AUTHOR	Yi Luo, Miriam Tedeschi, Dimitrios Gkouskos, Andrea Resmini
TITLE	Co-Constructing Agency in Human-AI Interactions
YEAR	2024
DOI	10.1145/3677045.3685459
VERSION	Author's accepted manuscript
CITATION	Yi Luo, Miriam Tedeschi, Dimitrios Gkouskos, and Andrea Resmini. 2024. Co-Constructing Agency in Human-AI Interactions. In Adjunct Proceedings of the 2024 Nordic Conference on Human-Computer Interaction (NordiCHI Adjunct 2024), October 13–16, 2024, Uppsala, Sweden. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3677045.3685459
LICENSE	CC BY

Co-Constructing Agency in Human-AI Interactions

Yi Luo
ISDD, Halmstad University

Miriam Tedeschi
Faculty of Law, University of Turku

Dimitrios Gkouskos
ISDD, Halmstad University

Andrea Resmini
ISDD, Halmstad University and GAME, University of Skövde

As AI becomes ever more embedded into daily life and spaces, the discussion of whether AI empowers or disempowers human agency has garnered significant attention. However, AI agency and human agency may not necessarily represent opposing extremes on a continuum. Previous studies have suggested that humans and AI followed different decision-making strategies which lead to results complementary to each other [3, 8, 10]. The aim of this workshop is to explore the role of design in the co-construction of agency in human-AI interactions. By bringing together an interdisciplinary group of researchers in a half-day workshop, we intend to explore design issues and factors that shape agency in collaborative relationships involving humans and AI technologies. Workshop participants will be asked to submit cases where there is evident tension between human and AI agency, and, using Research Through Design, work in exploring issues of human-AI agency in scenarios fueled by the cases submitted to this workshop. As AI technologies move from emergence to being entangled in the everyday, it is increasingly important to thoughtfully design their place in our lives.

1 Introduction

The agency of humans is being challenged by the increasingly pervasive presence of AI and tech in everyday lives and spaces. AI is often a black-box technology that does not readily afford interaction or intervention by human agents. The research community's response is to pursue work related to Human-Centered AI (HCAI) [20, 21] in an attempt to bring AI technologies in line with other technical solutions in terms of human-centeredness. Some scholars caution that it is imperative to examine the influence of AI in society as this technology can shape our view of reality, having a significant impact on our actions, freedom, and agency, even leading to loss of freedom [25].

Understanding where the threshold between human and AI agency lies in current AI-augmented experiences and exploring in what ways design research can support a more human-centered approach to establishing and modulating this threshold is a burgeoning field of research. The traditional model of this threshold is a one-dimensional scale from human control to computer automation [19], which indicates that an increase in human control leads to a decrease in computer automation. Recent studies have blurred this binary view of human and AI agency, where AI systems can be perceived junior colleagues and teammates, augmenting human skill and labor [9, 14]. Shneiderman has suggested a two-dimensional HCAI framework, placing human control and computer automation as x-axis and y-axis respectively, which allows for envisioning scenarios of high human agency and high AI agency [20]. As agency has been recognized as a multi-faceted concept with aspects including but not limited to freedom to self-causality, experience, and materiality [2], we call for contextualizing agency in specific cases to exemplify a multi-dimensional view of agency that paves the path for relevant future research.

In this workshop, we intend to explore the human and AI agency threshold not as an abstract, intangible concept, but rather as a material, collaborative space where agency-in-the-middle is co-produced and can be experimented with. Workshop participants will be accordingly challenged with four overlapping research questions:

- How can the threshold between human and AI agency be identified?
- How is this threshold made visible and material in different socio-spatial settings?
- In what way should this threshold adapt to different contexts, and why?
- Who is responsible for the increase or decrease of agency of either humans or AI?

We will invite participants to bring cases of contested socio-spatial settings where agency is negotiated between human and AI actors. The cases will be shared during the workshop: the participants will present their own views and experience of the threshold between human and AI agency and discuss how they imagine that design can remodulate such agency as being co-constructed rather than pre-constructed, often as a "human vs AI" opposition. Examples of such cases from the organizers' extant research work include: AI-entangled patient and caregiver decision-making in healthcare settings; use of AI in the design and use of vehicles, including autonomous driving; the use of AI in the generation of data doubles for personalization (from social media to serving advertisement). Participants will be encouraged to share experiences in large-scale socio-spatial settings, such as contested algorithmic decision-making interventions leading to segregation, marginalization, and injustice at the city level, all the way to more private, affective, emotional spaces of interaction between the human body and AI, such as AI-augmented personalized healthcare.

