



HAL
open science

Towards Co-Designing a Continuous-Learning Human-AI Interface: A Case Study in Online Grooming Detection

Peter Daish, Nicholas Micallef, Nuria Lorenzo-Dus, Adeline Paiement, Deepak Sahoo

► **To cite this version:**

Peter Daish, Nicholas Micallef, Nuria Lorenzo-Dus, Adeline Paiement, Deepak Sahoo. Towards Co-Designing a Continuous-Learning Human-AI Interface: A Case Study in Online Grooming Detection. International Conference on Advanced Visual Interfaces (AVI) workshop - Designing and Building Hybrid Human-AI Systems (SYNERGY 2024), Jun 2024, Genoa, Italy. <hal-04556796v2>

HAL Id: hal-04556796

<https://hal.science/hal-04556796v2>

Submitted on 15 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Towards Co-Designing a Continuous-Learning Human-AI Interface: A Case Study in Online Grooming Detection

Peter Daish^{1,*}, Nicholas Micallef¹, Nuria Lorenzo-Dus², Adeline Paiement³ and Deepak Sahoo¹

¹Department of Computer Science, Swansea University, Swansea, Wales, UK

²Department of Applied Linguistics, Swansea University, Swansea, Wales, UK

³Université de Toulon, Aix Marseille Univ, CNRS, LIS, Marseille, France

Abstract

Interest is growing in using Human-Centered design to enhance compatibility of AI within human environments. These design techniques are valid for eliciting human-centered design requirements, however, they often paint the scenario that AI interaction design is a one-way process in which user behaviour is captured to improve interactions and user experiences. Such an approach does not consider real-world settings in which Human-AI environments involve multiple stakeholders, with contrasting needs, which could impact the interactivity, usability and usefulness of Human-AI environments. In this paper, we present a framework for incorporating multiple-stakeholders perspectives into the design of Human-AI environments, designed to establish a common dialogue between end-users' needs for Human-AI interaction and AI developers' practical limitations. This is a work in progress project and in our future work we plan to follow this iterative prototyping approach to develop a real-world continuous-learning Human-AI detection system for online grooming.

Keywords

Human-AI, Continuous-learning, Human-centered design, Online grooming

1. Introduction

Humans are increasingly embedding Artificial Intelligence (AI) technologies into their everyday lives, with use-cases ranging from personal through to professional settings. These AI technologies are either designed to replace humans, or else to support them, depending on the stances taken by various stakeholders in the development of the project. As an example, autonomous vacuum cleaners might typically be designed with the mutual understanding of the designers, developers and end-users that the robot will replace the human in the task of vacuuming around

Proceedings of the 1st International Workshop on Designing and Building Hybrid Human-AI Systems (SYNERGY 2024), Arenzano (Genoa), Italy, June 03, 2024


*Corresponding author.

✉ peter.daish@swansea.ac.uk (P. Daish); nicholas.micallef@swansea.ac.uk (N. Micallef); n.lorenzo-dus@swansea.ac.uk (N. Lorenzo-Dus); adeline.paiement@univ-tln.fr (A. Paiement); d.r.sahoo@swansea.ac.uk (D. Sahoo)

🆔 0009-0006-6812-1791 (P. Daish); 0000-0002-2683-8042 (N. Micallef); 0000-0001-5114-1514 (A. Paiement); 0000-0002-4421-7549 (D. Sahoo)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

their home [1]. By contrast, AI clinical decision-support systems are developed with the mutual understanding of developers and clinicians that AI should provide a supporting role to the expert clinician - with whom the ultimate clinical decision lay - for enhanced decision-making [2]. Both systems convey a consensus of the roles of the human and AI within the hybrid Human-AI environment, however, the design of both systems could look very different depending on which stakeholders' perspectives are given precedence. An AI-centered vacuum cleaner may be excellent at recognising dirt within the room, but it has not been designed to accommodate the human-factors associated with cleaning the house, meaning it may: start disruptive cleaning cycles, potentially cause a safety concern ¹ or otherwise be incompatible with dynamic personal living environments [1] based on its technology-centered design [3]. The incompatibility of AI-powered systems with dynamic environments, could result in a high attrition rate in the sustained use of AI-based technologies over time [1]. Prioritizing the end-users' concerns might yield a different Human-AI interaction scenario: one which focuses on empowering - rather than replacing - humans through human-centered AI design philosophies (HCAI) [4]. To achieve this, AI developers and end-users can be modeled as stakeholders in the design of Human-AI interfaces, each with their own set of interests, which may align or compete with one another. This necessitates a paradigm shift from traditional HCI research approaches, wherein researchers study stakeholders merely as 'users', towards a more inclusive participatory or co-design protocol that aims to solve human-issues through the co-creation of solutions [5, 6].

