Designing Supportive Chatbots for Blood Donors

Helena M. Müller
Technical University of Darmstadt
mueller@is.tu-darmstadt.de

Melanie Reuter-Oppermann
University of Twente & Technical University of Darmstadt
m.n.reuter-oppermann@utwente.nl

Abstract

Healthcare systems worldwide rely on voluntary blood donations, as blood cannot be produced artificially, but is needed for many surgeries and treatments. In countries like South Africa or Ghana, currently less than 1 % of the population donates blood. Donor education, mobilisation and management are therefore crucial. Chatbots offer easy access to information for all types of donors as well as for blood services to educate (potential) donors. By applying the design science research approach and grounding our research on behaviour change models, we have developed a chatbot for all donor types in South Africa and Ghana. In this work, we present an instantiation of the chatbot and its positive evaluation with non-, first-time, lapsed and regular donors of both countries.

Keywords: Chatbot, Design science, Blood donation

1. Introduction

Blood products are an essential resource in healthcare systems worldwide that cannot be produced artificially, but only have a short shelf-life. Sufficient and timely donations as well as an efficient use of donations are crucial. Within the BISKIT project, an information and decision support system for blood supply chain management in South Africa and Ghana is developed (Horstkemper et al., 2021). In these countries, less than 1 % of the population donates blood (SANBS, 2023) and fulfilling the average daily demand for blood products is already challenging (Vermeulen et al., 2019). In Ghana, many donations are made by first-time donors who often do not return due to missing information on when and where to donate again (Asamoah-Akuoko et al., 2021). According to Western Cape Blood Service, same is true in South Africa, where additional donations due to education programs at schools decreased significantly because of the COVID-19 pandemic. Therefore, successful and efficient blood donor management is important. Sufficient new and first-time donors must be educated and motivated to voluntarily donate blood on a regular basis. The BISKIT tool aims to offer efficient ways to blood services to reach, interact with and provide information to all (potential) blood donors.

We argue that digital tools like chatbots can support educational programs when on-site visits are not possible and offer easy access to information, potentially reaching a vast majority of the population. Previous research has investigated the design of a chatbot for blood donors in Germany together with the potential benefits of offering such a chatbot to all donor types, non- (nd), first-time (fd), lapsed (ld) and regular donors (rd) (Müller and Reuter-Oppermann, 2022b). Diederich et al. (2019) have shown that inducing behaviour can be achieved by conversations with a chatbot when designed appropriately. Consequently, by applying the design science research (DSR) approach, we want to address the following research question with our work:

How to design chatbots to support potential blood donors and promote sustained blood donation behaviour?

In this paper, we focus on the user perspective on a chatbot that can be integrated into a website, for example, supporting behaviour change in a mainly passive way.

The remainder of this paper is structured as follows. In Section 2, we present the related work on existing blood donation chatbots. The following Section 3 summarises the design science research project. The design requirements, design principles and the instantiation of the chatbot are described in Section 4, followed by the user evaluation in Section 5. We close with a summary and an outlook on future research in Section 6.

2. Chatbots targeting blood donation

From its start in the 1960s with the rule-based psychotherapeutic chatbot ELIZA (Weizenbaum, 1966)

URI: https://hdl.handle.net/10125/106801
978-0-9981331-7-1

till now with the hype around ChatGPT embodying generative artificial intelligence, the interest in chatbots increased permanently. This is not only due to technological advances, but also due to the ease of use of chatbots enabling interaction through text messages in natural language. Thereby, they emulate human conversation even though there is a software program behind (Dale, 2016). Depending on their scope of application, chatbots can be used domain-specifically (e.g., in customer service) like ELIZA or as a general-purpose technology like ChatGPT (Gnewuch et al., 2017). They support users in searching for relevant information and take over simple tasks like booking an appointment through their easy access to available systems (Morana et al., 2017). Chatbots offer advantages to both, their providers and users, by running cost effectively with short resolution times on 24 hours a day seven days a week (Gnewuch et al., 2017). Along with their versatility through many use cases in various industries like banking and air travel, chatbots are spreading rapidly since they can be easily implemented on websites and messenger platforms, often used as massaging apps on smartphones.

With their roots in healthcare via ELIZA, one of the opportunities of chatbots is being perceived as anthropomorphic to give the feeling of a human contact (Verhagen et al., 2014). Following the "Computers Are Social Actors" paradigm (Nass et al., 1994), this is achieved through the incorporation of social cues, which are design features derived from interpersonal communication (e.g., small talk and emojis) (Feine et al., 2019, Gnewuch et al., 2017). Social cues trigger subconscious social responses by the users that positively or negatively influence the degree of interaction with a chatbot, depending on the elicited user perceptions and expectations towards it (Nass et al., 1994, Nass and Moon, 2000). There are numerous design elements that can be varied to stimulate perceptions and expectations of anthropomorphism towards the chatbot. While for the different interaction phases (i.e., before, during, after, in case of "error"/no intent detection), general design guidelines exist (e.g., Amershi et al. (2019), Weber and Ludwig (2020)), there are no precise guidelines for the human-like design of chatbots, which makes it difficult to select appropriate design features (Feine et al., 2019). Nevertheless, for the systematic selection of social cues, Feine et al. (2019) developed a taxonomy that categorises conversational agents' human-like cues into verbal, visual, auditory and invisible social cues with ten subcategories in total.