Using a Research through Design (RtD) approach, the workshop participants will be guided to identify and formalize selected socio-spatial settings, and experiment with the threshold paying attention to trade-offs, unintended consequences, and possible externalities. What happens if more agency is given to humans rather than AI in each explored context, what kind of agencies and spaces are created as a result. The objective of the workshop is not to identify one single appropriate threshold, a general solution to the problem of human-AI agency, but rather produce a set of scenarios and prototypes that explore the processes through which the threshold could be co-constructed, making it visible and material in a variety of socio-spatial settings. At the end of the workshop, such scenarios and prototypes will be discussed to finalize novel, productive questions and reflections that will allow research to move forward.

1.1 Background

Human agency is posthuman and sociotechnical from the outset [17]. The human body, rather than a unity or fixed, stable identity, is an ongoing assemblage of naturaltechnical [13] in-becoming elements, with different degrees and intensities of agency. Tech and AI elements are thus not simply passive extensions of the human body, but rather they have their own technicity and agency, a force exceeding itself and the initial intentions of their designers [22]. The result of such ongoing assemblage is a complex entanglement of multiple human and non-human entities intra-acting [1] and co-constructing agencies in-the-middle [6], leading to different intensities and degrees of agentic (un)predictability in a variety of socio-spatial settings. This is especially evident with AI products and services.

When different entities—e.g. humans and AI—meet and interact, their encounter produces frictions and ruptures. Such frictions are the seamless negotiations of the agency occurring in-the-middle the human and non-human entities, which, most of the time, remain invisible and unspoken of. At times such negotiations are experienced as effortless (e.g. the human giving up on personal or sensitive data in exchange for online services); other times ruptures make the materiality of the threshold explicit. It may be e.g. a human unsuccessfully trying to enter a website and their double being repeatedly marked as 'robot' or 'bot'. On a broader spatial scale, it may be levels of decision-making where algorithms co-plan contested decisions leading to redlining some city-spaces [18]. In general, the threshold of human and AI agency can be experienced as spatial and material when the negotiation is unsuccessful, or requires additional efforts on the side of the humans.

This is the reason why the specific socio-spatial settings where the negotiation occurs are central in this workshop. Workshop participants will be invited to identify and discuss cases with contested agency between humans and AI to make agency's threshold becomes visible/ material/spatial. Once the threshold is pushed or forced into the range of visibility, the participants will be able to experience and unpack it using RtD, identifying levels of responsibility towards the (un)balanced agency/power between entities, and rehearsing how design might help direct this co-constructed agency.

1.2 Methods

This workshop will use a theme-based discussion followed by an RtD prototyping session to develop artifacts and knowledge to address the agency related questions mentioned earlier. The workshop's methods build on the understanding that ontologies (realities) and epistemologies (how/with what methods we study realities) are tightly co-implicated. Thus, by discussing the cases that participants submit to the workshop, and by actively prototyping design artifacts that target and unpack specific agency-related dimensions, we are also feeding and building up an alternative, tangible, spatio-material, conceptualization of the threshold of humans and AI agency.

1.2.1 Research Through Design and Skewing

Research through design is a research methodology where design is used as the method of knowledge production [24]. Containing a rich tradition of design experimentation in various settings, RtD allows for tentative, speculative and mindful design explorations, including use of RtD to explore (academic) areas of concern [11] making RtD a suitable choice for this workshop. Examples of RtD are diverse, ranging from the design of a radio to experiment with music meta-data as design material [16] to providing a space for students to learn about ethical issues [12], to using speculative watercolor drawings as conversation mediators in design ideation [23]. Documenting RtD is a vital part of the process as design processes can be messy, unpredictable and hard to capture post-fact. Recent work by Desjardins and Kay exemplifies the possibilities in using various materials in capturing not just the outcomes of RtD processes but the process (the "through") itself [7]. We intend to use these insights to ensure that the RtD work in the workshop is documented, enabling future steps. We also bring structure to the RtD workshop, which is imperative due to the limited available time, by applying an ideation method called Skewing.

