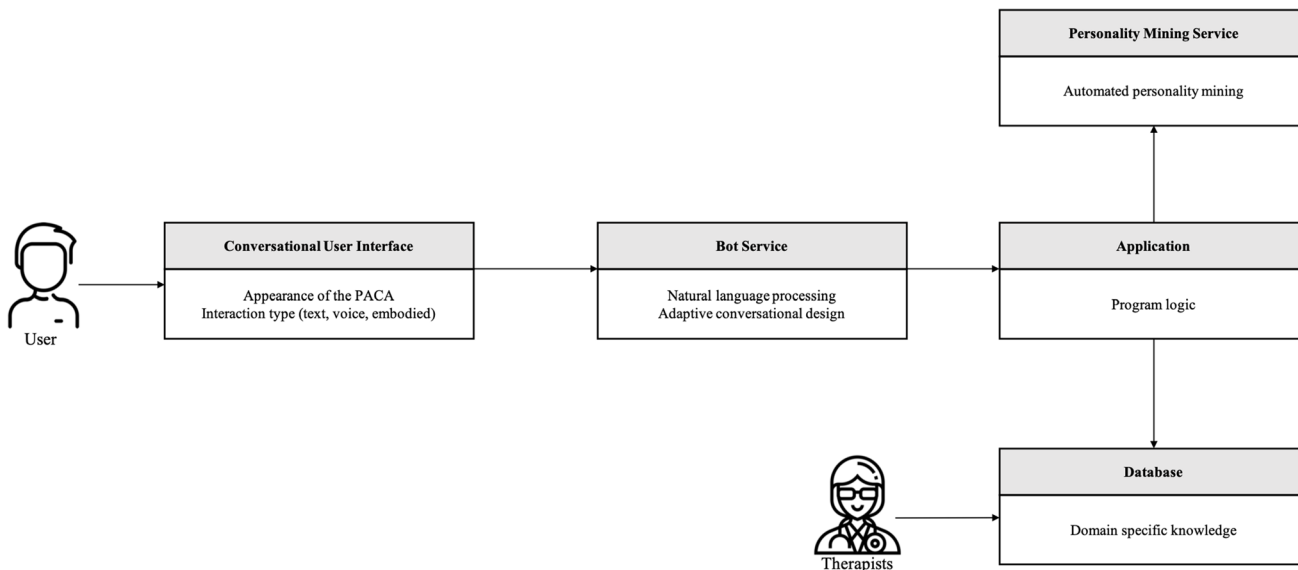

Fig. 3 Expository instantiation

and Word Count (LIWC), which quickly gained prominence in linguistic analysis (Chung & Pennebaker, 2012; Pennebaker et al., 2015). Further improvements have been achieved using so-called word embeddings, such as Google's word2vec or Stanford's GloVe (Pennington et al., 2014). The advantage of such models is that semantic similarity between words is determined unsupervised, whereas LIWC relies on human judges and psychologists to determine the meaning of words (Arnoux et al., 2017; Rice & Zorn, 2021). Arnoux et al. (2017) suggest that models with word embeddings predict personality even better than models based on LIWC. IBM developed Watson Personality Insights, a commercial software package, as a service for ready-to-use personality predictions based on GloVe and Twitter posts (Arnoux et al., 2017). These models or services can be used to determine the personality traits of the user from their interactions. With the user's personality traits identified, the PACA can adapt to the user according to the program logic (application).

The database contains the entire knowledge base of the PACA, ranging from communication forms, anthropomorphic expressions, and other cues to therapeutic methods and skills. The system also constantly learns from past interactions with all its users, as the application not only accesses the database, but also regularly updates it with information. The database also represents the learning aspect of a PACA, as it stores not only all therapeutic methods and knowledge bases, but also the knowledge gained from past interactions. For secure handling of the user's sensitive data, it is important that all data is protected with the latest technology (DP3), such as multiple factor authentication and data encryption. This should not only be ensured across all components, but also be communicated to the user via the bot services, so that they can build trust and rapport with the PACA (DP3). We previously published a paper that describes the implementation of a PACA within another application domain (Ahmad et al., 2020a).

6 Discussion

6.1 Theoretical and Practical Contributions

We contribute to the body of design knowledge for CAs by providing evaluated DPs for the design of a PACA. The six derived DPs can be divided into two categories. The first category contains *DP1: Principle of Proactive Support, DP2: Principle of Competence* and *DP3: Principle of*

Transparency. These three principles represent the foundation of a PACA and can be considered as basic prerequisites when designing a PACA. Proactive support, competence and transparency are particularly important design elements in the context of mental health care, as emphasized in our expert panel's reports. However, it is the second category of DPs that transforms a CA into a PACA: *DP4: Principle of Social Role*, *DP5: Principle of Anthropomorphism* and *DP6: Principle of Personality Adaptivity* are the principles that enable adaptation in the form of customization and personalization to user preferences, and consequently provide a more tailored service based on users' individual needs and personalities. As evaluated by the experts, DP6 in particular – which represents a necessary requirement for a PACA – is a crucial design element. The DPs do not build on each other and can be considered as stand-alone design elements. However, we believe that only a combination of all DPs provides the best possible results, since the six DPs cover all important criteria to design a PACA for mental health care.