However, to the best of our knowledge, regarding the literature and existing chatbots, no chatbot was developed for the use in South Africa or Ghana so far and only few studies have addressed the topic of designing chatbots for blood donors. Among them are the chatbot designed by Roman et al. (2020) for a Brazilian blood donation centre, the Facebook Messenger bot of the Canadian Blood Services (Canadian Blood Services, 2017) as well as "Clara", the chatbot of the American National Red Cross appearing on its website in the form of a female doctor (The American National Red Cross, 2022). For donor education and mobilisation purposes, they are all able but limited to answer specific questions about blood donation and its process as well as booking appointments. On the one hand, they enable free text input and provide guidance to the user in the form of buttons and website links but, on the other hand, even though they are partly represented by avatars and names, they overall only have very few social cues (e.g., no use of small talk, response delays and emojis), which may lead to a lower social presence of the chatbot negatively influencing the quality of interaction (Nass and Moon, 2000).

In addition, thorough evaluations of the design of blood donation chatbots are missing. Even though Roman et al. (2020) made use of a well elaborated standard questionnaire for the evaluation of their designed chatbot, this questionnaire is applicable to any tool and not devised for the specific evaluation of chatbots. Therefore, in this paper, we build upon their work and extend the evaluation by another standard questionnaire specifically aimed at chatbots as well as the collection of detailed feedback by asking specific questions about our design. Compared to Roman et al. (2020), our sample size was more than seven times higher, having more than twice as much interaction time with the chatbot with almost three times as many implemented intents.

Putting all this together, we argue that chatbots appropriately designed for and evaluated with potential blood donors might serve as a persuasive and natural way to support all types of donors and promote sustained blood donation behaviour (i.e., transition to and retention of regular donors).

3. Design science research project

Responding to the call by Burditt et al. (2009) for more theory-based recruitment and retention interventions with regard to blood donation due to the lack of effective practical solutions, we apply the DSR methodology (Hevner et al., 2004) particularly suited to address this challenge and answer our research question, since we target tailored blood donor management to better match demand and supply of blood donations

as a real-world problem by iteratively designing and evaluating a chatbot in the specific context of potential blood donors in South Africa and Ghana. We partly consulted the latter from our three blood service collaborators from those regions to ensure the coverage of all four types of donors as regular donors are rather rare and to increase the relevance of our research. This means they are either from the National Blood Service Ghana (NBSG), the Western Cape Blood Service (WCBS) in South Africa or the South African National Blood Service (SANBS) that provides its service to all of South Africa's provinces except of the Western Cape province. In tandem with the help of the platform Clickworker, from 25 February 2023 to 6 March 2023, we randomly recruited a total of 371 potential blood donors, 226 coming from South Africa and 145 from Ghana. As depicted in Figure 1, our DSR project with three subsequent design cycles is based on the framework proposed by Kuechler and Vaishnavi (2008) due to its clear and straightforward structure compared to other DSR frameworks. As this study focuses on our last design cycle, in the following, we only summarise design cycle three.

In our third design cycle, we balanced the rigor and relevance of our research by replicating our results of the second design cycle with our real end users. We started our last cycle with further thinking about additional design features to address the issues raised in the focus group discussions with our blood service experts. Consequently, we extended our design by new chatbot features and instantiated them as an updated version into our prototype developed with Google Dialogflow providing natural language processing capabilities for user intent detection and a custom-built web interface providing convenient access. We then evaluated our artefact with 371 (potential) blood donors from South Africa and Ghana via an online survey, quantitatively by their rating of the chatbot with the help of standardised questionnaires and qualitatively by their responses to our questions concerning our design.

4. Designing chatbots for blood donor support

For supporting potential blood donors and promoting sustained blood donation behaviour, knowledge about their motivators and barriers grounded on existing (i.e., theory of planned behaviour and transtheoretical model) as well as derived (i.e., user archetypes: nd, fd, ld, rd) behavioural change models is crucial. With this, we ensure a user-centred design of the chatbot following the three design principles (DPs) as shown in Figure 2 with their instantiation depicted in Figure 3. Considering the practical perspective, based on expert interviews and focus group discussions conducted with our African blood service partners and findings in the literature (Batis and Albarrak, 2021), we identified eleven key challenges (C1 - C11) in the chatbot design process (see Figure 2). Most of them (C1 - C7) refer to the second design requirement (DR2) as for the experts donor education is most important and promises the highest added value. Additionally, they assumed the chatbot to be a reactive system rather than a proactive one, whose abilities are mainly covered by DR1 and DR3.

In total, the first five challenges (C1 – C5) are addressed by **DR2.1**, describing the chatbot's ability to answer a wide variation of user questions concerning blood donation and its process. First, according to WCBS and Asamoah-Akuoko et al. (2021), Head of Research and Development and her colleagues from NBSG, as first-time donors face high uncertainties after donation, their return rates are very low in South Africa (ca. 50 %) and especially in Ghana (15,2 % within six months) (**C1**). Second, the Head of Marketing and PR of WCBS mentioned that most of the questions reaching customer service every day (ca. 90 %) are standard questions such as "Do I qualify as a blood donor?" and "Where can I donate blood?" (**C2**). However, the Senior Manager Business Intelligence of SANBS raised the concern that even with FAQ coverage, especially medical questions must still be treated with caution in order not to confuse donors. This undermines the suggestion the Head of Donor Management at WCBS made regarding the chatbot providing specific answers to certain deferral issues (e.g., medication, surgery, travelling abroad) by asking follow-up questions such as "What kind of surgery did you have?". Nevertheless, the Head of Marketing and PR of WCBS also highlighted to put emphasis on medical issues (e.g., avoid people coming in vain through clarification of Hb level) and proposed to point to the right contact person for sensitive user requests instead (**C3**). According to both targeted blood services, the unpredictable course of the COVID-19 pandemic is another important issue to consider, causing a lot of uncertainty among blood donors, for example, with regard to constantly changing regulations (**C4**). Moreover, the employees of NBSG emphasised that people of African culture have very strong perceptions and beliefs, including myths and misconceptions preventing them from donating blood (**C5**). Therefore, we consulted their list of collected assumptions wrongly made about donating blood such as that HIV or other infections can be contracted or that vegetarians are iron deficient and cannot give blood. Consequently, we extended the

