Take Back Control: User Privacy and Transparency Concerns in Personalized Conversational Agents

Iris Hendrickx[a], Jelte van Waterschoot[b], Arif Khan[a], Louis ten Bosch[a], Catia Cucchiarini[a] and Helmer Strik[a]

[a]Centre for Language Studies, Centre for Language and Speech Technology, Radboud University, Erasmusplein 1, 6500 HD, Nijmegen, The Netherlands

[b]Human Media Interaction, University of Twente, Drienerlolaan 5, 7522 NB, Enschede, The Netherlands

Abstract

We reflect on user privacy concerns, transparency and informed consent for long-term interactions with personalized conversational agents. We argue that the common practice of asking users to sign an informed consent form is insufficient to accommodate the privacy concerns of the user. We propose that long-term engaging personalized conversational agents must include an explicit mechanism in their conversations to allow users to have control over their personal information and to have transparency, i.e. about what is stored and who is allowed to view the stored personal information.

Keywords

dialogue systems, conversational agents, privacy concerns, user consent, personalization

1. Introduction

An important aspect in developing conversational agents (CAs) that engage in personalized, long-term interactions is how to ensure privacy and transparency. Privacy issues have received considerable attention in recent years and the introduction of GDPR in European countries is a clear way of addressing at least part of these issues. GDPR requires that users be thoroughly informed about the use that is made of their personal data and they give explicit consent for including their personal information in all sorts of systems and databases. Usually this is done upfront, before users start participating in a study or using a given application. However, with conversational agents we might face a different situation in which users are, sometimes unconsciously, continuously providing personal information to the CA that then incorporates these data without asking for further consent. This might lead to particularly worrying scenarios in which confidential information is shared and accessed by third parties without specific approval on the part of the user. In this paper we argue that this aspect should be explicitly addressed when designing CAs, making it possible for users to check and approve that the information and personal data they share with their CA can be stored and incorporated for subsequent use. In other words, we claim that users should take back control and give explicit consent at different

Joint Proceedings of the ACM IUI 2021 Workshops, April 13-17, 2021, College Station, Texas, USA

i.hendrickx@let.ru.nl (I. Hendrickx); j.b.vanwaterschoot@utwente.nl (J.v. Waterschoot); a.khan@let.ru.nl (A. Khan); l.tenbosch@let.ru.nl (L.t. Bosch); c.cucchiarini@let.ru.nl (C. Cucchiarini); w.strik@let.ru.nl (H. Strik)

0000-0002-9351-6449 (I. Hendrickx); 0000-0002-3361-2105 (J.v. Waterschoot)

stages of their conversational exchanges with CAs. We focus here on two aspects of personalized long-term engaging conversations with CAs:

1. Conflict between privacy and personalization
2. User control on and user access to personal information stored by the CA

2. Context

We focus on spoken personalized long-term engaging conversations with older adults to collect and reflect on their overall health and well-being within the BLISS project [1]. We are not so much interested in clinical health but in a broader concept of well-being and happiness that also comprises concepts such as quality of life, social participation and the overall capability to self-manage [2]. The use of conversational agents in healthcare facilities is growing [3]. Several studies have been conducted to check for the widespread acceptance of such technology in healthcare facilities. Razavi et al. have designed a dialogue manager to conduct conversations with older adults in order to receive feedback on their non-verbal communication such that to improve their communication skills [4]. Abdollahi et al. have showed the use of social robots in elderly homes to help patients in long-term engagements with the robots and to improve their quality of life in patients having dementia and depression [5]. They incorporated humor and other features that make robots likable for longer interactions so that subjects interest is intact for longer period of time.

In order to have improved technology that performs well for a broader and variety of audience, the technology must be personalized to a particular person.However, a personalized conversational agent might require disclosure of information that users feel reluctant to share. For example when a CA asks about topics like previous life experiences, accidents or any particular health issues. These could possibly trigger unwanted negative emotions of the user and unwanted ethical issues [6, 7].

3. Personalization

Dialogue personalization leads to higher satisfaction of users[8] which in turn can lead to long-term engagement conversations with the agent [9].