Despite renewed efforts to place humans at the forefront of AI research and development, there are still few established norms on how to balance multiple-stakeholders' perspectives in the design and development of such systems. Whilst current HCAI research propose design values to be upheld during the development of AI [7, 8], these are typically ambiguous and provide no specific guidelines for the development of Human-AI technology in the multi-stakeholder scenario we describe. Thus, it is unclear how to achieve human-centered AI ideals in practice. This phenomenon is not new, however, with distinct parallels to the elusive 'art' of requirements engineering having been documented in AI development harking back to the initial rise of expert-system and rule-based AI [9]. It is also unclear what the best-practices are to develop 'human-centered' Human-AI interfaces for systems wherein the interaction designers have limited access to the end-users due to their availability. This is of particular concern in settings where AI researchers and developers work with public sector agencies - who typically can provide only limited access to their time and data and are unable to participate in detailed contextual inquiries- to incorporate AI technology into their professional workflows. Thus, our research question was: **How can we develop a human-centered iterative prototyping framework to be used to design a continuous-learning Human-AI environment that involves multiple stakeholders?**

In this paper, we propose an iterative prototyping framework for designing an interface for a continuous-learning feature within a Human-AI environment that incorporates multiple stakeholders' perspectives into the design, including limited access to end-users. The case-study that we will use in future work to validate our framework is the design of a 'continuous-learning' feature for an AI online grooming detection tool that will be used by law-enforcement agencies. The tool will aid law enforcement's online grooming detection work. The AI algorithm for this

¹<https://www.bbc.com/news/world-asia-67354709>

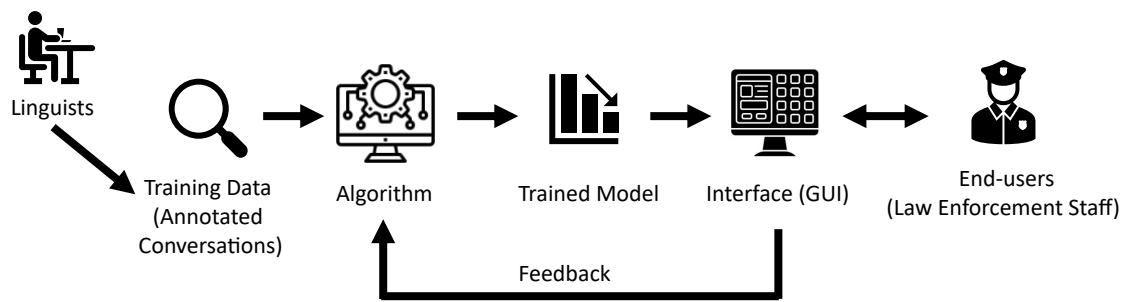


Figure 1: Continuous-learning technique within the Human-AI environment.

tool uses a “reinforcement-learning” approach to learn, which means it can improve at detecting cases of grooming upon receiving feedback from its end-users. To facilitate quality feedback being fed into the AI algorithm for further training, the tool is adding “continuous-learning” to allow end-users to provide feedback about the performance of the AI model, to adapt the model to real-world performance. Figure 1 shows how continuous-learning will be integrated within the Human-AI environment in this scenario.

2. Background

2.1. Humans ‘in’ and ‘on’ the loop

Typical approaches to Human-AI hybrid decision-making processes can be found in so-called ‘Human-In-The-Loop’ (HITL) and ‘Human-on-the-Loop’ (HOTL) design paradigms for interactive Machine-Learning. HITL systems typically include humans as participants in decisions alongside AI and in scenarios where the AI utilises a reinforcement algorithm, humans assess the quality of AI output and provide corresponding feedback to the AI’s model for real-time learning [10]. HOTL systems also maintain human supervision over algorithmic decision-making behaviour, however, the human is further removed from the (largely autonomous) decision-making process and only intervenes when they believe the machine has made a mistake [11].

