Facilitating Responsible Use of Artificial Intelligence in the Public Sector Through Guidelines for Human-AI Interaction

Robert Thor Arnarsson[1*], Mats Ellingsen[1*], Anh Nguyen Pham[1*], Jarl Sterkenburg[1*], Gard Cappelen[1¨], Stefan Hochwarter[2^], Stefan Schmager[3^]

[1]Norwegian University of Science and Technology, Department of Computer Science, Norway, *jjsterke@stud.ntnu.no*
[2]University of Oslo, Department of Informatics, Norway
[3]University of Agder, Department of Information Systems, Norway

* These authors share first authorship
¨ These authors share second authorship
^ These authors share senior authorship

Abstract

When developing Artificial Intelligence (AI) systems and implementing them in public services, guidelines for human-AI interaction can be instrumental in ensuring responsible use of these systems and AI intelligibility and responsibility, protecting the citizens' rights and interests. Although there are as of yet no such guidelines established that are directed at the public sector, tech companies like Microsoft have established guidelines for the commercial sector. To investigate how commercial guidelines may help guide the development of public services that utilize AI, we have explored how *Microsoft Human-AI Interaction Guidelines* can be amended and used to guide the development of a sick leave duration prediction service. By developing a web application prototype for the specific case and conducting usability testing, we found that the guidelines needed amendments to better ensure responsible use of AI in a public service. Three considerations were discovered to be important when amending them: (i) *the way users interact with the AI system*, (ii) *the diversity of citizens making use of the service*, and (iii) *the public service context*.

Introduction and background

The utilization of Artificial Intelligence (AI) is increasingly being introduced to the public sector, and its use may improve the delivery of public services to citizens. In the commercial sector, prominent actors such as Microsoft, Google and Apple have crafted AI guidelines to ensure responsible and human-centered development and implementation of AI systems. There is a lack of guidelines and regulation aimed at implementing AI in public services, and the need for such guidelines is more urgent than ever. Recently, members of the Norwegian Parliament issued a proposal to ask the Norwegian government to introduce a moratorium in the public sector on adopting new commercial tools based on AI, until regulations for the introduction and use of such tools have been developed (Fylkesnes & Hussein, 2023). This paper aims to contribute to a research effort addressing the need for such human-centered guidelines for the implementation of AI systems in public services. Specifically, as actors in the commercial sector already have established such guidelines, our research question is: *How can commercial AI guidelines be used to facilitate the responsible use of AI in public services for citizens?*

The empirical case for this study was an online public service related to receiving support during sick leave. Specifically, it addresses a scenario where a public service offers utilization of an AI system to predict sick leave duration, to better plan for long term support. The central issue was how to design, with AI intelligibility and responsibility in mind, a user experience where users would receive information about personal information used, the process depending on the user's choice, how the AI system works, and finally choose to consent or dissent to the utilization of the AI system. The Norwegian Labour and Welfare Administration (NAV) has expressed an intention of implementing the utilization of an AI system in such a public service (Nav – sluttrapport, 2022). Schmager (2022) addresses a similar case and explores how Google's PAIR guidelines can be amended to the public service context. Our research addresses the application and amendment of commercial guidelines for developing public services. The novelty of our research is a specific consideration of the *Microsoft Human-AI Interaction Guidelines* (MHAII guidelines). These guidelines were chosen because of their emphasis on user interaction and basis in academic literature, as Wright et al. (2020) indicates in a comparative analysis.

Our research is a subproject of the AI4Users project, consisting of researchers from the University of Agder (UiA), the University of Oslo (UiO), and the Norwegian University of Science Technology (NTNU). The research and work of the AI4Users project addresses the responsible use of AI in public services, especially the aspects of intelligibility and accountability, in relation to the needs of different user groups (Vassilakopoulou *et al.*, 2022).

Methods

The research follows an Action Design Research (ADR) methodology, and our focus is the stage of 'Building, Intervention, and Evaluation' (BIE) (Sein et al., 2022). To evaluate how the selected MHAII guidelines would work in a public service, we used them to guide development of a prototype for the empirical case and conducted usability testing to evaluate both the prototype and guidelines. This cycle was repeated for iterative development and evaluation. The prototype was developed as a web application by using an agile approach that facilitated regular user testing of new features and adoption of the protype.