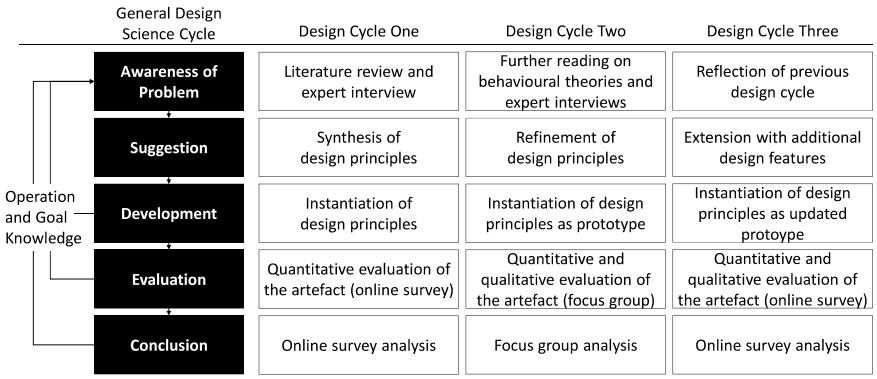

Figure 1. Design cycles with respective research activities.

chatbot's training phrases and responses of user intents already (indirectly) addressing these misconceptions and additionally integrated two more intents on Google Dialogflow for whole coverage (i.e., one concerning deferrals and the other one with regard to nutrition). In order to address the high variability of user input, we also created a default fallback response in case intent detection was not possible, together with the default welcome intent leading to 86 intents in total.

In line with the interviews conducted by Batis and Albarrak (2021), blood donation generally is not very attractive among people and only few are aware of its impact (**C6**). This challenge is addressed by **DR2.2** targeting the chatbot's ability to raise awareness about blood donation and its benefits. During the focus group discussion in Ghana, the Head of Planning of NBSG proposed to deepen the knowledge on the user's topic of interest to his/her fullest satisfaction. Besides the further integration of links to videos and documents, by applying the card sorting method, we therefore additionally included dialogue continuity suggestions in the form of buttons for triggering four related follow-up questions to eleven of our 86 modelled intents on Google Dialogflow.

The blood service employees interviewed by Batis and Albarrak (2021) also pointed out that it is generally important to provide guidance on pre-donation requirements to prevent people coming in vain, which is related to the aim raised by the Senior Manager Business Intelligence to avoid frustrating deferrals due to potential losses of blood donors (**C7**). This is addressed by **DR2.3** referring to the chatbot's ability to inform about blood donation requirements before the appointment takes place. A self-test for checking the own eligibility to give blood by clicking through the individual blood donation criteria and deciding whether they are met or not was already included.

Even though the online survey of our first design cycle revealed that mostly lapsed donors tend to use a chatbot for appointment making in comparison to the other donor groups (Müller and Reuter-Oppermann, 2022b), the Head of Marketing and PR of WCBS proposed to integrate appointment booking for plasma donations that are predominantly done by regular donors who might also benefit from making appointments quickly and easily (**C8**). This is addressed by **DR1.1** describing the chatbot's ability of providing assistance in registering and setting up appointments. Our former version of the chatbot already assisted in this task through entities on Google Dialogflow representing annotations for parameters like place and time queried during the chatbot conversation as well as contexts allowing sequential dialogues in case users respond with a simple "yes" or "no".

For communication and information exchange, the Head of Marketing and PR of WCBS suggested to enable pleasant human-like interactions similar to what donors expect from customer service (**C9**), while the CIO of SANBS also highlighted to put emphasis on the attraction of younger donors who are underrepresented in South Africa (Vermeulen et al., 2019) (**C10**). These challenges are addressed by **DR3.5** targeting the chatbot's ability to display social cues embodying a mix of different tailored characteristics. To support these characteristics, we purposefully selected a set of social cues for the human-like design of our chatbot based on the taxonomy of Feine et al. (2019) mentioned in Section 2. As a chatbot per se is a text-based conversational agent, we eliminated the auditory category as well as all subcategories belonging to embodied conversational agents. With regard to verbal cues, referring to the content, Amershi et al. (2019) as well as Weber and Ludwig (2020) for example recommend that at the start of an interaction, a chatbot should be able to present its capabilities to let users know its purpose when interacting with a

chatbot for the first time. In addition, with regard to trust and privacy issues, the chatbot should be able to explain the handling of user information (Weber and Ludwig, 2020). Referring to the style, especially appealing to younger people, we decided to apply informal language for instance. Concerning visual cues, Roman et al. (2020) for example find empirical evidence in an experiment that the blood donation chatbot representation in the form of an avatar, in this case a mascot of the blood donation centre in Brazil, increases user satisfaction as well as trust and at the same time decreases the inhibition threshold of the users. Considering invisible social cues, Gnewuch et al. (2017) for example demonstrated that, despite longer latency times for the user by using response delays that simulate the chatbot's thinking and typing, dynamic response times also positively contribute to the chatbot's anthropomorphic perception.