Our work also offers several practical contributions. We generated design knowledge in the form of prescriptive knowledge, and we provide guidance for CA designers and developers. Based on our experts' evaluation, we argue that instantiating our DPs to design a PACA should improve interaction with users who seek support for their mental health care. While it is unlikely that PACAs that mimic human interactions could ever completely replace human psychotherapists, they may be a promising source of support for a wide range of user groups and different situations. First, for users who want to receive social support as an everyday social interaction to reduce loneliness or prevent early mental health issues, PACAs can surpass availability of and ease the burden on health care providers by being accessible to provide counseling at any time. Especially in the current COVID-19 pandemic, the need for social support is of immense importance, and the lack of such support can otherwise profoundly affect a person's quality of life. Second, for users who are undergoing therapy for more severe mental health issues and are receiving treatment from human professionals, a PACA may also be beneficial by delivering added value to therapeutic interventions. Since PACAs are not susceptible to forgetfulness or fatigue, they can be used as additional support systems for both the patient and the provider, by offering periodic check-ups, for example. In these cases, however, it is crucial for the PACA to be monitored by human professionals to ensure patient safety, as proposed by our experts.

6.2 Limitations and Ethical Considerations

A number of limitations have to be considered with respect to our qualitative study and expert interviews. A number of our study participants indicated that they have no experience with mental health related therapies. Hence, the USs we gathered would probably have led to different requirements if all participants had a similar amount of experience with mental health related therapies. However, as our intention was to include a non-clinical population, that is people who might want to receive social support as an everyday social interaction, we did not exclude participants from our study who indicated that they have no experience with mental health therapies. Another limitation is the small number of experts who evaluated the DPs. All our experts have less than 10 years of experience in their field and are about the same age. Experts with significantly more years of experience (e.g., < 20 years) might have evaluated the DPs differently.

Despite the cost, reach, and scalability advantages of CAs over human counsellors and therapists, it is important to note that there may be several drawbacks to using CAs in mental health care. First, these CAs could cause users to isolate themselves even more from the outside world. Since it can be easier to establish a relationship with these CAs than with another human (e.g., a therapist), users may lose interest in meeting or spending time with with their human friends and family members (Skjuve et al., 2021). This is even more problematic as many CAs are deliberately designed to look and act like humans (e.g., by giving them names and avatars), which further blurs the line between humans and machines (Porra et al., 2019). Second, these CAs are developed and operated by companies with business interests. Therefore, some users worry that the sensitive information they present in their conversations with the CAs could be intentionally or unintentionally shared with third parties (Bae Brandtzæg et al., 2021). Finally, as users have no control over their own CA, they become very dependent on the company that operates the CA, and some are afraid that their CAs could be deleted at some point (Skjuve et al., 2021). When taking all these considerations together, it is important to keep these drawbacks in mind when designing PACAs. Nevertheless, we believe that the potential advantages outweigh the potential drawbacks when these CAs are designed appropriately.

7 Conclusion

Though being a fruitful area with large practical potential, the adoption of CAs for mental health care is associated with some challenges. Literature issues involve the inability of current CAs to capture and adapt dynamically to user personality, as well as to provide responses and reactions tailored in accordance with users' individual characteristics to an adequate extent. In addition, CAs do not live up to their full potential yet, as they have been shown to have limited conversational abilities, particularly when it comes to longer and more complex interactions. These issues, however, are important factors that need to be taken into consideration when addressing primary issues from the application domain, such as the increased necessity for health-related services, the shortage of health care providers, or the paramount importance of patient safety. Motivated by the lack of design knowledge for PACAs, we proposed and evaluated DPs for designing PACAs that improve interaction, specifically for users who seek support for their mental health. We focused on two steps to answer our research question: Based on our kernel theories we first identified current issues from the AD and LIs, as well as derived USs, through a qualitative study. Within this first step of our problem space, we then derived MRs. On the basis of step one, and as part of our solution space, we proposed DPs for PACAs in step two. We then conducted expert interviews with psychologists and psychiatrists to evaluate the derived DPs, and adjusted and refined our final DPs based on their feedback. In a last step, we transferred our defined DPs to an expository instantiation for the purpose of better visualization.