Three iterations of usability testing provided insights into usability and the effects of using the guidelines in the different iterations of the prototype and were performed in two ways. The first approach was to share a link to the prototype hosted online, letting the users test the prototype on their own. The second approach involved usability testing in a semi-controlled environment, usually in person in the user's home or through digital communication platforms Users were asked to be transparent about their thoughts and intentions, in line with the *Think-aloud technique* (Sharp *et al.*, 2019). A facilitator would observe and log comments and the user's interaction with the prototype. Both approaches ended with the user filling out a questionnaire that was specific to the functionality implemented and the guidelines used in each iteration of the prototype. Questions were either open ended, dichotomous, or using a five-point Likert Scale (Sharp *et al.*, 2019). The method for selecting participants for the usability tests was convenience sampling (Sharp *et al.*, 2019). The first, second and third round of usability testing had 11, 23 and 24 participants.

Findings

We found that the guidelines helped narrow in on features to implement in the prototype, and how these features should be designed. It was however discovered a need to amend the chosen guidelines to better fit the nature of the empirical case. We also found that the MHAII guidelines were not sufficient in accounting for all ethical and legal considerations of a public service. For example, as Schmager (2022) points out, the government-citizen relationship includes rights aimed at protecting the citizen. Also, since public services are meant to address citizens' needs, they must account for a diverse range of administrative and technical literacy. Rights are not mentioned in any MHAII guideline, and accounting for a user group's diversity in administrative and technical literacy is not mentioned explicitly. Taking this into account, we found three considerations to be particularly important when amending the guidelines: (i) *the way users interact with the AI system*, (ii) *the diversity of*

citizens making use of the service, and (iii) *the public service context*. The amendments were stated in a table (see table 1.1), together with what features in the prototype adhered to the guideline and the source of the guideline.

Table I: An example of an amended guideline.

3	*Avoid impeding refusal* To ensure the users' freedom of choice, the user should not feel disincentivized towards declining the usage of the AI system.	-Not mentioning downsides of rejecting the usage of the AI model. -The buttons for consenting and not consenting have the same visual hierarchy. -The summary page provides a description of how the user can change their choice.	MHAII Guideline 8: *Support efficient dismissal.*

Usability testing also provided interesting findings. In usability testing iterations 2 and 3, users were asked what feature of the prototype was most instrumental for the user to understand how the AI model worked. In iteration 2, the share of users that thought illustrations, a sandbox feature or the textual content was most instrumental were almost equally large, with 35%, 35% and 30% in favor of each category, respectively. In iteration 3, we added feature importance charts to the prototype and changed the question to multiple choice. 54% found illustrations instrumental, 50% a sandbox feature, 25% a description of feature importance, and 63% the textual content. The user's correct understanding of who or what was responsible for the usage of the AI system and its results gradually increased with each iteration, from 64% in iteration 1 to 74% in iteration 2, and finally 96% in iteration 3. 52% answered that they felt an incentive to accept the usage of the AI system in iteration 2. In iteration 3 we changed the phrasing of the question, and 33% answered that they felt pressure to accept.

Concluding discussion

The resulting prototype based on the case, the selected MHAII guidelines, and user feedback, indicates that commercial responsible AI guidelines can be used in the public sector, but must be amended to ensure AI intelligibility and accountability. The first amendment consideration (i), *the way users interact with the AI system*, was necessary because the MHAII guidelines assume direct interaction between the AI system and the user. Our empirical case does however not include direct interaction between the citizen and the system. Instead, the user consents or dissents to a case handler's usage of such a system to process the user's personal information. This consideration led us to reformulate guidelines that address direct usage in a way that kept their essence intact, attempting to instead address the act of deciding whether to allow such a system. The second consideration (ii), *the diversity of citizens making use of*

the service, was deemed necessary as none of the MHAII guidelines specifies how information should be conveyed to accommodate for diversity. Because many public services are used by a wide range of users, the guidelines were amended to emphasize the need for information and interactive elements that take this into account. The final consideration (iii), *the public service context,* was necessary since the MHAII guidelines were developed for a commercial context, which differs from the Norwegian public service context in multiple ways. For example, the government-citizen relationship makes the importance of citizens' rights in relation to the service important to highlight, and the MHAII guidelines have no mention of rights. Public services are also less avoidable than commercial ones. For instance, people on sick leave might be dependent on receiving support to cover fundamental needs. We accounted for this in the amendment of our chosen guidelines by being transparent about benefits and disadvantages of allowing the usage of the AI system, both in the public and individual context.