With regard to the individual donor groups, according to the Head of Marketing and PR of WCBS, finding the right strategy to address them appropriately is not always easy, especially when looking at the cumbersome process of regaining lapsed donors by sending SMS back and forth (**C11**). This is addressed by **DR3.6** referring to the chatbot's ability to facilitate direct follow-up conversations and act like a companion according to the user's stage of change to foster donor engagement. To make this proactive reaction of the chatbot fully adapted to the user possible, we plan to integrate the chatbot into a blood donation app that provides additional functionalities for the individual types of donors that positively influence blood donor behaviour change directly after the user's initial stage of change is determined through a short self-test (Müller and Reuter-Oppermann, 2022a). Like a companion (or "blood buddy", which is why we named it "Bloody"), the chatbot can provide stage-matched guidance and feedback to the user to support him/her to positively develop in his/her blood donor career (Amoyal et al., 2013).

5. Evaluation

5.1. Evaluation methodology

Due to the need of more real-world evaluations of DSR artefacts (Peffers et al., 2012, Venable et al., 2016), we evaluated our chatbot prototype through the involvement of our target group of potential blood donors from South Africa and Ghana. In order to induce sustained blood donation behaviour and overall increase the supply of blood donations to better match demand, it is a prerequisite for the chatbot to be continuously used by as many potential donors as possible. According to the unified IS continuance model of Bhattacherjee and Lin (2015) three constructs directly and positively influence IS continuance: subjective norm, perceived usefulness and satisfaction. The latter two are encompassed by the broader term user experience (UX), which includes many other quality attributes besides usability (e.g., functionality) (Laugwitz et al., 2008). Hence, we evaluated our chatbot by conducting a UX assessment via an online survey with potential South African and Ghanaian blood donors to get feedback on how they perceive the chatbot and how it can be improved.

In this work, similar to Roman et al. (2020), we used the User Experience Questionnaire (UEQ) based on the UX framework of Zarour and Alharbi (2017). It equally considers pragmatic as well as hedonic product quality aspects, using the UX framework of Hassenzahl (2001) (Laugwitz et al., 2008). We additionally applied the Chatbot Usability Questionnaire (CUQ), which is based on the chatbot UX categories by Martín et al. (2017) and specifically tailored to the UX assessment of chatbots (Holmes et al., 2019). The UEQ includes the following six scales: attractiveness (i.e., extent of likeability of the product), perspicuity (i.e., extent of familiarity with the product), efficiency (i.e., extent of usefulness of the product), dependability (i.e., extent of user control), stimulation (i.e., extent of motivation to use the product), and novelty (i.e., extent of innovativeness of the product). They are rated through in total 26 item pairs of antonymous adjectives describing the product (e.g., "annoying" vs. "enjoyable"). Seven stage scales are provided for each pair allowing the user to respectively select its level of agreement with one of the presented items. The CUQ is composed of 16 statements, of which half reflect positive aspects and the other half negative aspects of chatbot usability. For each statement, the respondents decide on their level of agreement via a five-point Likert scale from 1 = "strongly disagree" to 5 = "strongly agree". Additionally, to get qualitative feedback on our proposed design, we purposefully selected questions we discussed at the focus group workshops with our African blood service experts.

In total, 371 potential blood donors from South Africa (226) and Ghana (145) participated in our online survey (i.e., nd = 191 (51,48 %), fd = 74 (19,95 %), ld = 72 (19,41 %), rd = 34 (9,16 %)). Of the 371 participants, 295 (79,51 %) have already used and 76 (20,49 %) never used a chatbot before. Our sample represents a fairly balanced gender (i.e., m = 172 (46,36 %) and f = 199 (53,64 %)) as well as age mix (from 18 to 57 years), ensuring comprehensive feedback