2.2. Continuous-Learning

The roots of the HITL concept of AI continuous-learning can be found in the field of robotics, wherein the pursuit of the ability of robots to continuously learn and adapt to their environments over time was seen as pivotal in their usefulness in real-world dynamic environments [12]. Not only does the re-introduction of the human perspective help to democratise AI [13], it has also been shown to help maintain the performance of AI systems over time [14, 15], in addition to enabling them to learn new concepts [16] through reinforcement-learning with human feedback (RLHF) techniques.

Supervised reinforcement loops have also proven effective in tailoring AI models to local

contextual data, through transfer-learning fine tuning techniques [14]. This is important, since there is often limited data availability for locally deployed contexts (such as UK-based grooming data). Pre-training an algorithm on related large datasets and then fine-tuning on local domain data over time, merges the concept of transfer-learning with continuous learning and could enable the AI to maintain its performance despite the introduction of new concepts [17, 18, 14].

Whilst HITL environments have many different possible configurations, they typically revolve around an AI-centered design, where an AI agent makes predictions (or decision recommendations) and their human counterpart uses the AI's advice to come to a decision. Whilst this traditional HITL configuration has proven effective in continuous-learning systems [14], it relies on the human to spot erroneous predictions in real-time. Furthermore, owing to the lack of transparency in many blackbox algorithmic systems, the human supervisor has little understanding as to why the AI made its (seemingly erroneous) decision, or at best, has to be trained to comprehend technical AI explainability techniques.

Additionally, recent work has explored a more natural feedback loop for continuous-learning systems, where the system has been designed to improve through human-centered Human-AI interaction [19, 20], rather than direct annotation of training data. However, little is understood regarding how best to design this interactive environment from a multiple stakeholder perspective, which we attempt to address with our proposed framework.

2.3. Intelligent UIs and Their Lack of Human-Centered Design

There is ample literature that explores how RLHF techniques can be used for the automated adaptation of User Interfaces (UIs) for end-users [21, 22, 23], however, best practices in UI design to support RLHF from the end-users' perspective are largely unexplored. Indeed, even among those that aim to support RLHF with what are often considered to be "human-centered" techniques - such as Explainable AI (XAI) - often take an AI-centered approach [24]. This means that, whilst a continuous-learning system may be able to 'explain' its reasoning or else be capable of intelligently switching interfaces for increased efficiency, little is known as to why the new interface is more effective for end-users or, indeed, whether the end-users find the feedback system suitable for the target domain or given their technical proficiency in understanding AI behaviour. Finally, in contexts where data annotators, algorithm developers and end-users' experience of the target domain is asymmetrical, best practices for establishing how to incorporate the expertise of each stakeholder within the human-centered approach to RLHF are yet to be established.

2.4. Iterative Co-Design

Multiple techniques are available for Human-AI interaction designers to capture design-requirements from end-users, including: ethnographic field work [25] and contextual inquiries [26]. Whilst both of these methods help developers and designers understand the use-case of the tool within the target domain, they are still typically used as one-dimensional requirement gathering tools, owing to their largely observational nature. As a result, it is, up to the designers and developers of the Human-AI environment to identify - in their limited experience of the domain by observing or through limited interactions with end-users - which

features should be included in the Human-AI environment, how they should look and moreover, how they should function. Co-design protocols have been mooted as a means of bringing end-users back into the design process, to innovate empowering designs crafted through the lived experiences of end-users in a manner which hasn't been pre-determined by technological constraints [27]. Literature in studying the co-design of Human-AI environments is emerging, however, these works either focus on empowering users to create AI models themselves [28], or conclude with user-centered role-play scenarios which do not consider the technological feasibility of the designs produced for use in Human-AI environments [29, 30]. As a result and despite the co-design protocol facilitating interaction between designers and end-users, there remains a gap in knowledge regarding how these techniques can be deployed to facilitate Human-AI environment design in settings with multiple stakeholders whose perspectives and opinions on the operating criteria of a system are asymmetrical.

Whilst evidence in their use in the design of Human-AI interaction is yet to emerge, prior work in interactive design has explored the potential of provocative prototypes 'provotypes' to explore tensions between multiple stakeholders' perspectives on the intended and actual use of a system (or device) [31]. Provotypes are tangible artifacts designed to embody tensions in perspective between stakeholders, enabling collaborative deliberation and analysis of these by all stakeholders [32, 33]. A 'provotype' methodology utilises multiple techniques to tease out tensions between stakeholders (and thus inform the design of the 'provotype'), such as interviews and ethnographic field work. However, 'provotypes' are typically deployed in scenarios where stakeholders share a common domain of expertise [32], meaning that their use in provoking tensions in systems designed to embed knowledge across multiple domains is currently unexplored.