According to the DSR contribution framework proposed by Gregor and Hevner (2013), our work can be classified as an improvement, as we address a known problem with a new solution. We provide prescriptive design knowledge by deriving and evaluating DPs for PACAs in mental health care. Our DPs contribute to the body of design knowledge for CAs and provide guidance for practitioners, such as designers, developers, and mental health organizations, on how to design PACAs that can better support their users. Instantiating these principles may improve interaction with users who seek support for mental health issues. We believe that our design approach could also be a valuable starting point for the design of PACAs in other domains.
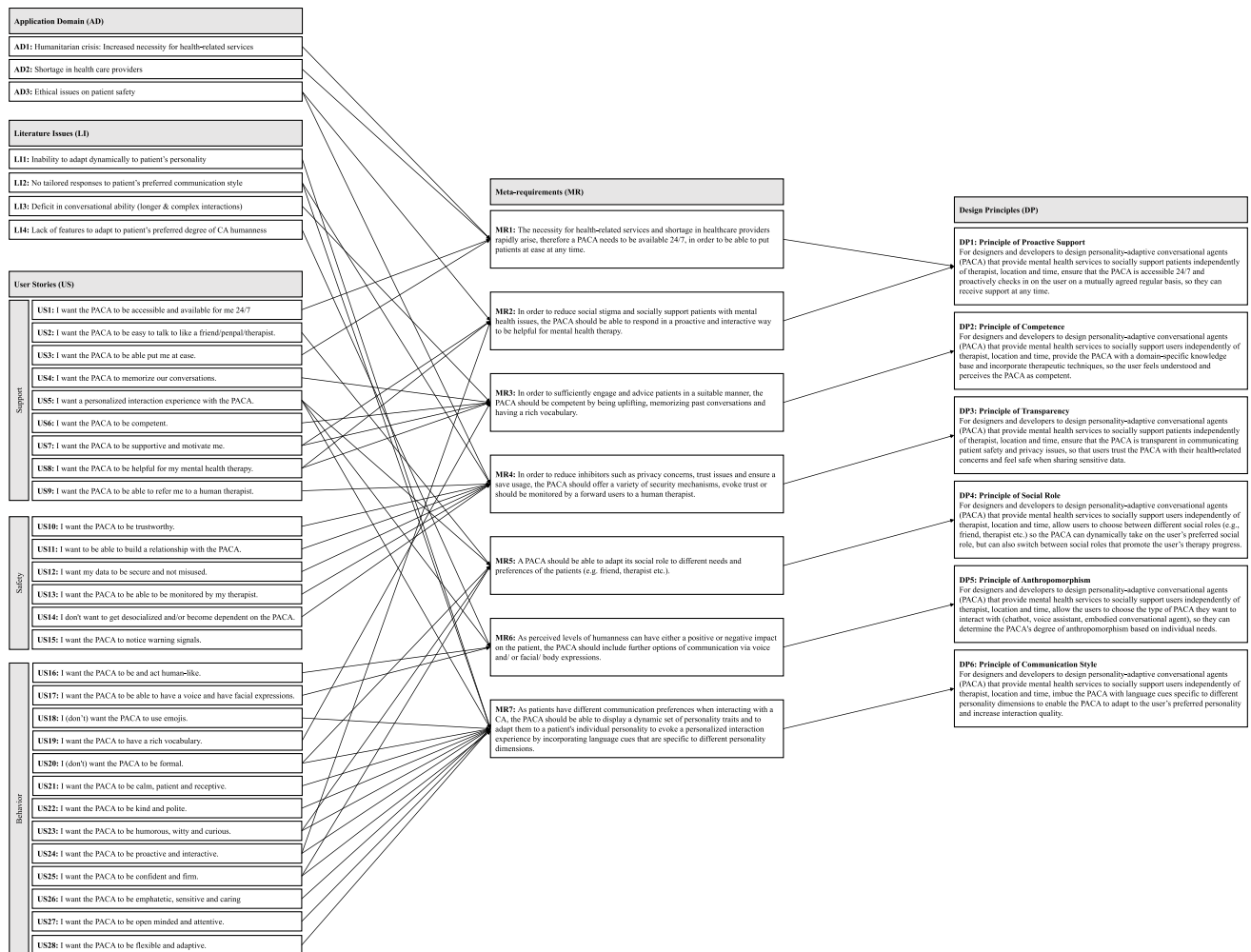
Appendix

Figure 4

Application Domain (AD)

AD1: Humanitarian crisis: Increased necessity for health-related services

AD2: Shortage in health care providers

AD3: Ethical issues on patient safety

Literature Issues (LI)

LI1: Inability to adapt dynamically to patient's personality

LI2: No tailored responses to patient's preferred communication style

LI3: Deficit in conversational ability (longer & complex interactions)

LI4: Lack of features to adapt to patient's preferred degree of CA humanness

User Stories (US)

Support

US1: I want the PACA to be accessible and available for me 24/7

US2: I want the PACA to be easy to talk to like a friend/penpal/therapist.

US3: I want the PACA to be able put me at ease.

US4: I want the PACA to memorize our conversations.

US5: I want a personalized interaction experience with the PACA.

US6: I want the PACA to be competent.