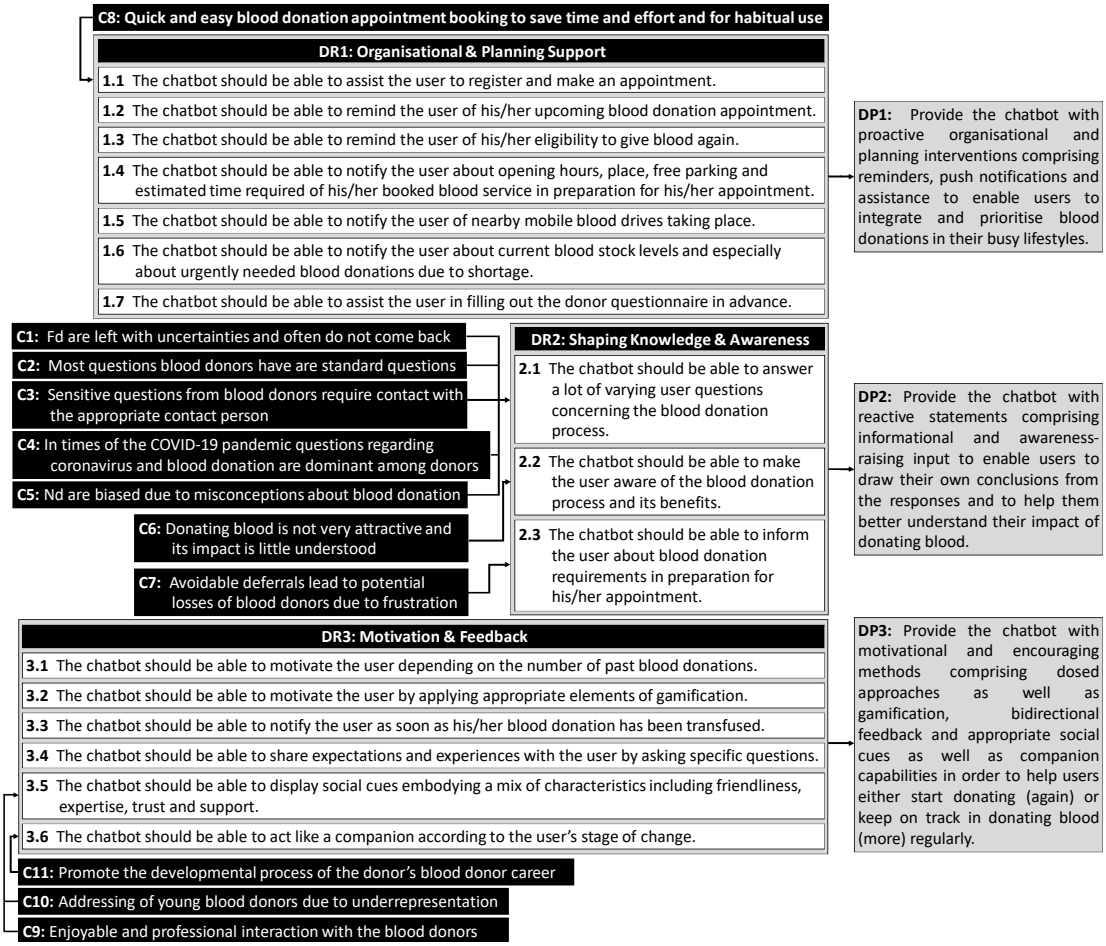

Figure 2. Cs, DRs and DPs for blood donation chatbots (based on Müller and Reuter-Oppermann (2022b)).

from diverse perspectives. On average, each chatbot experiment lasted half an hour and was structured as follows. First, the instructions comprising a brief definition of chatbots, our objectives and the procedure of the experiment were displayed. Then, participants were asked for their agreement with our privacy policy and the use of Google's Dialogflow and Forms services for interaction with our chatbot and allowing us to conduct our online survey. Next, they were asked three control questions to make sure that they understood our chatbot experiment correctly. If they answered them correctly, each participant freely interacted with our chatbot empathising with the role of being a potential blood donor for 15 minutes. Afterwards, the chatbot provided a survey link for the participants in order to open to answer questions related to their demographics, the UEQ, the CUQ and their feedback on the selected questions from our previous focus group discussions with the blood service experts.

5.2. Results and discussion

Overall, the potential blood donors from South Africa and Ghana generally rated the UX of our chatbot positively, which we revealed by conducting data analyses through the analytical Excel tools provided for each questionnaire (i.e., UEQ and CUQ) to obtain scores for comparison. Regarding the seven-stage semantic differential scales of the UEQ, the gradations are equivalent to -3, -2, -1, 0, 1, 2, 3 with scores ranging from -3.000 to 3.000 according to the user ratings. If the average rating is lower than –0.800, the UX is perceived negatively, whereas it is perceived positively when the mean is higher than 0.800 and neutral if in between. Table 1 shows the results per item pair and the overall scales. All of the overall scales as well as item pairs were rated with a mean higher than 0.800 besides those referring to novelty (except "dull/creative" (mean = 1.315)) and the pair with the items "unpredictable" and "predictable" (mean = 0.337). For them, the

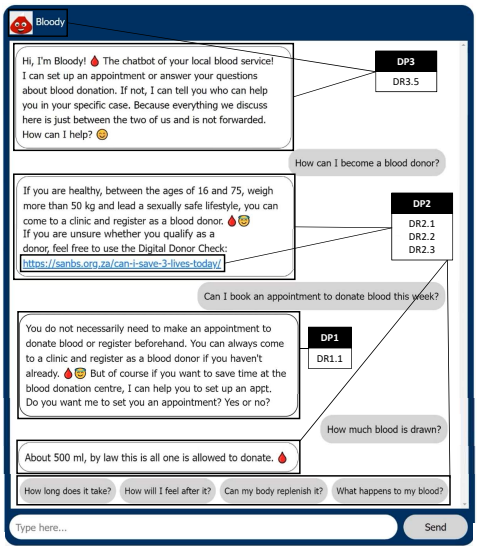

Figure 3. Instantiated DPs.

user responses are distributed rather fairly (i.e., neutral rating). For the item pair "unpredictable/predictable", one possible reason might be that a chatbot represents a "blank canvas where the content and features of the underlying service are mostly hidden from the user" (Følstad and Brandtzæg, 2017, p. 41). The unfamiliarity with chatbot features is especially true for users who have never used a chatbot before and with its content for users who are not or never have been into blood donation and its process, which is the case for most of our participants being non-donors. For the overall novelty scale (mean = 0.623) and its corresponding item pairs with means below 0.800 (i.e., "conventional/inventive" (mean = 0.385), "usual/leading edge" (mean = 0.156) and "conservative/innovative" (mean = 0.636)), this might be due to the fact that ca. 80 % of our participants were already familiar with chatbot usage and it was nothing new to them. Consequently, it is not surprising that among the six scales, the perspicuity scale achieved the highest mean (1.956), which was mainly influenced by the chatbot being perceived as easy (2.189 = second highest mean among all item pairs). Closely followed by the attractiveness scale with the second highest mean (1.881), primarily influenced by the chatbot being perceived as friendly (2.251 = highest mean among all item pairs) and good (2.086 = third highest mean among all item pairs). This lets us assume that our design generally is user-friendly and that our participants from South Africa and Ghana generally liked the chatbot, indicating that even though they are very used to this technology, according to their rating, it might add value after its release. In the presentation of our results, we have not differentiated between the rating of the

Table 1. Results of the UEQ (N = 371).