To address the various challenges outlined in this section, we propose a methodological framework inspired by iterative co-design techniques to incorporate multiple-stakeholders' perspectives - all of whom are experts in their own domains - into the design of a human-centered, hybrid, Human-AI continuous-learning system to aid expert decision-makers. In the next section, we describe our proposed methodological framework to incorporate linguistic, law-enforcement and AI expertise into the design of a UI capable of eliciting useful feedback from end-users to continuously evolve AI's capabilities in online grooming detection.

3. Method

In this section, we detail the proposed methodological framework that we will follow to evaluate our iterative prototyping approach. The following subsections describe our proposed iterative prototyping framework and how we will recruit participants for our evaluation.

3.1. Proposed Iterative Prototyping Framework

As illustrated in Figure 2, our iterative prototyping framework consists of two phases. In the first phase we will hold interviews with all stakeholders to capture their perspectives and expectations on how the continuous-learning feature should function. We will start this phase by interviewing linguistics experts because they are the stakeholders that initiated the development of the tool and are experts in the scientific pursuit of online grooming detection.

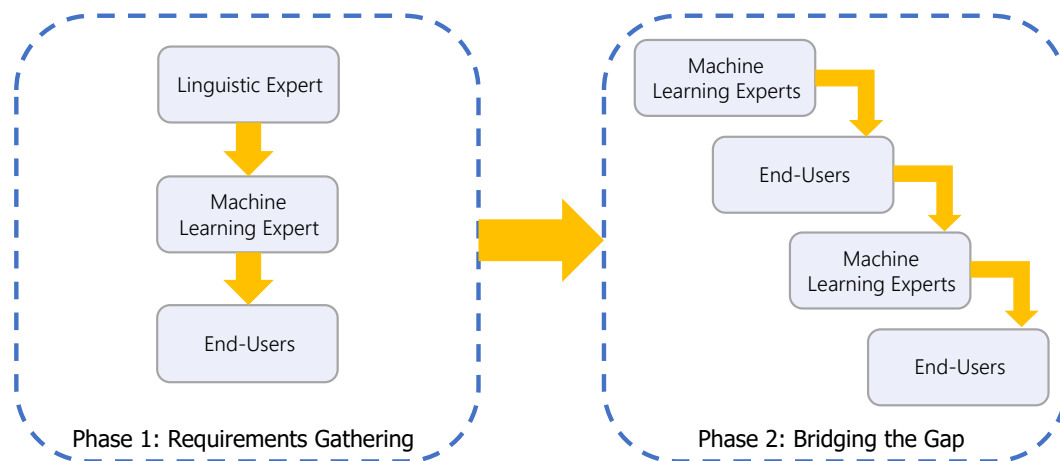


Figure 2: Iterative Prototyping Framework proposed by this research.

Afterwards, we will interview machine learning experts, since the continuous-learning feature will be integrated within the machine learning process, meaning that the perspectives of this stakeholder will be critical to the design process. To close Phase 1, we will interview the end-users (i.e., law enforcement staff) to capture their perspectives and expectations. Due to the different roles of the stakeholders, we will vary the questions that we will ask in the interviews. We will start the interviews by gathering general information about the users' professional roles and their associated tasks. Then our inquiry will turn towards their roles within the AI online grooming detection system: specifically, we will consult with the users on how they perceive an AI classification tool to work. We will solicit this feedback by showing them an initial screen of the tool. After establishing the users' mental models of the AI, we will ask them how they would expect to provide feedback to the AI classification of conversations. After eliciting their thoughts on the interactive feedback loop, we will show the users a number of prototype designs intended to convey both the AI's classification of the conversations and suggested continuous-learning interactions. Among these prototypes we will show designs developed through the iterative feedback that we will receive from Linguistics and Machine Learning experts. We selected this order of inquiry so that end-users will not be biased with pre-conceptions as to how the continuous-learning mechanism 'should' work.

The first phase should provide a better understanding of the stakeholders' perspectives and expectations of how they would like the continuous-learning technique to be integrated within the tool. We expect that Phase 1 will surface a gap between our stakeholders on how the continuous-learning Human-AI environment should work. For this assumption we included Phase 2 in our framework, which will attempt to bridge the gap. In Phase 2, we will use a different approach and ask more specific questions with the aim of achieving consensus across our stakeholders. We will start this phase by interviewing machine learning experts to discuss any differences in expectations between their technical needs and the end-users needs.