US7: I want the PACA to be supportive and motivate me.

US8: I want the PACA to be helpful for my mental health therapy.

US9: I want the PACA to be able to refer me to a human therapist.

Safety

US10: I want the PACA to be trustworthy.

US11: I want to be able to build a relationship with the PACA.

US12: I want my data to be secure and not misused.

US13: I want the PACA to be able to be monitored by my therapist.

US14: I don't want to get desocialized and/or become dependent on the PACA.

US15: I want the PACA to notice warning signals.

Behavior

US16: I want the PACA to be and act human-like.

US17: I want the PACA to be able to have a voice and have facial expressions.

US18: I (don't) want the PACA to use emojis.

US19: I want the PACA to have a rich vocabulary.

US20: I (don't) want the PACA to be formal.

US21: I want the PACA to be calm, patient and receptive.

US22: I want the PACA to be kind and polite.

US23: I want the PACA to be humorous, witty and curious.

US24: I want the PACA to be proactive and interactive.

US25: I want the PACA to be confident and firm.

US26: I want the PACA to be emphatetic, sensitive and caring

US27: I want the PACA to be open minded and attentive.

US28: I want the PACA to be flexible and adaptive.

Meta-requirements (MR)

MR1: The necessity for health-related services and shortage in healthcare providers rapidly arise, therefore a PACA needs to be available 24/7, in order to be able to put patients at ease at any time.

MR2: In order to reduce social stigma and socially support patients with mental health issues, the PACA should be able to respond in a proactive and interactive way to be helpful for mental health therapy.

MR3: In order to sufficiently engage and advice patients in a suitable manner, the PACA should be competent by being uplifting, memorizing past conversations and having a rich vocabulary.

MR4: In order to reduce inhibitors such as privacy concerns, trust issues and ensure a save usage, the PACA should offer a variety of security mechanisms, evoke trust or should be monitored by a forward users to a human therapist.

MR5: A PACA should be able to adapt its social role to different needs and preferences of the patients (e.g. friend, therapist etc.).

MR6: As perceived levels of humanness can have either a positive or negative impact on the patient, the PACA should include further options of communication via voice and/ or facial/ body expressions.

MR7: As patients have different communication preferences when interacting with a CA, the PACA should be able to display a dynamic set of personality traits and to adapt them to a patient's individual personality to evoke a personalized interaction experience by incorporating language cues that are specific to different personality dimensions.

Design Principles (DP)

DP1: Principle of Proactive Support
For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support patients independently of therapist, location and time, ensure that the PACA is accessible 24/7 and proactively checks in on the user on a mutually agreed regular basis, so they can receive support at any time.

DP2: Principle of Competence
For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support users independently of therapist, location and time, provide the PACA with a domain-specific knowledge base and incorporate therapeutic techniques, so the user feels understood and perceives the PACA as competent.

DP3: Principle of Transparency
For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support patients independently of therapist, location and time, ensure that the PACA is transparent in communicating patient safety and privacy issues, so that users trust the PACA with their health-related concerns and feel safe when sharing sensitive data.

DP4: Principle of Social Role
For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support users independently of therapist, location and time, allow users to choose between different social roles (e.g., friend, therapist etc.) so the PACA can dynamically take on the type of PACA they want to interact with, but can also switch between social roles that promote the user's therapy progress.

DP5: Principle of Anthropomorphism
For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support users independently of therapist, location and time, allow the users to choose the type of PACA they want to interact with (chatbot, voice assistant, embodied conversational agent), so they can determine the PACA's degree of anthropomorphism based on individual needs.

DP6: Principle of Communication Style
For designers and developers to design personality-adaptive conversational agents (PACA) that provide mental health services to socially support users independently of therapist, location and time, imbue the PACA with language cues specific to different personality dimensions to enable the PACA to adapt to the user's preferred personality and increase interaction quality.

Fig. 4 Mapping diagram of the derivation and construction of the MRs and DPs (Möller et al., 2020)

Declarations

References

Abd-alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). An overview of the features of Chatbots in mental Health: A scoping review. *International Journal of Medical Informatics, 132*, 103978 s.

Abd-Alrazaq, A. A., Alajlani, M., Ali, N., Denecke, K., Bewick, B. M., & Househ, M. (2021). Perceptions and opinions of patients about mental Health Chatbots: Scoping review. *Journal of Medical Internet Research, 23*(1), e17828.

Ahmad, R., Siemon, D., Fernau, D., & Robra-Bissantz, S. (2020a). "Introducing" Raffi": A Personality Adaptive Conversational Agent.," in PACIS, p. 28.

Ahmad, R., Siemon, D., & Robra-Bissantz, S. (2020b). ExtraBot vs IntroBot: The Influence of Linguistic Cues on Communication Satisfaction. In B. B. Anderson, J. Thatcher, R. D. Meservy, K. Chudoba, K. J. Fadel, & S. Brown (Eds.), *26th Americas Conference on Information Systems, AMCIS 2020, Virtual Conference, August 15–17, 2020*. Association for Information Systems.