Personalization can take many different forms and levels. For example, some levels of personalization are: conversation style (informal or formal, level of language complexity), choice of agent voice (accent, age and gender) and the content of conversations [10, 6].

The source of information to be used for personalization can be based on previous conversations or on user information and preferences that were explicitly encoded prior to the user-agent conversations [11].

In this paper we focus on one particular aspect of personalization, namely the creation of a user profile containing personal identifiable information (PII) and possibly uniquely identifiable information of the user.

Personalized conversational agents that allow for free language input are still rather uncommon and problematic. Montenegro et al. performed a systematic literature review of CAs in health (March 2019) and found that "[...] improvements to conversational agents related to interactions, interfaces and models of learning, with focus on facilitating user engagement." [12, p. 66] are necessary. Kocaballi et al. conducted a survey on personalized conversational agents in the healthcare domain that allow for free language input and were properly evaluated with users [10]. The authors only found 13 studies matching these criteria. Personalization is aimed a individuals (and not at groups), expressed at content level Laranjo et al. conclude that patient safety is

not evaluated in the discussed studies and is something that deserves more attention [13]. Kocaballi et al. states that "implicit methods used for gathering user information need to be clearly communicated to the users, since such methods often run automatically in the background, not being noticed by users. To this end, the model of informed consent for information systems may prove useful for considering various factors involved in collecting personal information" [10, 14]. We argue that a single consent form signed prior to the user-agents conversations is insufficient.

4. Challenges

In work on spoken CAs we observe a fundamental conflict between restrictions imposed by privacy regulations on the one hand, and personalization as a feature of the interaction on the other. Obviously this paradox is not limited to CAs and has been discussed for other applications as well [15]. As we aim to develop agents that have previous knowledge about their interlocutors and are capable of adapting to their needs, we need to explicitly model personal profiles and store information from previous conversations. Personalization also contributes to long-term engagement and is crucial for establishing credibility and trust of the CA [16], which in turn are essential for actually having an effective CA that can provide insights in personal health and even stimulate behavioural change like in coaching and counseling agents. Trust appears to be an essential element of long-term conversational engagement [9] as is needed in health related CAs. Having a personalized CA that has knowledge about previous conversations helps building such trust.

Not all personal identifiable information is equally private. Kim and Ko defined multiple levels of private information, depending on the needs of the users [17]. People are willing to share their favorite color and name of first pet much more likely than a voice signal or email [18]. Additionally, users prefer to share the most sensitive PII only with people they are intimate with (immediate family) or people who need it (doctors) [19]. In Figure 1, we plot two dimensions of sensitivity and intimacy regarding PII.

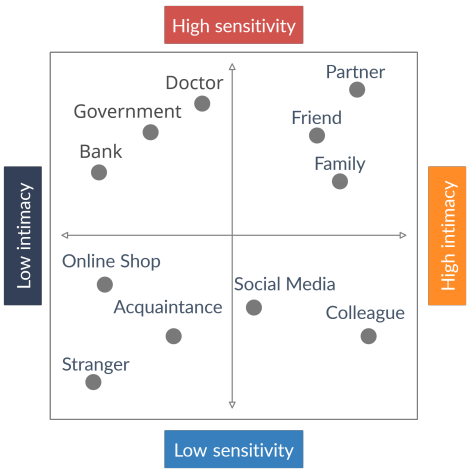

Figure 1: Two possible dimensions of user control for personal identifiable information: sensitivity and intimacy, based on [17] and [19].

The systematic review study on informed consent in biomedical research by De Sutter et al. [20] provides us with list of high-level recommendations that are largely applicable to our study but that need concrete implementation to be practically usable. For example, "pay attention to transparency regarding the use of participants' health information and their right to control the sharing and use of this information" (Table 1 from [20]). All this control might be difficult to reconcile with privacy regulations that were introduced to protect personal data such as GDPR, which require that information should not be traceable. So, for each piece of personal identifiable information that is incorporated in the

CA, users need to give previous explicit consent, which is excellent, but also makes the procedure lengthy and tedious. As we create conversational agents for research projects, we follow the regulations of proper research ethics and give participants a clear explanation of the research they engage in followed by signing a consent form, a very necessary step. After giving consent, the user has no longer influence on what is stored from the conversations. So explicit consent given upfront may not be sufficient. Users should also remain in control of their own data, especially when deployed in the field where a loss of control negatively impacts acceptance of a CA [11].