Afterwards, we will interview end-users to learn more about how they perceive providing feedback to the machine learning model through the continuous-learning technique. To confirm that we managed to bridge any gaps between these two stakeholders, we will conduct a second round of interviews with machine learning experts and end-users (see Figure 2).

3.2. Participant Recruitment Strategy

Our collaborators will assist us in defining the sample that will participate in our evaluation. The Principal Investigator of the project will introduce us to machine learning experts and various end-users that will likely use the designed. To maximize our research efforts, we seek to recruit highly experienced professionals with more than 10 years experience in their specific domain. We plan to involve the same set of end-users in Phase 1 and Phase 2 so that the end-users would be able to see how the design of the continuous learning feature evolves based on their feedback.

3.3. Data Analysis Strategy

To analyze the qualitative data that we will collect during the interviews, we will follow an inductive thematic analysis approach by using an adapted version of the method used by Braun and Clarke [34]. We will conduct the analysis in the following stages: familiarization with data, identification of themes, review of themes, discussion and finalization of themes. In the first two stages the researcher will review the transcripts and extract the themes. Subsequently, meetings will be held with a second researcher to review and discuss the themes. The meetings will define the themes that will form the higher-level concepts that connect the information that will be extracted from the transcripts.

4. Concluding Remarks

This work is in progress and requires some discussion to further improve our proposed iterative prototyping framework. Within this paper, we presented a gap in the literature regarding the current limit of academic understanding regarding Human-AI interaction scenarios which model the Users and the AI developers and Linguistics Experts as stakeholders in a multi-stakeholder project, each with their own perspectives and understanding of the target domain. Given inspiration from AI-centered literature, we anticipate that typical software development workflows might surface gaps in understanding and perspective between stakeholders and we envision the Human-AI interaction designers role as mediator to find compromises between technological limitations and human-centered design principles. In our future work, we will validate our framework through a case-study in Human-AI online grooming detection that will contribute towards exploring the current knowledge gap in literature regarding driving human-centered Human-AI interaction for Continuous-Learning systems when working with multiple stakeholders.

Acknowledgments

This study is part of project DRAGON-S, a research and development project that is headquartered in the UK that is primarily funded by Safe Online Initiative at End Violence.

References

- [1] F. Vaussard, J. Fink, V. Bauwens, P. Rétornaz, D. Hamel, P. Dillenbourg, F. Mondada, Lessons learned from robotic vacuum cleaners entering the home ecosystem, *Robotics and Autonomous Systems* 62 (2014) 376–391. doi:10.1016/j.robot.2013.09.014.
- [2] C. Giordano, M. Brennan, B. Mohamed, P. Rashidi, F. Modave, P. Tighe, Accessing artificial intelligence for clinical decision-making, *Frontiers in Digital Health* 3 (2021). URL: <https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2021.645232>. doi:10.3389/fdgth.2021.645232.
- [3] S. Doncieux, R. Chatila, S. Straube, F. Kirchner, Human-centered ai and robotics, *AI Perspectives* 4 (2022) 1. doi:10.1186/s42467-021-00014-x.
- [4] C. P. University of Maryland, B. Shneiderman, Human-centered artificial intelligence: Three fresh ideas, *AIS Transactions on Human-Computer Interaction* (2020) 109–124. doi:10.17705/1thci.00131.
- [5] P. Nemirovsky, Redefining digital audience: Models and actions, in: *IFIP TC13 International Conference on Human-Computer Interaction, 2003*. URL: <https://api.semanticscholar.org/CorpusID:11060703>.
- [6] M. Maceli, M. E. Atwood, “human crafters” once again: Supporting users as designers in continuous co-design, in: Y. Dittrich, M. Burnett, A. Mørch, D. Redmiles (Eds.), *End-User Development*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 9–24.
- [7] W. Xu, Z. Gao, M. Dainoff, An hcai methodological framework: Putting it into action to enable human-centered ai (2023). URL: <https://arxiv.org/abs/2311.16027>. doi:10.48550/ARXIV.2311.16027.
- [8] B. Shneiderman, Human-centered ai: ensuring human control while increasing automation, in: *Proceedings of the 5th Workshop on Human Factors in Hypertext, HUMAN '22*, Association for Computing Machinery, New York, NY, USA, 2022. URL: <https://doi.org/10.1145/3538882.3542790>. doi:10.1145/3538882.3542790, event-place: Barcelona, Spain.
- [9] J. K. Ang, S. B. Leong, C. F. Lee, U. K. Yusof, Requirement engineering techniques in developing expert systems, in: *2011 IEEE Symposium on Computers & Informatics, 2011*, pp. 640–645. doi:10.1109/ISCI.2011.5958991.
- [10] G. Li, R. Gomez, K. Nakamura, B. He, Human-centered reinforcement learning: A survey, *IEEE Transactions on Human-Machine Systems* 49 (2019) 337–349. doi:10.1109/THMS.2019.2912447.
- [11] N. Li, S. Adepu, E. Kang, D. Garlan, Explanations for human-on-the-loop: a probabilistic model checking approach, in: *Proceedings of the IEEE/ACM 15th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, ACM, Seoul Republic of Korea, 2020*, p. 181–187. URL: <https://dl.acm.org/doi/10.1145/3387939.3391592>. doi:10.1145/3387939.3391592.