Ahmad, R., Siemon, D., Gnewuch, U., & Robra-Bissantz, S. (2021a). The benefits and caveats of personality-adaptive conversational agents in mental Health care. AMCIS.

Ahmad, R., Siemon, D., & Robra-Bissantz, S. (2021b). "Communicating with machines: Conversational agents with personality and the role of extraversion," in Proceedings of the 54th Hawaii International Conference on System Sciences, p. 4043.

Allport, G. W. (1961). Pattern and growth in personality.

Al-Natour, S., Benbasat, I., & Cenfetelli, R. T. (2005). "The Role of Similarity in E-Commerce Interactions: The Case of Online Shopping Assistants," SIGHCI 2005 Proceedings, p. 4.

Arnoux, P.-H., Xu, A., Boyette, N., Mahmud, J., Akkiraju, R., & Sinha, V. (2017). "25 Tweets to Know You: A New Model to Predict Personality with Social Media," in Proceedings of the International AAAI Conference on Web and Social Media (Vol. 11).

Babbie, E. R. (2020). The practice of social research, Cengage learning.

Bae Brandtzæg, P. B., Skjuve, M., Kristoffer Dysthe, K. K., & Følstad, A. (2021). "When the social becomes non-human: Young People's perception of social support in Chatbots," in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–13.

Baskerville, R., & Pries-Heje, J. (2010). Explanatory design theory. *Business & Information Systems Engineering, 2*(5), 271–282.

Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2019). *The next generation: Chatbots in clinical psychology and psychotherapy to Foster mental Health – A scoping review* (pp. 1–13). Karger Publishers.

Bouchet, F., & Sansonnet, J.-P. (2012). "Intelligent Agents with Personality: From Adjectives to Behavioral Schemes," in Cognitively Informed Intelligent Interfaces: Systems Design and Development, IGI Global, pp. 177–200.

Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: A new approach to personality in a digital world. *Current Opinion in Behavioral Sciences, 18*, 63–68.

Brendel, A. B., Mirbabaie, M., Lembcke, T.-B., & Hofeditz, L. (2021). Ethical management of artificial intelligence. *Sustainability, 13*(4), 1974.

Chakrabarti, C., & Luger, G. F. (2015). Artificial conversations for customer service chatter bots: Architecture, algorithms, and evaluation metrics. *Expert Systems with Applications, 42*(20), 6878–6897.

Chung, C. K., & Pennebaker, J. W. (2012). Linguistic Inquiry and Word Count (LIWC): Pronounced 'Luke,'... and Other Useful Facts. In *Applied Natural Language Processing: Identification, Investigation and Resolution* (pp. 206–229).

Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spillane, B., Gilmartin, E., Murad, C., Munteanu, C., Wade, V., & Cowan, B. R. (2019). "What Makes a Good Conversation?: Challenges in Designing Truly Conversational Agents," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow Scotland Uk: ACM, May 2, pp. 1–12.

D'Alfonso, S. (2020). AI in mental Health. *Current Opinion in Psychology, 36*, 112–117.

Diederich, S., Brendel, A., Morana, S., & Kolbe, L. (2022). On the design of and interaction with conversational agents: an organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems, 23*(1), 96–138.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies, 132*, 138–161.

Ferrucci, D. A. (2012). Introduction to 'This Is Watson. *IBM Journal of Research and Development, 56*(3.4), 1–1.

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research, 21*(5)). https://doi.org/10.2196/13216

Fogg, B. J. (2002). "Persuasive Technology: Using Computers to Change What We Think and Do," Ubiquity (2002:December), ACM New York, NY, USA, p. 2.

Gaffney, H., Mansell, W., & Tai, S. (2019). Conversational Agents in the Treatment of Mental Health Problems: Mixed-Method Systematic Review. *JMIR Mental Health, 6*(10), e14166.

Gnewuch, U., Morana, S., & Maedche, A. (2017). "Towards designing cooperative and social conversational agents for customer service," in Proceedings of the 38th International Conference on Information Systems (ICIS2017).

Gnewuch, U., Yu, M., & Maedche, A. (2020). "The Effect of Perceived Similarity in Dominance on Customer Self-Disclosure to Chatbots in Conversational Commerce," in Proceedings of the 28th European Conference on Information Systems (ECIS 2020).

Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). "Predicting Personality from Twitter," in Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference On, IEEE, pp. 149–156.

Goldberg, L. R. (1993). "The Structure of Phenotypic Personality Traits.," American Psychologist (48:1), American Psychological Association, p. 26.

Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H.-C., & Jeste, D. V. (2019). Artificial intelligence for mental Health and mental illnesses: An overview. *Current Psychiatry Reports, 21*(11), 116.

