1st Workshop on Data Craft as Boundary AI Practice

1st Workshop on Data Craft as Boundary AI Practice

Maria Luce Lupetti*
Politecnico di Torino
Turin, Italy
maria.lupetti@polito.it

Alessandro Ianniello
Delft University of Technology
Delft, Netherlands
a.ianniello@tudelft.nl

Willem van der Maden
IT University of Copenhagen
Copenhagen, Denmark
wiva@itu.dk

Juri Sanni
Politecnico di Torino
Turin, Italy
juri.sanni@polito.it

Eric Tron Gianet
Politecnico di Torino
Turin, Italy
erictrngnt@gmail.com

Minoru Matsui
Advanced Institute of Industrial Technology Tokyo
Tokyo, Japan
xerroxcopy@gmail.com

Dave Murray-Rust
Delft University of Technology
Delft, Netherlands
d.s.murray-rust@tudelft.nl

Abstract

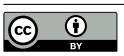
Data is central to AI performance, yet its curation remains an invisible and labour-intensive process, often leading to biases and reliability issues. While HCI has explored methods to improve data work, a growing body of research embraces data imperfections in an artistic vein to provoke reflection on human-AI relations. This workshop examines "Data Craft"—practices that creatively manipulate data to challenge conventional AI narratives and lower barriers to public engagement. By framing data craft as a boundary practice, we explore its potential to foster dialogue on AI capabilities, limitations, and societal impact. The workshop will investigate how data craft can be systematically integrated into participatory AI efforts, moving beyond artistic spaces to inform public debate. Through hands-on exploration, we aim to uncover strategies for leveraging data craft to engage diverse communities in shaping AI discourse.

Keywords

Boundary AI Objects, Data Work, Data Craft, Participatory AI, Human-AI Interaction

ACM Reference Format:
Maria Luce Lupetti, Alessandro Ianniello, Willem van der Maden, Juri Sanni, Eric Tron Gianet, Minoru Matsui, and Dave Murray-Rust. 2025. 1st Workshop on Data Craft as Boundary AI Practice. In *CHItaly 2025: 16th Biannual Conference of the Italian SIGCHI Chapter (CHItaly 2025), October 06–10, 2025, Salerno, Italy.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3750069.3757880

*Corresponding author: Maria Luce Lupetti, maria.lupetti@polito.it



CHItaly 2025, Salerno, Italy

ACM ISBN 979-8-4007-2102-1/25/10
https://doi.org/10.1145/3750069.3757880

1 Introduction

Data is a critical component of the good functioning of Artificial Intelligence (AI) algorithms and related applications. As the popular saying *"Garbage in, garbage out"* underscores, if the quality of data we train AI with is low, the quality of AI behaviour and output will be poor [11], and the coverage of the dataset shapes the behaviour of the model. However, despite its name[1] it is important to acknowledge that data is not a given resource. It is rather a "raw material" that needs human labor to be extracted, curated, and maintained, all actions that together constitute the complex—yet often invisible—sociotechnical practice space of *Data Work* [3, 21].

Choices made at each step of data work pipelines have downstream effects on the behaviour of an AI model and on the generalizability and reliability of the results [4, 7, 19]—what Sambasivan et al. [23] call *Data Cascades*. Understanding and mitigating data cascades is critical, especially when it comes to AI being applied in sensitive settings and domains, such as health care, hiring, or driving safety, where consequences can be serious. For instance, research has found that algorithms used to identify candidates for high-risk care management programs routinely failed to refer people from racial minorities for these beneficial services [13]. Relatedly, the recent Dutch childcare benefit scandal put the spotlight on how AI algorithms employed by public administration often make use of historical data and nationality data to make an assessment about the probability of citizens to commit fraud, often leading to concrete discriminatory actions towards non-Dutch citizens [15]. This is a problem further exacerbated by the fact that data work remains extremely time-consuming and expensive [19, 23], leading to the use of datasets often lacking diversity and poor annotation quality.

As creating appropriate datasets and accurate data work pipelines becomes more and more important, the field of human-computer interaction (HCI) increasingly engages with investigations of how best to support these practices [19]. HCI strategies vary from the development of dedicated interfaces, measurement mechanisms,

[1] Data comes for *dare*, that which is given; alternative framings as *capta* more precisely highlight the selective nature of data, see e.g. [9].

and incentives, to advocating for and supporting the participation of broader communities, especially the fragile and vulnerable ones [23], in the production, curation, and annotation of datasets [19]. An example is PRISM [17], a dataset that collects preferences and feedback of 1500 diverse participants from 75 countries on 8,011 live conversations with 21 Large Language Models (LLMs) and aims to make manifest the complexity of aligning LLMs to human preferences when the human is not an individual but a plurality [17].