- [12] S. Thrun, T. M. Mitchell, Lifelong robot learning, *Robotics and Autonomous Systems* 15 (1995) 25–46. URL: <https://www.sciencedirect.com/science/article/pii/092188909500004Y>. doi:[https://doi.org/10.1016/0921-8890\(95\)00004-Y](https://doi.org/10.1016/0921-8890(95)00004-Y), the Biology and Technology of Intelligent Autonomous Agents.
- [13] S. Amershi, M. Cakmak, W. B. Knox, T. Kulesza, Power to the people: The role of humans in interactive machine learning, *AI Magazine* 35 (2014) 105–120. doi:<https://doi.org/10.1609/aimag.v35i4.2513>.
- [14] O. S. Pianykh, G. Langs, M. Dewey, D. R. Enzmann, C. J. Herold, S. O. Schoenberg, J. A. Brink, Continuous learning ai in radiology: Implementation principles and early applications, *Radiology* 297 (2020) 6–14. doi:[10.1148/radiol.2020200038](https://doi.org/10.1148/radiol.2020200038).
- [15] A. Holzinger, M. Plass, M. Kickmeier-Rust, K. Holzinger, G. C. Crişan, C.-M. Pintea, V. Palade, Interactive machine learning: experimental evidence for the human in the algorithmic loop, *Applied Intelligence* 49 (2019) 2401–2414. doi:[10.1007/s10489-018-1361-5](https://doi.org/10.1007/s10489-018-1361-5).
- [16] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, G. Zhang, Learning under concept drift: A review, *IEEE Transactions on Knowledge and Data Engineering* 31 (2019) 2346–2363. doi:[10.1109/TKDE.2018.2876857](https://doi.org/10.1109/TKDE.2018.2876857).
- [17] K. Yang, J. Tao, J. Lyu, C. Ge, J. Chen, Q. Li, W. Shen, X. Zhu, X. Li, Using human feedback to fine-tune diffusion models without any reward model (2024). URL: <http://arxiv.org/abs/2311.13231>, arXiv:2311.13231 [cs].
- [18] D. M. Ziegler, N. Stiennon, J. Wu, T. B. Brown, A. Radford, D. Amodei, P. Christiano, G. Irving, Fine-tuning language models from human preferences (2020). URL: <http://arxiv.org/abs/1909.08593>, arXiv:1909.08593 [cs, stat].
- [19] S. Counts, J. Neville, M. Wan, R. W. White, L. Yang, Learning from interaction with microsoft copilot (web), 2024. URL: <https://www.microsoft.com/en-us/research/blog/learning-from-interaction-with-microsoft-copilot-web/>.
- [20] Y.-C. Lin, J. Neville, J. W. Stokes, L. Yang, T. Safavi, M. Wan, S. Counts, S. Suri, R. Andersen, X. Xu, D. Gupta, S. K. Jauhar, X. Song, G. Buscher, S. Tiwary, B. Hecht, J. Teevan, Interpretable user satisfaction estimation for conversational systems with large language models (2024). URL: <https://arxiv.org/abs/2403.12388>. doi:[10.48550/ARXIV.2403.12388](https://doi.org/10.48550/ARXIV.2403.12388).
- [21] L. Zouhaier, Y. B. D. Hlaoui, L. B. Ayed, A reinforcement learning based approach of context-driven adaptive user interfaces, in: 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), 2021, pp. 1463–1468. doi:[10.1109/COMPSAC51774.2021.00217](https://doi.org/10.1109/COMPSAC51774.2021.00217).
- [22] K. Todi, G. Bailly, L. Leiva, A. Oulasvirta, Adapting user interfaces with model-based reinforcement learning, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, ACM, Yokohama Japan, 2021, p. 1–13. URL: <https://dl.acm.org/doi/10.1145/3411764.3445497>. doi:[10.1145/3411764.3445497](https://doi.org/10.1145/3411764.3445497).
- [23] T. Langerak, S. Christen, M. Albaba, C. Gebhardt, O. Hilliges, Marlui: Multi-agent reinforcement learning for adaptive uis (2022). URL: <https://arxiv.org/abs/2209.12660>. doi:[10.48550/ARXIV.2209.12660](https://doi.org/10.48550/ARXIV.2209.12660).
- [24] B. Ghai, Q. V. Liao, Y. Zhang, R. Bellamy, K. Mueller, Explainable active learning (xal): Toward ai explanations as interfaces for machine teachers, *Proc. ACM Hum.-Comput. Interact.* 4 (2021). URL: <https://doi.org/10.1145/3432934>. doi:[10.1145/3432934](https://doi.org/10.1145/3432934).

