Participatory Design of a Machine Learning Driven Visualization System for Non-Technical Stakeholders

Jesse Josua Benjamin
Human-Centered Computing
Freie Universität Berlin
Berlin, Germany
jesse.benjamin@fu-berlin.de

Christoph Kinkeldey
Human-Centered Computing
Freie Universität Berlin
Berlin, Germany
christoph.kinkeldey@fu-berlin.de

Claudia Müller-Birn
Human-Centered Computing
Freie Universität Berlin
Berlin, Germany
clmb@inf.fu-berlin.de

CCS CONCEPTS

• **Human-centered computing → Participatory design**; *HCI theory, concepts and models*; Visualization theory, concepts and paradigms.

1 INTRODUCTION

Promoting participation by non-technical stakeholders in the design and development of socio-technical systems is increasingly recognised as a necessity, particularly in what has been called the "third-wave" human-computer interaction (HCI) research [9]. However, with regards to the increasing implementation of machine learning (ML) algorithms in everyday systems, design researchers are faced with multiple challenges. One is that ML technologies require distinct, and as yet unelaborated, ways of prototyping to reflect the dynamics of "statistical intelligence" [7] at work in ML. Another challenge presents itself as a matter of methodological pacing: researching abstract issues such as opacity, interpretability or fairness is complicated with regards to data-driven technologies; where exact effects are hard to apprehend before the technologies are already implemented. A third is that, while ML and HCI research in fields such as Explainable AI and Interpretability has generated many ways of representing ML outputs (e.g. [8, 10, 17]), the methods proposed are almost exclusively oriented at stakeholders with formal ML education [16], or at designers using ML in their practice (e.g. [1]). In summary, a substantial gap exists regarding how, in the development of an actual ML-driven system, non-technical stakeholders may be enabled to articulate interpretability needs through design research methods.

In this workshop contribution, we present our participatory interpretability methodology, which we generated while developing an ML-driven visualization system over three years at a natural history research institution. Our goal was to understand the specific ways in which particular non-technical stakeholders interpret the representation of institution data (in our case, research projects) by ML, and thereby learn how we may support a multitude of possible interpretations.

MuC'20 Workshops, Magdeburg, Deutschland

https://doi.org/10.18420/muc2020-ws109-287

2 DEVELOPING OUR PARTICIPATORY INTERPRETABILITY METHODOLOGY

The methodology we will present at the workshop took shape over the course of a three-year research project in which we cooperated with a major German research institution, the Museum für Naturkunde in Berlin[1]. The goal of the research project was to find a way to foster knowledge exchange and collaboration between institution employees by means of data visualization of research activities. For the institution, this was an ongoing challenge due to a multiplicity of factors; for instance the haphazard distribution of facilities throughout various buildings, the lack of communal spaces, and the absence of communication measures for the large and diverse staff (approximately 250 permanent members of staff with at least 400 temporary members such as students and visiting fellows) across 4 major research departments. Additionally, not all stakeholders at the institution are directly involved in research activities in the main domain of natural science–for example, activities such as Citizen Science initiatives, lectures and field tours. Therefore, the visualization of research activities in our use case would have to be oriented not only at one specific group of stakeholders and their way of doing things, but rather diverse communities-of-practice [19] that envelope various socio-material practices [11].

2.1 Preparatory Research into Stakeholder Perspectives

As initial steps in the development of our methodology, we conducted *participatory data modelling workshops* and *semi-structured interviews*. The former were conceptualized to understand what stakeholders saw as the 'formal' prerequisites for what constitutes research activities at the institution, whereas the latter was deemed necessary for a more foundational understanding of the actual lived experience of various stakeholders with research activities.

2.1.1 Participatory Data Modelling. In order to gain an initial insight into the context and the most relevant types of data for research activities, we conducted participatory data modelling workshops with institution employees from various organizational positions in three iterations. This method allowed us to discern what specific 'formal' attributes exist for research activities in the mind of stakeholders. By way of successive iterations, we could integrate multiple stakeholders in defining what data model should drive our data visualization. For each workshop, we provided paper prototype for a data entry interface. By simulating entering data for an actual research activity, groups of stakeholders were able to review and

[1]https://www.museumfuernaturkunde.berlin/en, accessed 06/05/2020.

critique the chosen categories for the interface, and by extension, the hypothetical data model underlying those categories.

2.1.2 Semi-Structured Interviews. To discern the actual lived experience of being implicitly or explicitly involved with research activities, we subsequently carried out semi-structured interviews with stakeholders from a variety of organizational departments and positions. Based on an analysis of these interviews, we could form hypotheses on how research activities integrated into the day-to-day practices of employees at the institution, and, more importantly, what kind of visualizations may promote discovering potentials for knowledge exchange and collaboration.[2] Among our main observations, we found that the hierarchical organization was seen as a hindrance to knowledge exchange, with stakeholders pointing out that research groups rarely know about activities within their own department, let alone those of others. Therefore, while the hierarchy was identified as a key attribute for research activities (i.e., what department a research project 'belongs' to) in the prior workshops, the findings from our interviews implied that we should find different ways to visualize the activities.