Skewing artifact properties is an ideation method that uses a framework to re-design existing design artifacts into speculative versions with potentially extreme characteristics, enabling explorations of the outskirts of a design space [15]. Skewing is selected here as an ideation method for two reasons: firstly, the method allows direct use of a framework in ideation, and this will enable the workshop participants to apply dimensions of human-AI agency co-construction to a particular design artifact, and secondly, skewing encourages exploration of a design space, which will enable workshop participants to rehearse potential instances of agency negotiation in their setting of choice. The steps of the Skewing method are described in detail in Lundgren and Gkouskos, 2013, with a short summary presented here: 1. Analysis of an existing design based on the chosen framework; 2. skew by changing one or more properties of the framework; 3. ideate (and prototype) how the artifact is changed due to the skewing. 4. explore further by constructing negative and positive scenarios of use. For the purposes of this workshop, we will adapt the Skewing method by providing the existing design to the workshop participants, and by asking each group to use the dimensions of agency they identified in a previous part of the workshop as their framework for skewing.

2 Workshop Plan

The workshop is a half-day activity, planned to be in-person as we will be working with physical materials. The workshop is organized into three distinct phases, that together enable workshop participants to explicate cases where agency thresholds between AI and humans are evidently negotiated, resulting in visible seams [5].

2.1 Pre-workshop activities

- Call for participation: The organizers will create and distribute a website of call for participation through multiple channels. The strategies of promoting participation can be found in Section 4. The participants will submit a case with (potentially contested) agency between human and AI.
- Workshop participants will be invited to submit cases where AI and human agency are evidently negotiated, elevating issues around agency co-construction. For example, is it the control of a vehicle? Control over one's social media feed? Control over healthcare decisions? What is communicated from each actor (human and AI) and what interaction is afforded, if any?
- Review and synthesize: The organizers will review the submitted cases and select workshop participants. The selected cases will then be analyzed and synthesized to find common themes to be discussed in the workshop.

2.2 During the workshop

1. Introduction: Workshop participants will be introduced to the structure of the workshop, and to each other. Participants will then be divided by the workshop organizers into groups based on shared characteristics of their workshop submissions.
2. Discussion on dimensions of agency: The first activity is to deconstruct agency as it relates to the cases brought by the participants, and identify dimensions of agency for their particular group. One of the outcomes of this phase is a set of dimensions of agency co-construction for each of the workshop groups. These dimensions will be used as input in the RtD phase of the workshop.
3. Research Through Design work: This part of the workshop is informed by the rich tradition of RtD and combines the Skewing artifact properties ideation method (Lundgren & Gkouskos, 2013) and low fidelity prototyping to explore the agency dimensions from the first part of the workshop through the iterative re-design of design artifacts (one for each group) that the workshop organizers will select based on the workshop submissions. Moderators in each work group will ensure that the work, including design decisions is documented for future work and to support the presentation session that follows.
4. Presentations and Discussion: Each group will present the artifacts that they have designed during the RtD session along with a description of agency co-construction that is designed into these artifacts. The workshop will conclude with a summarizing discussion, outlining results and future research directions for agency co-construction between humans and AI.

After the workshop, the organizers will keep a digital copy of all the materials created in the workshop and make it accessible to all participants. The organizers also plan to analyze the materials and publish the findings to potential venues.

2.3 Schedule

Table 1 presents the workshop schedule.

Table 1: Workshop schedule

Time	Activity	Goal
9:00-10:00	Introduction	The organizers introduce the setting of the workshop and participants get familiar with each other.
10:00-11:30	Group work: Research through Design	Participants identify and explore dimensions of agency in related to their cases through prototyping.
11:30-12:00	Presentation and wrap up	The organizers and participants reflect and raise key takeaways from the workshop.

3 Intended Audience

The workshop will be limited to up to 20 participants to ensure adequate facilitation. The organizers plan to invite scholars from different backgrounds, including design, computer science, human geography, humanities, and sociology, to foster an interdisciplinary discussion on agency in the age of AI in different socio-spatial settings. Participants will be selected by the organisers on the basis of quality and originality of the work, as well as the fit with the workshop's topic. Each selected participant will have to prepare and send a case of (contested) agency involving humans and AI that must include the following elements: (1) Description of the socio-spatial settings in which the interaction between humans and AI is taking place; (2) How or if the interaction is producing tensions or frictions; (3) In what ways, if at all, the interaction evolves in time; (4) What kind of impact is the interaction having on the socio-spatial settings.