Furthermore, it has been shown that an interactive presentation of project information helps to improve understanding within the informed consent process [21].

5. Possible solutions

A possible way of ensuring more control on the part of the users would be to build in the dialogue a form of explicit consent or consolidation of the information provided by the user after each conversational exchange. In this way it is possible that information or data that were mentioned by users, but that they actually prefer not to share are not included in the dialogue history. One solution that we plan to investigate involves the use of summarized content reflection at the end of each conversation as a mechanism for transparent user control and explicit feedback on the correctness of the stored user information. We plan to compare the effectiveness and appreciation of such an automatic conversational summarizing method against giving participants an opportunity to correct a list of the extracted facts. We view a summary as a way of efficient "dialogue condensation"; it is not desirable to have the user verify each and every spoken utterance, but only the most important information nuggets. Classifying the type of user information (social, physical health, financial etc.) could be a helpful step to decide on which information should be stored and presented to the user.

Another useful feature would be to establish layers of accessibility for personal data, so that the information contained in the CA is accessed to different degrees by different individuals or groups like the users themselves, their family and/or care providers. Storing all this sensitive information poses enormous challenges in terms of safety and security. This might be less of problem when large health providers are involved, but for individual users this could be a bridge too far.

The security issue emerged clearly during the COVID-19 crisis which caused a shift to online solutions that are much more vulnerable. In our own research the effects of the pandemic led to an unexpected struggle. We had developed a software architecture that works on device, but this could not be used because we were not allowed to conduct experiments on the premises. So we had to switch to an online version, which caused a lot of additional work and in fact led to a less robust solution in terms of security.

6. Conclusion

Privacy regulations and personalization as a feature of a CA cannot be easily accommodated. The development of an efficient and user-friendly CA requires knowledge of the user, and as a consequence personal user profiles and information from previous conversations must be stored and re-used in a flexible way. Since this may include medical data, privacy is a matter of utmost concern.

All bits of personal identifiable information stored in the system require a form of consent, preferably explicitly by the user instead of im-

plicitly by for example agreeing on embarking on a conversation with a CA. In addition, controlling the data and accessibility of personal data is an aspect that needs careful consideration. This counts for both the design and architecture of the system but also (and, even more importantly) in the information flow along the entire data path during interaction, especially in the case where external web services are employed for the online functioning of the CA.

We argued that static and dynamic management by the user of stored personal identifiable information within agents can provide an adequate and powerful solution for resolving the privacy versus personalization conflict.

Acknowledgments

This work is part of the research programme Data2-Person with project no. 628.011.029, which is (partially) financed by the Dutch Research Council (NWO). We thank the reviewers for their valuable suggestions.

References

[1] J. van Waterschoot, I. Hendrickx, A. Khan, E. Klabbers, M. de Korte, H. Strik, C. Cucchiarini, M. Theune, BLISS: An agent for collecting spoken dialogue data about health and well-being, in: Proceedings of the 12th Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2020, pp. 449–458.

[2] M. Huber, et al., How should we define health?, BMJ (Online) 343 (2011) 1–3. doi:10.1136/bmj.d4163.

[3] A. K. Ostrowski, D. DiPaola, E. Partridge, H. W. Park, C. Breazeal, Older Adults Living With Social Robots: Promoting Social Connectedness in Long-Term Communities, IEEE Robotics Automation Magazine 26 (2019) 59–70. doi:10.1109/MRA.2019.2905234, conferenceName: IEEE Robotics Automation Magazine.

[4] S. Z. Razavi, L. K. Schubert, B. Kane, M. R. Ali, K. A. V. Orden, T. Ma, Dialogue design and management for multi-session casual conversation with older adults, in: Joint Proceedings of the ACM IUI 2019 Workshops, ACM, 2019, pp. 1–9.

[5] H. Abdollahi, A. Mollahosseini, J. T. Lane, M. H. Mahoor, A pilot study on using an intelligent life-like robot as a companion for elderly individuals with dementia and depression, in: 2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids), 2017, pp. 541–546. doi:10.1109/HUMANOIDS.2017.8246925, ISSN: 2164-0580.