Graham, S. A., Lee, E. E., Jeste, D. V., Van Patten, R., Twamley, E. W., Nebeker, C., Yamada, Y., Kim, H.-C., & Depp, C. A. (2020). Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: A conceptual review. *Psychiatry Research, 284*, 112732.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Q, 37*(2), 337–356.

Gregor, S., Jones, D., & et al. (2007). The Anatomy of a Design Theory, Association for Information Systems.

Gregor, S., Chandra Kruse, L., & Seidel, S. (2020). Research perspectives: The anatomy of a design principle. *Journal of the Association for Information Systems, 21*(6), 2.

Grudin, J., & Jacques, R. (2019). "Chatbots, Humbots, and the quest for artificial general intelligence," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19, Glasgow, Scotland Uk: ACM Press, pp. 1–11.

Grünzig, S.-D., Baumeister, H., Bengel, J., Ebert, D., & Krämer, L. (2018). "Effectiveness and Acceptance of a Web-Based Depression Intervention during Waiting Time for Outpatient Psychotherapy: Study Protocol for a Randomized Controlled Trial," Trials (19:1), Springer, pp. 1–11.

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian Journal of Information Systems, 19*(2), 4.

Hevner, A. R. (2020). The duality of science: Knowledge in information systems research. *Journal of Information Technology*, 0268396220945714.

Hevner, A., March, S. T., Park, J., & Ram, S. (2004). Design science research in information systems. *MIS Quarterly, 28*(1), 75–105.

Iivari, J. (2015). Distinguishing and contrasting two strategies for design science research. *European Journal of Information Systems, 24*(1), 107–115.

Iivari, J., Hansen, M. R. P., & Haj-Bolouri, A. (2021). A proposal for minimum reusability evaluation of design principles. *European Journal of Information Systems, 30*(3), 286–303.

Jones, S. P., Patel, V., Saxena, S., Radcliffe, N., Ali Al-Marri, S., & Darzi, A. (2014). How google's 'ten things we know to be true'could guide the development of mental health mobile apps. *Health Affairs, 33*(9), 1603–1611.

Junglas, I. A., Johnson, N. A., & Spitzmüller, C. (2008). Personality traits and concern for privacy: An empirical study in the context of location-based services. *European Journal of Information Systems, 17*(4), 387–402.

Kampman, O., Siddique, F. B., Yang, Y., & Fung, P. (2019). Adapting a Virtual Agent to User Personality. In*Advanced Social Interaction with Agents* (pp. 111–118). Springer.

Kerr, I. R. (2003). Bots, babes and the Californication of commerce. *University of Ottawa Law and Technology Journal, 1*, 285.

Kim, S. Y., Schmitt, B. H., & Thalmann, N. M. (2019). Eliza in the Uncanny Valley: Anthropomorphizing consumer robots increases their perceived warmth but decreases liking. *Marketing Letters, 30*(1), 1–12.

Kocaballi, A. B., Berkovsky, S., Quiroz, J. C., Laranjo, L., Tong, H. L., Rezazadegan, D., Briatore, A., & Coiera, E. (2019). The personalization of conversational agents in Health care: Systematic review. *Journal of Medical Internet Research, 21*(11), e15360.

Kocaballi, A. B., Laranjo, L., Quiroz, J., Rezazadegan, D., Kocielnik, R., Clark, L., Liao, V., Park, S., Moore, R., & Miner, A. (2020). Conversational agents for Health and wellbeing.

Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S., & Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association, 25*(9), 1248–1258.

Liu, K., & Picard, R. W. (2005). "Embedded Empathy in Continuous, Interactive Health Assessment," in CHI Workshop on HCI Challenges in Health Assessment (Vol. 1), Citeseer, p. 3.

Luxton, D. D. (2014). Recommendations for the ethical use and Design of Artificial Intelligent Care Providers. *Artificial Intelligence in Medicine, 62*(1), 1–10.

Luxton, D. D. (2020). Ethical implications of conversational agents in global public Health. *Bulletin of the World Health Organization, 98*(4), 285.

Maedche, A., Gregor, S., Morana, S., & Feine, J. (2019). Conceptualization of the problem space in design science research. In*International Conference on Design Science Research in Information Systems and Technology* (pp. 18–31). Springer.

Mairesse, F., & Walker, M. A. (2010). Towards personality-based user adaptation: Psychologically informed stylistic language generation. *User Modeling and User-Adapted Interaction, 20*(3), 227–278.

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30*, 457–500.

Mayring, P. (2014). Qualitative content analysis: Theoretical Foundation, Basic Procedures and Software Solution.

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist, 52*(5), 509.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60*(2), 175–215.

McTear, M., Callejas, Z., & Griol, D. (2016). *The conversational Interface: Talking to smart devices*. Springer.

Möller, F., Guggenberger, T. M., & Otto, B. (2020). Towards a method for design principle development in information systems. In*International Conference on Design Science Research in Information Systems and Technology* (pp. 208–220). Springer.