4 Promotional Strategy

The organizers will actively attract and recruit participants through different channels. Research mailing lists will be used, social media channels, as well as the research and professional networks of each organizer. A website for the workshop will be created, where the organizers will share the call for participation; updated schedule of the workshop; description of the activities that will be carried out during the workshop. Participants will be updated via email of important deadlines and eventual changes.

5 Organizer Bios

Yi Luo is a Ph.D. student in informatics and experience design at Halmstad University interested in designing technology to enhance people's everyday life. Her background is a mixture of human-computer interaction and psychology with working experience in designing IoT devices with multimodal interaction. Her thesis aims at how to design AI/ML that supports human-AI collaboration in the experience design process.

Miriam Tedeschi is a researcher at the Faculty of Law, University of Turku, Finland, and a fellow at the Westminster Law & Theory Lab, London. A docent in human geography, Miriam's work combines empirical methods, primarily ethnographic and autoethnographic approaches, with non-representational geographies and posthuman theories. She currently serves as the Principal Investigator for the Academy of Finland research project JuDiCe – Justice in Digital Space.

Dimitrios Gkouskos is a senior lecturer in the department of Intelligent Systems and Digital Design at Halmstad University, where he works with design research on topics ranging from use of artificial intelligence in design processes, to social sustainability and research through design. Dimitrios' research interests center around research through design methodologies, interaction design education, and using design to include people in the shaping of emergent technologies.

Andrea Resmini is associate professor of experience design and information architecture in the Department of Intelligent Systems and Digital Design at Halmstad University, where he works with the conceptualization and design of blended spaces and blended experiences, and a researcher in exploratory and speculative game making at the GAME Research Lab, University of Skövde. An architect turned information architect turned educator, Andrea is a two-time past president of the Information Architecture Institute, a founding member of Architecta, the Italian Society for Information Architecture, a co-founder of the World IA Day international conference, the editor-in-chief of the Journal of Information Architecture, and the author of Pervasive Information Architecture (2011), Reframing Information Architecture (2014), and Advances in Information Architecture (2021).

ACKNOWLEDGEMENTS

REFERENCES

[1] Karen Barad. 2011. Nature's Queer Performativity. *Qui Parle* 19, 2 (2011), 121–158. https://doi.org/10.5250/quiparle.19.2.0121

[2] Dan Bennett, Oussama Metatla, Anne Roudaut, and Elisa D. Mekler. 2023. How does HCI Understand Human Agency and Autonomy? In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, April 19, 2023. ACM, Hamburg Germany, 1–18. https://doi.org/10.1145/3544548.3580651

[3] Nigel Bosch and Sidney K. D'Mello. 2022. Can Computers Outperform Humans in Detecting User Zone-Outs? Implications for Intelligent Interfaces. *ACM Trans. Comput.-Hum. Interact.* 29, 2 (April 2022), 1–33. https://doi.org/10.1145/3481889

[4] Matthew Chalmers and Areti Galani. 2004. Seamful interweaving: heterogeneity in the theory and design of interactive systems. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques* (*DIS '04*), August 01, 2004. Association for Computing Machinery, New York, NY, USA, 243–252. https://doi.org/10.1145/1013115.1013149

[5] Augusto Cusinato and Andreas Philippopoulos-Mihalopoulos (Eds.). 2016. *Knowledge-creating Milieus in Europe: Firms, Cities, Territories*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-45173-7

[6] Audrey Desjardins, Heidi R. Biggs, Cayla Key, and Jeremy E. Viny. 2020. IoT Data in the Home: Observing Entanglements and Drawing New Encounters. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (*CHI '20*), April 23, 2020. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376342

[7] Mingming Fan, Ke Wu, Jian Zhao, Yue Li, Winter Wei, and Khai N. Truong. 2020. VisTA: Integrating Machine Intelligence with Visualization to Support the Investigation of Think-Aloud Sessions. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (January 2020), 343–352. https://doi.org/10.1109/TVCG.2019.2934797

[8] Mingming Fan, Xianyou Yang, TszTung Yu, Q. Vera Liao, and Jian Zhao. 2022. Human-AI Collaboration for UX Evaluation: Effects of Explanation and Synchronization. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1 (March 2022), 1–32. https://doi.org/10.1145/3512943

[9] Jessica L. Feuston and Jed R. Brubaker. 2021. Putting Tools in Their Place: The Role of Time and Perspective in Human-AI Collaboration for Qualitative Analysis. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (October 2021), 1–25. https://doi.org/10.1145/3479856

[10] Elisa Giaccardi. 2019. Histories and futures of research through design: from Prototypes to Connected things. *International Journal of Design* 13, 3 (2019), 139–155.