- [25] A. Rapp, In search for design elements: A new perspective for employing ethnography in human-computer interaction design research, *International Journal of Human-Computer Interaction* 37 (2021) 783–802. doi:10.1080/10447318.2020.1843296.
- [26] K. L. Smart, Contextual inquiry as a method of information design, in: *Content and Complexity*, Routledge, 2014, pp. 217–244.
- [27] K. Niedderer, D. Harrison, J. Gosling, M. Craven, A. Blackler, R. Losada, T. Cid, Working with Experts with Experience: Charting Co-production and Co-design in the Development of HCI-Based Design, *Human-Computer Interaction Series*, Springer International Publishing, Cham, 2020, p. 303–320. URL: http://link.springer.com/10.1007/978-3-030-32835-1_19. doi:10.1007/978-3-030-32835-1_19.
- [28] M. K. Lee, D. Kusbit, A. Kahng, J. T. Kim, X. Yuan, A. Chan, D. See, R. Noothigattu, S. Lee, A. Psomas, A. D. Procaccia, Webuildai: Participatory framework for algorithmic governance, *Proceedings of the ACM on Human-Computer Interaction* 3 (2019) 1–35. doi:10.1145/3359283.
- [29] A. Berge, F. Guribye, S.-L. S. Fotland, G. Fonnes, I. H. Johansen, C. Trattner, Designing for control in nurse-ai collaboration during emergency medical calls, in: *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, ACM, Pittsburgh PA USA, 2023, p. 1339–1352. URL: <https://dl.acm.org/doi/10.1145/3563657.3596110>. doi:10.1145/3563657.3596110.
- [30] C. Bull, H. Aljasim, D. Zytka, Designing opportunistic social matching systems for women’s safety during face-to-face social encounters, in: *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*, ACM, Virtual Event USA, 2021, p. 23–26. URL: <https://dl.acm.org/doi/10.1145/3462204.3481751>. doi:10.1145/3462204.3481751.
- [31] L. Boer, J. Donovan, Provotypes for participatory innovation, in: *Proceedings of the Designing Interactive Systems Conference, DIS ’12*, Association for Computing Machinery, New York, NY, USA, 2012, p. 388–397. URL: <https://doi.org/10.1145/2317956.2318014>. doi:10.1145/2317956.2318014.
- [32] F. Casnati, A. Ianniello, A. Romani, Provocation through narratives: New speculative design tools for human-non-human collaborations, in: F. Zanella, G. Bosoni, E. Di Stefano, G. L. Iannilli, G. Matteucci, R. Messori, R. Trocchianesi (Eds.), *Multidisciplinary Aspects of Design*, Springer Nature Switzerland, Cham, 2024, p. 747–755.
- [33] P. V. K. Andersen, W. S. Mosleh, Conflicts in co-design: engaging with tangible artefacts in multi-stakeholder collaboration, *CoDesign* 17 (2021) 473–492. doi:10.1080/15710882.2020.1740279.
- [34] V. Braun, V. Clarke, Using thematic analysis in psychology, *Qualitative research in psychology* 3 (2006) 77–101.