Moon, Y., & Nass, C. (1996). How 'real' are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research, 23*(6), 651–674.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103.

Nass, C., Steuer, J., Tauber, E., & Reeder, H. (1993). "Anthropomorphism, Agency, and Ethopoeia: Computers as Social Actors," in INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems, CHI '93, New York, NY, USA: Association for Computing Machinery, April 1, pp. 111–112. https://doi.org/10.1145/259964.260137.

Nass, C., Steuer, J., & Tauber, E. R. (1994). "Computers are social actors," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, pp. 72–78.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies, 43*(2), 223–239.

Natale, S. (2018). If software is narrative: Joseph Weizenbaum, artificial intelligence and the biographies of ELIZA. *New Media & Society, 21*, 146144481880498.

Nißen, M. K., Selimi, D., Janssen, A., Cardona, D. R., Breitner, M. H., Kowatsch, T., & von Wangenheim, F. (2021). See you soon again, Chatbot? A design taxonomy to characterize user-Chatbot relationships with different time horizons. *Computers in Human Behavior*, 107043.

Pennebaker, J. W. (2011). *The secret life of pronouns: How our words reflect who we are*. Bloomsbury.

Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition & Emotion, 10*(6) Taylor & Francis, 601–626.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015.

Pennington, J., Socher, R., & Manning, C. (2014). "Glove: Global vectors for word representation," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543.

Peters, O. (2013). Critics of Digitalisation: Against the Tide: Warners, Sceptics, Scaremongers, Apocalypticists: 20 Portraits, Studien Und Berichte Der Arbeitsstelle Fernstudienforschung Der Carl von Ossietzky Universität Oldenburg, Oldenburg: BIS-Verlag der Carl von Ossietzky Universität Oldenburg.

Porra, J., Lacity, M., & Parks, M. S. (2019). "'Can computer based human-likeness endanger humanness?' – A philosophical and ethical perspective on digital assistants expressing feelings they Can't have"", Information Systems Frontiers.

Prakash, A. V., & Das, S. (2020). Intelligent conversational agents in mental healthcare services: A thematic analysis of user perceptions. *Pacific Asia Journal of the Association for Information Systems, 12*(2), 1.

Purao, S., Chandra Kruse, L., & Maedche, A. (2020). The origins of design principles: Where do… they all come from?

Ranjbartabar, H., Richards, D., Kutay, C., & Mascarenhas, S. (2018). "Sarah the virtual advisor to reduce study stress," In Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, pp. 1829–1831.

Rice, D. R., & Zorn, C. (2021). Corpus-based dictionaries for sentiment analysis of specialized vocabularies. *Political Science Research and Methods, 9*(1) Cambridge University Press, 20–35.

Rothe, H., Wessel, L., & Barquet, A. P. (2020). Accumulating design knowledge: A mechanisms-based approach. *Journal of the Association for Information Systems, 21*(3), 1.

Schuetzler, R., Grimes, G., Giboney, J., & Nunamaker, J. (2018). "The influence of conversational agents on socially desirable responding," Proceedings of the 51st Hawaii International Conference on System Sciences, pp. 283–292.

Shah, H., Warwick, K., Vallverdú, J., & Wu, D. (2016). Can machines talk? Comparison of Eliza with modern dialogue systems. *Computers in Human Behavior, 58*, 278–295.

Shum, H., He, X., & Li, D. (2018). From Eliza to XiaoIce: Challenges and opportunities with social Chatbots. *Frontiers of Information Technology & Electronic Engineering, 19*(1), 10–26.

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My Chatbot companion-a study of human-Chatbot relationships. *International Journal of Human-Computer Studies, 149*), Elsevier, 102601.

Smith, K., Masthoff, J., Tintarev, N., & Wendy, M. (2015). Adapting emotional support to personality for Carers experiencing. *Stress.* https://doi.org/10.13140/RG.2.1.3898.9929

Stieger, M., Nißen, M., Rüegger, D., Kowatsch, T., Flückiger, C., and Allemand, M. 2018. "PEACH, a smartphone-and conversational agent-based coaching intervention for intentional personality change: Study protocol of a randomized, wait-list controlled trial," BMC Psychology (6:1), Springer, pp. 1–15.

Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., & Loggarakis, A. (2020). User experiences of social support from companion Chatbots in everyday contexts: Thematic analysis. *Journal of Medical Internet Research, 22*(3), e16235.

Torous, J., Myrick, K. J., Rauseo-Ricupero, N., & Firth, J. (2020). Digital mental health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. *JMIR Mental Health, 7*(3), e18848. https://doi.org/10.2196/18848

Venable, J. (2006). "The Role of Theory and Theorising in Design Science Research," in Proceedings of the 1st International Conference on Design Science in Information Systems and Technology (DESRIST 2006), Citeseer, pp. 1–18.