Semantic differential rating	Mean ± SD	95% CI	Scale/Quality
annoying/enjoyable	1.704 ± 1.383	0.141	Attractiveness
not understandable/understandable	1.784 ± 1.304	0.133	Perspicuity
dull/creative	1.315 ± 1.512	0.154	Novelty
difficult to learn/easy to learn	1.900 ± 1.473	0.150	Perspicuity
inferior/valuable	1.884 ± 1.408	0.143	Stimulation
boring/exciting	1.499 ± 1.344	0.137	Stimulation
not interesting/interesting	1.714 ± 1.294	0.132	Stimulation
unpredictable/predictable	0.337 ± 1.788	0.182	Dependability
slow/fast	1.666 ± 1.405	0.143	Efficiency
conventional/inventive	0.385 ± 1.825	0.186	Novelty
obstructive/supportive	1.865 ± 1.145	0.117	Dependability
bad/good	2.086 ± 1.187	0.121	Attractiveness
complicated/easy	2.189 ± 1.111	0.113	Perspicuity
unlikable/pleasing	1.895 ± 1.177	0.120	Attractiveness
usual/leading edge	0.156 ± 1.820	0.185	Novelty
unpleasant/pleasant	1.838 ± 1.174	0.120	Attractiveness
not secure/secure	1.658 ± 1.302	0.133	Dependability
demotivating/motivating	1.795 ± 1.184	0.120	Stimulation
does not meet expectations/meets...	1.523 ± 1.437	0.146	Dependability
inefficient/efficient	1.717 ± 1.275	0.130	Efficiency
confusing/clear	1.951 ± 1.242	0.126	Perspicuity
impractical/practical	1.803 ± 1.240	0.126	Efficiency
cluttered/organised	1.987 ± 1.193	0.121	Efficiency
unattractive/attractive	1.512 ± 1.344	0.137	Attractiveness
unfriendly/friendly	2.251 ± 0.972	0.099	Attractiveness
conservative/innovative	0.636 ± 1.913	0.195	Novelty
Overall Attractiveness Scale	1.881 ± 0.979	0.100	Attractiveness
Overall Perspicuity Scale	1.956 ± 0.989	0.101	Pragmatic
Overall Efficiency Scale	1.793 ± 1.018	0.104	Pragmatic
Overall Dependability Scale	1.346 ± 0.893	0.091	Pragmatic
Overall Stimulation Scale	1.723 ± 1.075	0.109	Hedonic
Overall Novelty Scale	0.623 ± 1.174	0.120	Hedonic

SD = Standard deviation, CI = Confidence interval

Table 2. Results of the CUQ (mean ± SD).

No.	Statement	Score
1	The chatbot's personality was realistic and engaging.	3.8 ± 1.1
2	The chatbot seemed too robotic.	2.7 ± 1.2
3	The chatbot was welcoming during initial setup.	4.2 ± 1.0
4	The chatbot seemed very unfriendly.	1.6 ± 1.0
5	The chatbot explained its scope and purpose well.	4.2 ± 0.9
6	The chatbot gave no indication of its purpose.	1.7 ± 1.2
7	The chatbot was easy to navigate.	4.4 ± 0.9
8	It would be easy to get confused when using the chatbot.	1.9 ± 1.1
9	The chatbot understood me well.	3.7 ± 1.1
10	The chatbot failed to recognise a lot of my inputs.	2.4 ± 1.2
11	The chatbot response were useful, appropriate and informative.	4.3 ± 0.8
12	The chatbot responses were irrelevant.	1.7 ± 1.0
13	The chatbt coped well with any errors or mistakes.	3.7 ± 1.1
14	The chatbot seemed unable to handle any errors.	2.1 ± 1.1
15	The chatbot was very easy to use.	4.6 ± 0.7
16	The chatbot was very complex.	1.8 ± 1.1

participants from South Africa and Ghana as well as between the individual donor types (i.e., nd, fd, ld, rd), as the respective results only differ marginally from those displayed. The same applies to the presentation of the CUQ results.

Regarding the 16 statements of the CUQ, the odd-numbered relate to the positive usability aspects of the chatbot, which are supported by the users when the level of agreement is rather high (i.e., towards 5 = "strongly agree"). The even-numbered statements relate to the negative usability aspects of the chatbot, which are not supported by the users when the level of agreement is rather low (i.e., towards 1 = "strongly disagree"). Comparing the level of agreements of our participants from South Africa and Ghana shown in Table 2, it can easily be seen that they agreed to the positive usability aspects and rather disagreed to the negative ones (with a slight tendency towards neutrality

regarding their perception of the chatbot being robotic (no. 2)). Overall, this means that our participants rated the usability of the chatbot rather positively than negatively.

The fact that the chatbot has potential to add value for blood donation was also reflected in the participants' qualitative statements. Concerning our general questions (i.e., "What did you like or dislike about the chatbot?", "Would you use the chatbot again, why or why not?" and "What improvements or extensions would you make?"), the overall feedback on the chatbot our survey participants wrote was rather positive. Two non-donors stated that *"the chatbot, Bloody if I should use its name, was quite interesting. I like the fact that it encourages and gives facts about blood donation. What I dislike was the fact that it didn't understand some of my questions"* and that *"the chatbot generally was able to communicate its purpose lucidly. Moreover, any question I asked was answered to satisfaction. However, what I dislike was that the bot didn't give me the opportunity to introduce myself. Overall, I liked the performance of the chatbot"*. Additionally, one of the regular donors mentioned that *"the chatbot was fast and information was realistic. Some answers are not straight forward but was relevant and on point. Love how it gives suggestions after it responds to certain questions, it was relevant"*. Almost all of our participants would use the chatbot again to either learn more about donating blood, because it provides *"new, relevant and useful information"* or due to the fact that the chatbot *"makes it easy to even set up appointments or find about where to go"*. Others prefer how *"effective and efficient"* it is compared to *"using Google and reading a 43-page article just for a piece of information"* or *"getting in touch with a human consultant[, which] can be time-consuming"*. As improvements, one participant suggested to implement *"a menu with already made questions [users can choose from]"*, which ideally also covers disease-related questions many participants wish to get answered. Others asked for more personalised messaging with the chatbot being *"more empathetic and relat[ing] to the user's feelings"* or *"able to acknowledge the person it is chatting with[, which] can go a long way to create an atmosphere of friendliness"*.