2.2 Participatory Interpretability Design Workshop

Based on our findings, we hypothesized that an ML-driven visualization system would be most promising for the use case. By connecting research activities based on their thematic similarities as discerned via Natural Language Processing (NLP), we developed a scatter-plot cluster-visualization of research projects in which the proximity and grouping of points would indicate thematic similarity. By foregoing the hierarchy as a visual and semantic structure for our visualization, we sought to offer an alternative view on the research institution that the multiplicity of stakeholders may interpret given their own presuppositions. However, as indicated above, promoting this exact dynamic is an open challenge in HCI and ML research, which are predominantly expert-oriented [13, 16] and often studied in crowd-working settings (e.g., [5]), which do not map to the particularities of real life socio-material contexts. Building on recent discussions regarding its applicability in the development of ML-driven systems (e.g., [4, 14, 15]), we therefore looked at participatory design research for inspiration in developing a method for understanding interpretability for non-technical stakeholders in our specific use case context.

In developing our participatory design method, which forms the main part of our contribution to the workshop, we asked ourselves: how can we discern and support the specific ways in which non-technical stakeholders interpret ML outputs? Reviewing recent participatory design research in the area of Internet of Things [3, 6], we wanted to make sure that design artefacts for our participatory method would represent the actual NLP pipeline we developed rather than abstract and potentially misleading proxies. We therefore firstly searched the technical literature for 'interpretability techniques', which are algorithmic methods which extract information from ML pipeline steps [12]. Reflecting on related work in design research, specifically Vallgåda and Redström's "computational composites" [18], we transformed interpretability techniques suitable for our NLP pipeline into tangible artefacts: transparencies which were printed with visual indicators of specific steps (e.g., indicating the most important words for a specific cluster), and could be overlayed onto a print-out of our ML-driven cluster visualization.

We chose to deploy the artefacts in a participatory design workshop with non-technical stakeholders from varying institutional positions and fields of expertise. We recruited six participants (2 female, 4 male, self-reported) which formed two groups. In the first part of the workshop, participants would interact with the designed transparencies to solve typical visual analytics tasks in order to familiarize themselves with the material. Subsequently, we asked participants to emulate the NLP pipeline by using transparencies as they saw fit to place two new projects, i.e. not represented in the cluster-visualization, within the cluster-visualization. In the second part of the workshop, participants were given prototyping material kits and were asked to represent research activities at the institution in a 'self-explanatory' manner.

We found that the use of physical artefacts and the interplay between the two workshop parts led to significant insights into what constitutes interpretability in our use case. The interplay is particularly noteworthy: in the second part, we came to understand explicit presuppositions for interpretation at work within the organizational context, which we could then relate back to the actual interactions with interpretability techniques in the first part. Among our main findings, we found that participants with diverse institutional backgrounds could use a technique visualizing uncertainty of cluster assignments to great effect, reflecting on how technological decision-making maps to the 'natural' assumptions of socio-material relations at the institution. We furthermore found out exactly which of the latter could potentially enrich interpretations of the ML-driven visualization, prompting us to adapt our visualization design. The insights from the workshop have directly informed the prototype of our ML-driven visualization system. Therefore, we argue that this approach is highly promising for future work in ML interpretability for non-technical stakeholders, and suggest that presenting our work at the workshop will lead to promising research directions.

3 CONTRIBUTION TO THE WORKSHOP

Due to our experience of pursuing participation in the development and design of an ML-driven visualization system over a long period of time, we are particularly eager to discuss how our approach could be transferred to other contexts. Furthermore, we aim to reflect on the degree of participation we were able to promote throughout the diverse and longitudinal deployment of research methods. Lastly, we will also discuss how our approach to ML interpretability for non-technical stakeholders could also be used in participatory design of likewise emerging technologies, such as Internet of Things.

ACKNOWLEDGMENTS

This work is supported by the German Federal Ministry of Education and Research, grant 03IO1633 ("IKON – Wissenstransferkonzept für Forschungsinhalte, -methoden und -kompetenzen in Forschungsmuseen").

[2] We published all findings in a technical report [2].

REFERENCES

[1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 3:1–3:13. https://doi.org/10.1145/3290605.3300233 event-place: Glasgow, Scotland Uk.