Völkel, S. T., Kempf, P., & Hussmann, H. (2020). "Personalised chats with voice assistants: The user perspective," in Proceedings of the 2nd Conference on Conversational User Interfaces, pp. 1–4.

Völkel, S. T., Meindl, S., & Hussmann, H. (2021). "Manipulating and evaluating levels of personality perceptions of voice assistants through enactment-based dialogue design," in CUI 2021-3rd Conference on Conversational User Interfaces, pp. 1–12.

Vom Brocke, J., Winter, R., Hevner, A., & Maedche, A. (2020). Special issue editorial–accumulation and evolution of design knowledge in design science research: A journey through time and space. *Journal of the Association for Information Systems, 21*(3), 9.

Wasil, A. R., Palermo, E., Lorenzo-Luaces, L., & DeRubeis, R. (2021). Is there an app for that? A review of popular mental health and wellness apps, PsyArXiv. https://doi.org/10.31234/osf.io/su4ar

Weizenbaum, J. (1966). ELIZA—A Computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36–45.

WHO. (2017). *Depression and other common mental disorders: Global Health estimates*. World Health Organization.

WHO. (2021). "WHO Executive Board Stresses Need for Improved Response to Mental Health Impact of Public Health Emergencies". https://www.who.int/news/item/11-02-2021-who-executive-board-stresses-need-for-improved-response-to-mental-health-impact-of-public-health-emergencies. Accessed April 20, 2021.

Wibhowo, C., & Sanjaya, R. (2021). "Virtual assistant to suicide prevention in individuals with borderline personality disorder," in 2021 International Conference on Computer & Information Sciences (ICCOINS), IEEE, pp. 234–237.

Woebot Health. (2021). "Woebot". https://woebothealth.com/products-pipeline/. Accessed Feb 28, 2021.

Wysa, io. (2021). "Wysa". https://www.wysa.io/. Accessed Feb 28, 2021.

X2, A. (2021). "Tess". https://www.x2ai.com/. Accessed Feb 28, 2021.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality, 44*(3), 363–373.

Yorita, A., Egerton, S., Oakman, J., Chan, C., & Kubota, N. (2019). "Self-adapting Chatbot personalities for better peer support," in 2019 IEEE international conference on systems, Man and Cybernetics (SMC), IEEE, pp. 4094–4100.

Zalake, M. (2020). "Advisor: Agent-based intervention leveraging individual differences to support mental wellbeing of college students," in Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–8.

Rangina Ahmad received her B.Sc. and M.Sc. degrees in business information systems from the Technische Universität Braunschweig in Germany, where she is currently pursuing her Ph.D. degree. Since 2018, she has been a Research Associate at the Chair of Information Management at the Technische Universität Braunschweig. Her research focuses on topics such as human-AI interaction, personality psychology, and e-services. Her work has been published in leading information systems conferences, such as Americas Conference on Information Systems and Hawaii International Conference on Information Systems

Dominik Siemon is an Associate Professor with the Department of Software Engineering, Lappeenranta-Lahti University of Technology (LUT University), Finland. He studied business information systems at the Technische Universität Braunschweig in Germany, where he received his Dr. rer. pol. (Ph.D.) degree in business information systems. He worked as a Research Assistant and a Postdoctoral Researcher at the Technische Universität Braunschweig and a part-time Professor of business information systems and digital business at IU International University. His mainly design-oriented research in the field of information systems addresses collaboration and interaction with intelligent systems, conversational agents, innovation management, collaboration technology, and creativity. His work has been published in leading conferences, such as the International Conference on Information Systems, and in journals, such as Education and Information Technologies, AIS Transactions on Human-Computer Interaction, and the Communications of the Association for Information Systems

Ulrich Gnewuch is a Postdoctoral Researcher at the Institute of Information Systems and Marketing at Karlsruhe Institute of Technology (KIT), Germany. He studied Information Systems at the University of Mannheim and received his PhD degree from KIT. His research focuses on the design of conversational user interfaces and digital assistants. His research has been published in journals such as the International Journal of Human-Computer Studies and Computers in Human Behavior. He serves as the vice chair for teaching resources of the AIS Special Interest Group on Human-Computer Interaction (SIGHCI)

Susanne Robra-Bissantz is Professor for Information Management and has been head of the Institute of Information Systems and the Chair of Information Management at the Technische Universität Braunschweig in Germany since 2007. After receiving her doctorate in Economic and Social Sciences, she worked as a Research Assistant and habilitated at the Chair of Business Administration at the Friedrich–Alexander University Erlangen–Nürnberg, Germany. As Vice President for Studies and Cooperation, she actively worked on new forms of teaching and examination and has implemented numerous third-party funded projects in cooperation with companies. Her design-oriented research focuses on holistic digital service development, e-collaboration and smart participation. She published her research at international conferences and in recognized journals