In order to confirm the three DPs, the participants had to decide for each DP if it was covered by the chatbot or not. Overall, at least about 75 % of our participants confirmed the chatbot's coverage for each DP. DP1 obtained the lowest level of agreement while DP2 achieved the highest one (approx. 80 % confirmed its coverage). Even regular donors perceived the chatbot as informative and able to raise awareness since one of their responses to the question "Did the chatbot provide you with informational and awareness-raising input?" was that *"the information I received has been insightful. Many doubts have been cleared about blood donation. These information you usually do not get clarity on from many centres, and that makes it more convincing for me and others who might use the chatbot to consider donating blood frequently"*. According to the participants statements, the most insightful information provided by the chatbot was that with every unit of blood donated up to three lives can be saved. This was also one of the main reasons why our participants confirmed that they had been motivated by the chatbot to donate blood (DP3). Another reason for confirming the question "Did the chatbot motivate you to donate blood?" was *"because I found out that the blood I donate will regenerate within 24 hours"*, as one of the first-time donors stated. Additionally, the statements *"because now I know the importance of me donating blood because I have universal blood"* and *"It's been a while since I donated, and the bot gave me inspiration to go back and help out"* of our lapsed donors indicate the chatbot's potential to induce blood donor behaviour change. This is also evident in the responses to our follow-up question "After interacting with the chatbot, do you consider donating blood in the near future (again)?" ("yes" = 297 (80,05 %), "no" = 44 (11,86 %), which was again mainly due to health reasons and "I do not know" = 30 (8,09 %)). Two non-donors agreed in saying *"because it feels easy to donate blood and not that painful"* and *"I have always wanted, and the information provided by the chatbot gave me more desire to do so"*. In addition, two first-time donors optimistically mentioned *"I will be donating every six weeks"* and *"I am at a rather healthy point in life, this would be the best time to give back and be noble. Chatbot did help, genuinely will look into setting a date"*. Even one of the lapsed donors stated that *"it has made me aware of all the importance of donating blood and has also made me a lot more comfortable to go and do the blood donation because I know all the facts and do's and don'ts which has really made it more appealing for me"*. The chatbot also has potential to foster sustained blood donation behaviour among regular donors as one of them highlighted that the *"chatbot shared light and more knowledge, places to go and what I need to know"*.

The only time opinions diverged between our South African and Ghanaian participants was on our final questions about the chatbot's human-like design, a phenomenon that we already experienced during previous focus group discussions with the South African and Ghanaian blood service experts. While the South

African participants had ambivalent opinions on our question "Did you perceive the chatbot as human-like, why or why not?", the Ghanaians predominantly perceived the design as anthropomorphic. On the one hand, participants stated that the chatbot *"feels like a friend in front"* and *"questions I asked it and answers it provided suggested some level of imagination and creativity"* but on the other hand, *"it's very robotic. The only thing that might make it have a little bit human-like is the use of emojis"* and that *"it lacks human emotional intelligence"*. The participants were also unsure about the chatbot's name with regard to the question "What did you think of the name and look of the chatbot? What would you change and how?" because for some of them *"Bloody has a negative connotation"* sounding a *"little scary"*, although most people were in favour of keeping the name as it is *"very simple to remember and to mention"* and *"it justifies what the chatbot is doing"*.

6. Conclusion and outlook

In this work, we have investigated the design of a chatbot to support potential blood donors and promote sustained blood donation behaviour. We focused on a chatbot that can be integrated into a website, e.g., supporting behaviour change in a mainly passive way and targeted the user perspective in the evaluation. In total, 371 (potential) blood donors representing all four donor types from South Africa (226) and Ghana (145) participated in our online survey. Overall, the evaluation of the chatbot was affirmative regarding our three DPs and comments of the participants were positive, expressing the chatbot's potential to change their blood donor behaviour. We contribute with a demonstration of how researchers, designers and developers can make use of our prescriptive design knowledge and put it into practice. However, even though we were able to built on previous research we performed for Germany (Müller and Reuter-Oppermann, 2022b), this only indicates the generalisability of the results to other countries and future cultural research is needed.

As a next step towards the deployment in practice, the chatbot will be integrated into the BISKIT information system, taking the feedback of the survey participants into account. The aim is that the chatbot will be available on the blood services' websites as well as within a dedicated BISKIT blood donor app to release the full potential of the designed chatbot.

In addition, we will investigate the usefulness and the resulting design of chatbots that support the blood services themselves in their daily tasks. This also includes the use of a chatbot as an interface to the BISKIT information and decision support system that, e.g., can display and monitor stock levels, determine optimised transport routes or simulate crisis scenarios.

Acknowledgement: This research was conducted within the project "Blood Information System for Crisis Intervention and Management" (BISKIT) funded by the German Federal Ministry of Research and Education.

References

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., & et al. (2019). Guidelines for Human-Ai Interaction. *CHI 2019 Proceedings*.

Amoyal, N., Robbins, M., Paiva, A., Burditt, C., Kessler, D., & Shaz, B. (2013). Measuring the processes of change for increasing blood donation in black adults. *Transfusion*, *53*(6), 1280–1290.

Asamoah-Akuoko, L., Ullum, H., Appiah, B., Hassall, O. W., Ndanu, T., Adongo, P., & Bates, I. (2021). Determinants of intention to return to donate blood among first-time blood donors in Ghana. *Vox Sanguinis*, *116*(3), 324–335.