[6] L. P. Vardoulakis, L. Ring, B. Barry, C. L. Sidner, T. Bickmore, Designing Relational Agents as Long Term Social Companions for Older Adults, in: Y. Nakano, M. Neff, A. Paiva, M. Walker (Eds.), Intelligent Virtual Agents, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2012, pp. 289–302. doi:10.1007/978-3-642-33197-8_30.

[7] A. van Maris, N. Zook, P. Caleb-Solly, M. Studley, A. Winfield, S. Dogramadzi, Designing Ethical Social Robots—A Longitudinal Field Study With Older Adults, Frontiers in Robotics and AI 7 (2020). doi:10.3389/frobt.2020.00001, publisher: Frontiers.

[8] V. Demberg, A. Winterboer, J. D. Moore, A strategy for information presentation in spoken dialog systems, Computational Linguistics 37 (2011) 489–539.

[9] T. W. Bickmore, R. W. Picard, Establishing and Maintaining Long-term Human-

computer Relationships, ACM Transactions on Computer-Human Interaction 12 (2005) 293–327. doi:10.1145/1067860.1067867.

[10] A. B. Kocaballi, S. Berkovsky, J. C. Quiroz, L. Laranjo, H. L. Tong, D. Rezazadegan, A. Briatore, E. Coiera, The personalization of conversational agents in health care: Systematic review, J Med Internet Res 21 (2019) e15360. doi:10.2196/15360.

[11] C. Tsiourti, E. Joly, C. Wings, M. B. Moussa, K. Wac, Virtual assistive companions for older adults: qualitative field study and design implications, in: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare, PervasiveHealth '14, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, 2014, pp. 57–64. doi:10.4108/icst.pervasivehealth.2014.254943.

[12] J. L. Z. Montenegro, C. A. da Costa, R. da Rosa Righi, Survey of conversational agents in health, Expert Systems with Applications 129 (2019) 56 – 67. doi:10.1016/j.eswa.2019.03.054.

[13] L. Laranjo, A. G. Dunn, H. L. Tong, A. B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A. Y. Lau, et al., Conversational agents in healthcare: a systematic review, Journal of the American Medical Informatics Association 25 (2018) 1248–1258.

[14] Y. O'Connor, W. Rowan, L. Lynch, C. Heavin, Privacy by Design: Informed Consent and Internet of Things for Smart Health, Procedia Computer Science 113 (2017) 653–658. doi:10.1016/j.procs.2017.08.329.

[15] N. F. Awad, M. S. Krishnan, The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization, MIS quarterly 30 (2006) 13–28. doi:10.2307/25148715.

[16] B. R. Cowan, N. Pantidi, D. Coyle, K. Morrissey, P. Clarke, S. Al-Shehri, D. Earley, N. Bandeira, " what can i help you with?" infrequent users' experiences of intelligent personal assistants, in: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, 2017, pp. 1–12.

[17] S.-H. Kim, I.-Y. Ko, How do multilevel privacy controls affect utility-privacy trade-offs when used in mobile applications?, ETRI Journal 40 (2018) 813–823. doi:https://doi.org/10.4218/etrij.2017-0259.

[18] V. Marmion, D. E. Millard, E. H. Gerding, S. V. Stevenage, The Willingness of Crowds: Cohort Disclosure Preferences for Personally Identifying Information, Proceedings of the International AAAI Conference on Web and Social Media 13 (2019) 358–368.

[19] A. Rapp, F. Cena, Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data, International Journal of Human-Computer Studies 94 (2016) 1–17. doi:10.1016/j.ijhcs.2016.05.006.

[20] E. De Sutter, D. Zaçe, S. Boccia, M. L. Di Pietro, D. Geerts, P. Borry, I. Huys, Implementation of electronic informed consent in biomedical research and stakeholders' perspectives: Systematic review, Journal of medical Internet research 22 (2020) e19129. doi:doi:10.2196/19129.

[21] E. E. Anderson, S. B. Newman, A. K. Matthews, Improving informed consent: Stakeholder views, AJOB empirical bioethics 8 (2017) 178–188.