[2] Jesse Josua Benjamin, Claudia Müller-Birn, and Christoph Kinkeldey. 2019. *Understanding Knowledge Transfer Activities at a Research Institution through Semi-Structured Interviews.* Technical Report TR-B-19-02. Freie Universität Berlin, Berlin. https://refubium.fu-berlin.de/handle/fub188/25265?locale-attribute=en

[3] Arne Berger, William Odom, Michael Storz, Andreas Bischof, Albrecht Kurze, and Eva Hornecker. 2019. The Inflatable Cat: Idiosyncratic Ideation of Smart Objects for the Home. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, Glasgow, Scotland Uk, 1–12. https://doi.org/10.1145/3290605.3300631

[4] Tone Bratteteig and Guri Verne. 2018. Does AI Make PD Obsolete?: Exploring Challenges from Artificial Intelligence to Participatory Design. In *Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2 (PDC '18)*. ACM, New York, NY, USA, 8:1–8:5. https://doi.org/10.1145/3210604.3210646 event-place: Hasselt and Genk, Belgium.

[5] Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O'Connell, Terrance Gray, F. Maxwell Harper, and Haiyi Zhu. 2019. Explaining Decision-Making Algorithms Through UI: Strategies to Help Non-Expert Stakeholders. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 559:1–559:12. https://doi.org/10.1145/3290605.3300789 event-place: Glasgow, Scotland Uk.

[6] Audrey Desjardins, Cayla Key, Heidi R. Biggs, and Kelsey Aschenbeck. 2019. Bespoke Booklets: A Method for Situated Co-Speculation. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19)*. ACM, New York, NY, USA, 697–709. https://doi.org/10.1145/3322276.3322311 event-place: San Diego, CA, USA.

[7] Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. UX Design Innovation: Challenges for Working with Machine Learning As a Design Material. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 278–288. https://doi.org/10.1145/3025453.3025739 event-place: Denver, Colorado, USA.

[8] David Gunning. 2016. Broad Agency Announcement : Explainable Artificial Intelligence (XAI). https://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf

[9] Steve Harrison, Phoebe Sengers, and Deborah Tatar. 2011. Making epistemological trouble: Third-paradigm HCI as successor science. *Interacting with Computers* 23, 5 (Sept. 2011), 385–392. https://doi.org/10.1016/j.intcom.2011.03.005

[10] Fred Hohman, Andrew Head, Rich Caruana, Robert DeLine, and Steven M. Drucker. 2019. Gamut: A Design Probe to Understand How Data Scientists Understand Machine Learning Models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 579:1–579:13. https://doi.org/10.1145/3290605.3300809 event-place: Glasgow, Scotland Uk.

[11] Paul M. Leonardi. 2012. *Materiality, Sociomateriality, and Socio-Technical Systems: What Do These Terms Mean? How are They Related? Do We Need Them?* SSRN Scholarly Paper ID 2129878. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=2129878

[12] Zachary C. Lipton. 2016. The Mythos of Model Interpretability. *arXiv:1606.03490 [cs, stat]* (June 2016). http://arxiv.org/abs/1606.03490 arXiv: 1606.03490.

[13] Zachary C. Lipton and Jacob Steinhardt. 2018. Troubling Trends in Machine Learning Scholarship. *arXiv:1807.03341 [cs, stat]* (July 2018). http://arxiv.org/abs/1807.03341 arXiv: 1807.03341.

[14] Daria Loi, Thomas Lodato, Christine T. Wolf, Raphael Arar, and Jeanette Blomberg. 2018. PD Manifesto for AI Futures. In *Proceedings of the 15th Participatory Design Conference: Short Papers, Situated Actions, Workshops and Tutorial - Volume 2 (PDC '18)*. ACM, New York, NY, USA, 48:1–48:4. https://doi.org/10.1145/3210604.3210614

[15] Daria Loi, Christine T. Wolf, Jeanette L. Blomberg, Raphael Arar, and Margot Brereton. 2019. Co-designing AI Futures: Integrating AI Ethics, Social Computing, and Design. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion (DIS '19 Companion)*. ACM, New York, NY, USA, 381–384. https://doi.org/10.1145/3301019.3320000 event-place: San Diego, CA, USA.

[16] Tim Miller, Piers Howe, and Liz Sonenberg. 2017. Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences. *arXiv:1712.00547 [cs]* (Dec. 2017). http://arxiv.org/abs/1712.00547 arXiv: 1712.00547.

[17] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. 2018. A Survey of Evaluation Methods and Measures for Interpretable Machine Learning. *arXiv:1811.11839 [cs]* (Nov. 2018). http://arxiv.org/abs/1811.11839 arXiv: 1811.11839.

[18] Anna Vallgåda and Johan Redström. 2007. Computational Composites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 513–522. https://doi.org/10.1145/1240624.1240706 event-place: San Jose, California, USA.

[19] Etienne Wenger. 1999. *Communities of practice: Learning, meaning, and identity.* Cambridge university press.