The usability test results confirmed the importance of some guidelines, and the amendments of these. The close to even distribution of users that found text, illustrations, or interactive sandbox most instrumental for understanding how the AI system worked underlines the importance of providing different explanations of AI system behavior suited for users with varying degrees of previous knowledge. One particularly interesting guideline throughout the project was MHAII guideline number 8, *"Support efficient dismissal"*, which emphasizes the ability to dismiss or ignore undesired AI system services when the system is not working as expected (Guideline 8: Support efficient dismissal, 2019). In relation to consideration (i), we amended this guideline to instead underline the importance of making it easy to dismiss the service altogether. Usability testing iteration 3 found that many users (33%) felt pressured to consent to the usage of the AI model. In light of consideration (iii), there are two main concerns. The first being the EU right to not be subjected to a decision solely based on automated processing which produces legal effects (Guidelines on Automated …, 2018). The second being the societal benefit of utilizing AI systems in a public service, which in turn also benefits individuals. The question then becomes, how much pressure to consent to the usage of AI is an admissible amount for the user to feel, considering the public benefits? We settled on amending the guideline to emphasize not disincentivizing declining the usage of the AI system, to still allow information about the benefits of using the system to be presented. In our final prototype, these considerations are taken into account by making the buttons for consent and dissent equal in visual hierarchy, and by explaining the benefits of the usage, such as shorter processing time.

References

'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01)', 2018, Retrieved May 21, 2023 from https://ec.europa.eu/newsroom/article29/items/612053/en

Fylkesnes, T. K., Hussein, M. (2023): 'Representantforslag om demokratisk kunstig intelligens', Dokument 8:232 S (2022–2023), Retrieved June 18, 2023 from https://www.stortinget.no/no/Saker-og-publikasjoner/Publikasjoner/Representantforslag/2022-2023/dok8-202223-232s/?fbclid=IwAR2pvLr_J5v8pnOr8NcbKZWnbFtI58bBjmr9IF4xoFMvyy-ddGy7c3RTyNs

'Guideline 8: Support efficient dismissal', 2019, Retrieved May 29, 2023 from https://www.microsoft.com/en-us/haxtoolkit/guideline/support-efficient-dismissal/

'Nav – sluttrapport', 2022, Retrieved May 29, 2023 from https://www.datatilsynet.no/regelverk-og-verktoy/sandkasse-for-kunstig-intelligens/ferdige-prosjekter-og-rapporter/nav-sluttrapport/

Schmager, S (2022): 'From commercial agreements to the social contract: human-centered ai guidelines for public services', *The 14th Mediterranean Conference on Information Systems (MCIS)*, October 2022. https://www.researchgate.net/publication/364345425_From_commercial_agreements_to_the_social_contract_Human-Centered_AI_guidelines_for_public_services

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., and Lindgren, R. (2011): 'Action design research', *MIS quarterly*, vol. 35, no. 1, March 2011, pp. 37-56. https://doi.org/10.2307/23043488

Sharp, H., Rogers, Y. and Preece, J. (2019): *Interaction design: Beyond human-computer interaction*, John Wiley & Sons, Indianapolis, 5th edition.

Vassilakopoulou, P., Parmiggiani, E., Shollo, A., & Grisot, M. (2022): 'Responsible AI: Concepts, critical perspectives and an Information Systems research agenda', Scandinavian Journal of Information Systems, vol. 34, no. 2, December 2022, pp. 89-112. https://aisel.aisnet.org/sjis/vol34/iss2/3

Wright, A. P., Wang, Z. J., Park, H., Guo, G, Sperrle, F., El-Assady, M., Endert, A., Keim, D. and Chau, D. H. (2020): 'A comparative analysis of industry human-AI interaction guidelines', arXiv preprint arXiv:2010.11761. https://arxiv.org/abs/2010.11761