[11] Dimitrios Gkouskos, Andrea Resmini, Charlotte Arghavan Shahlaei, and Jonathan Burgos. 2023. Teaching Design Ethics Through Tabletop Game Prototyping. 2023. . Retrieved April 18, 2024 from https://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-52899

[12] Donna Haraway. 1990. *Simians, Cyborgs, and Women: The Reinvention of Nature*. Routledge, New York. https://doi.org/10.4324/9780203873106

[13] Emily Kuang, Ehsan Jahangirzadeh Soure, Mingming Fan, Jian Zhao, and Kristen Shinohara. 2023. Collaboration with Conversational AI Assistants for UX Evaluation: Questions and How to Ask them (Voice vs. Text). In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, April 19, 2023. ACM, Hamburg Germany, 1–15. https://doi.org/10.1145/3544548.3581247

[14] Sus Lundgren and Dimitrios Gkouskos. 2013. Escaping the obvious: Skewing properties of interaction. *Nordes Conference Series* (June 2013). Retrieved from https://dl.designresearchsociety.org/nordes/nordes2013/researchpapers/2

[15] William Odom and Tijs Duel. 2018. On the Design of OLO Radio: Investigating Metadata as a Design Material. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (*CHI '18*), April 19, 2018. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3173574.3173678

<bib id="bib16"><number>[16]</number>Gillian Rose. 2017. Posthuman Agency in the Digitally Mediated City: Exteriorization, Individuation, Reinvention. *Annals of the American Association of Geographers* 107, 4 (July 2017), 779–793. https://doi.org/10.1080/24694452.2016.1270195</bib>

<bib id="bib17"><number>[17]</number>Sara Safransky. 2020. Geographies of Algorithmic Violence: Redlining the Smart City. *International Journal of Urban and Regional Research* 44, 2 (2020), 200–218. https://doi.org/10.1111/1468-2427.12833</bib>

<bib id="bib18"><number>[18]</number>T. B. Sheridan, W. L. Verplank, and T. L. Brooks. 1978. Human/computer control of undersea teleoperators. November 01, 1978. . Retrieved April 18, 2024 from https://ntrs.nasa.gov/citations/19790007441</bib>

<bib id="bib19"><number>[19]</number>Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (April 2020), 495–504. https://doi.org/10.1080/10447318.2020.1741118</bib>

<bib id="bib20"><number>[20]</number>Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2020. From Human-Human Collaboration to Human-AI Collaboration: Designing AI Systems That Can Work Together with People. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, April 25, 2020. ACM, Honolulu HI USA, 1–6. https://doi.org/10.1145/3334480.3381069</bib>

<bib id="bib21"><number>[21]</number>Mark A Wood. 2021. Rethinking how Technologies Harm. *The British Journal of Criminology* 61, 3 (May 2021), 627–647. https://doi.org/10.1093/bjc/azaa074</bib>

<bib id="bib22"><number>[22]</number>Paulina Yurman. 2021. Fluid Speculations: Drawing Artefacts in Watercolour as Experimentation in Research Through Design. In *Proceedings of the 13th Conference on Creativity and Cognition* (C&C '21), June 22, 2021. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3450741.3466777</bib>

<bib id="bib23"><number>[23]</number>John Zimmerman and Jodi Forlizzi. 2014. Research Through Design in HCI. In *Ways of Knowing in HCI*, Judith S. Olson and Wendy A. Kellogg (eds.). Springer New York, New York, NY, 167–189. https://doi.org/10.1007/978-1-4939-0378-8_8</bib>

<bib id="bib24"><number>[24]</number>Shoshana Zuboff. 2023. The Age of Surveillance Capitalism. In *Social Theory Re-Wired* (3rd ed.). Routledge.</bib>