Batis, A., & Albarrak, A. (2021). Preferences and features of a blood donation smartphone app: A multicenter mixed-methods study in Riyadh, Saudi Arabia. *Comput. Methods Programs. Biomed.*, Update 1, 100005.

Bhattacherjee, A., & Lin, C. (2015). A unified model of IT continuance: Three complementary perspectives and crossover effects. *Eur. J. Inf. Syst.*, *24*(4), 364–373.

Burditt, C., Robbins, M., Paiva, A., Velicer, W., Koblin, B., & Kessler, D. (2009). Motivation for blood donation among African Americans: Developing measures for stage of change, decisional balance, and self-efficacy constructs. *J. Behav. Med.*, *32*(5), 429–442.

Canadian Blood Services. (2017). Canadian Blood Services launches new tool to engage young donors. https://www.blood.ca/en/about-us/media/newsroom/canadian-blood-services-launches-new-tool-engage-young-donors

Dale, R. (2016). The return of the chatbots. *Nat. Lang. Eng.*, *22*(5), 811–817.

Diederich, S., Lichtenberg, S., Brendel, A. B., & Trang, S. (2019). Promoting sustainable mobility beliefs with persuasive and anthropomorphic design: Insights from an experiment with a conversational agent. *ICIS 2019 Proceedings*.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for

Conversational Agents. *Int. J. Hum. Comput. Stud.*, *132*, 138–161.

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *Interactions*, *24*, 38–42.

Gnewuch, U., Morana, S., & Maedche, A. (2017). Towards Designing Cooperative and Social Conversational Agents for Customer Service. *ICIS 2017 Proceedings*.

Hassenzahl, M. (2001). The Effect of Perceived Hedonic Quality on Product Appealingness. *Int. J. Hum-Comput. Int.*, *13*(4), 481–499.

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Q*, *28*(1), 75–105.

Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., & Mctear, M. (2019). Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? *ECCE 2019 Proceedings*.

Horstkemper, D., Reuter-Oppermann, M., Middelhoff, M., Widera, A., & Hellingrath, B. (2021). Improving blood supply chain crisis management by simulation-based optimization. *Simulation in Produktion und Logistik 2021: Erlangen, 15.-17. September 2021*.

Kuechler, W., & Vaishnavi, V. (2008). On theory development in design science research: Anatomy of a research project. *Eur. J. Inf. Syst*, *17*(5), 489–504.

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger (Ed.), *HCI and Usability for Education and Work. USAB 2008. LNCS* (pp. 63–76). Springer.

Martín, J., Muñoz-Romero, C., & Ábalos, N. (2017). Chatbottest – Improve your chatbots design. https://chatbottest.com/

Morana, S., Friemel, C., Gnewuch, U., Maedche, A., & Pfeiffer, J. (2017). Interaktion mit smarten Systemen – Aktueller Stand und zukünftige Entwicklungen im Bereich der Nutzerassistenz. *WuM*, *9*(5), 42–51.

Müller, H. M., & Reuter-Oppermann, M. (2022a). A Design Science Approach to Blood Donation Apps. In A. Drechsler, A. Gerber, & A. Hevner (Eds.), *The Transdisciplinary Reach of Design Science Research. DESRIST 2022. LNCS* (pp. 221–232). Springer.

Müller, H. M., & Reuter-Oppermann, M. (2022b). Chatblood - towards designing chatbots for blood donors. *ECIS 2022 Research-in-Progress Papers.*

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *J. Soc. Issues*, *56*(1), 81–103.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are Social Actors. *SIGCHI 1994 Proceedings*, 72–78.

Peffers, K., Rothenberger, M., Tuunanen, T., & Vaezi, R. (2012). Design Science Research Evaluation. In K. Peffers, M. Rothenberger, & B. Kuechler (Eds.), *Design Science Research in Information Systems. Advances in Theory and Practice. DESRIST 2012. LNCS* (pp. 398–410). Springer.

Roman, M., Bellei, E., Biduski, D., Pasqualotti, A., Araujo, C., & Marchi, A. (2020). "Hey assistant, how can i become a donor?" The case of a conversational agent designed to engage people in blood donation. *J. Biomed. Inform.*, *107*, 103461.

SANBS. (2023). Blood stock levels. https://sanbs.org.za/

The American National Red Cross. (2022). Meet Clara, the Blood Donation Chatbot. https://www.redcrossblood.org/donate-blood/dlp/meet-clara--the-blood-donation-chatbot-.html

Venable, J. R., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A Framework for Evaluation in Design Science Research. *Eur. J. Inf. Syst.*, *25*, 77–89.

Verhagen, T., Van Nes, J., Feldberg, F., & Van Dolen, W. (2014). Virtual Customer Service Agents: Using Social Presence and Personalization to Shape Online Service Encounters. *J. Comput-Mediat. Comm.*, *19*(3), 529–545.

Vermeulen, M., Lelie, N., Coleman, C., Sykes, W., Jacobs, G., Swanevelder, R., Busch, M., Zyl, Gert, G., Eduard, W., Alex, R., & Ravi. (2019). Assessment of HIV transfusion transmission risk in South Africa. *Transfusion*, *59*(1), 267–276.

Weber, P., & Ludwig, T. (2020). (Non-)Interacting with Conversational Agents: Perceptions and Motivations of Using Chatbots and Voice Assistants. *MuC 2020 Proceedings*, 321–331.

Weizenbaum, J. (1966). ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine. *Commun. ACM*, *9*(1), 36–45.

Zarour, M., & Alharbi, M. (2017). User experience framework that combines aspects, dimensions, and measurement methods. *Cogent. Eng.*, *4*, 1–25.