In stark contrast, a recent—smaller—stream of HCI research has been looking at and engaging with data work from a radically different standpoint. Here, data cascades are embraced and latent spaces are indulged in, with the explicit intention of engaging in reflections about our relationships with AI algorithms and data more specifically. For instance, Arzberger et al. [1] underscores the possible role of data curation as a reflexive practice in which the limits of algorithms and the biases inherited from poor datasets can become a resource for auto-confrontation. In this vein, in *Objective Portrait* [27] three designers annotated a collection of images with subjective labels, later used to teach an object detection model the designers' perspectives. These trained models were then used to *confront* their own design work as a way to prompt self-reflection. Similarly, Hoggenmueller et al. [14] leveraged text-to-image models in their "CreativeAI Postcards" method, deliberately embracing AI limitations to surface robot stereotypes and prompt critical reflection on designers' assumptions about robot appearances. Further, Turtle et al. [26] make manifest a critique of and encourage an engagement with the theme of human datafication. In particular, through the exploration titled *Undoing Gracia*, a Digital Twin simulation where the first author interacts with their AI twins within a fictional world, the authors underscore how humans are becoming encoded as data subjects whose existence is entangled with both the digital and physical realms. The theme of human-AI entangled experience of the world becomes manifest also through the work of Karmann [16] who designed *Paragraphica*, a camera that uses location data to prompt AI to generate a "photo" of a specific place and moment. Instead of centering on the self, however, here the work focuses on data as a means to "see through algorithmic eyes" and reflect on when data matters. As a matter of fact, we often uniquely think of it as something exclusively related to AI training, yet AI inference relies on data as well. This is even more evident in *Shadowplay* [20], an installation where visitors can use their bodies to create AI imagery in real-time. Here, the author puts the spotlight on the craft practices of prompting and invites a reflection on what we consider data (pixels in this case) and how this affects our interaction with AI. To some extent, most of these examples fit within the umbrella of Experiential AI—the use of creative practice to make the way algorithms work vivid and apparent [12].

Because of their radically different standpoint and ambitions, these exemplars of data work are better defined as *Data Craft* practices: approaches where authors express "freedom in curating the data, avoiding standardised measures of success and tweaking conventional steps of the process in order to get the desired outcome" [6]. The desired outcome here is critical: instead of attempting to *fix* datasets and prevent data cascades, the ambition of these works is to rather encourage *reflections on* and *reconceptualisations of* our entanglement with AI algorithms and the data these work with. In doing so, data craft lowers the barriers for the public to grasp and engage with matters of algorithmic capabilities and limitations, as well as hopes and fears we embed into these technologies. This stream of work relates to a large history of more-or-less critical data practices, from human data interaction studies [22], through the *ablative framework* for designing *from, with and by* data [25] to explicit explorations of feminist approaches to data (e.g. [8, 24]) and explicit use of curated data to expose the potential for bias in generative models [10].

In these reflexive and dialogic ambitions, we see a great potential that data craft can have to function as a *boundary practice*: "a practice whose purpose is to connect other practices by addressing conflicts, reconciling perspectives, and resolving differences" [28]. Data craft combines knowledge from diverse disciplinary domains and looks at the troubles that arise from the engagement with data, to seek either personal or collective confrontation. In doing so, these works help delineate borders, which are not to be seen as "markers of difference" but rather as "interfaces facilitating knowledge production" [18] about AI and data. This holds great potential, especially within the space of participatory AI research and practice, in which there is a growing call for both approaches to nurture public AI literacy, as well as to develop novel models of engagement [2, 5].

However, up to today, these data craft practices seem to remain bound to the space of designerly and artistic exploration, without explicit employment within the space of public debate and consultation about AI. Although many of these take the form of installations, are exhibited in galleries and festivals, and are experienced by multitudes of people, there is no track of the effects on the public perception of AI and matters of data. *Shadowplay* [20], for instance, has been experienced by thousands of people across Europe; Elwes' *Zizi Project* has been shown at SFMOMA and the V&A. But, *what did they learn? Did this interaction change the way these people think about AI and possible interactions with it?* And more broadly, *can data craft practices be systematically employed in participatory AI activities?*

1.1 The workshop

The workshop investigates the potential and challenges of data craft to function as a boundary practice. More specifically, through the creation of a collection of works, the workshop will unpack the venues, work pipelines, and modes of engagement to better understand data craft and its potential to enable debate about important topics surrounding AI and data. The goal is to create a dialogue about these practices and to understand whether and how data craft can bring a distinct value to the fields of HCI and AI. The workshop provides the community with a *collection of data craft exemplars*, and a series of critical reflections on the *alternative ways to practice data craft*, with the ultimate intent to set an *HCI research agenda around data craft as a boundary practice for participatory AI*. The program will combine lightning talks from two invited speakers, short presentations from workshop participants and a hands-on activity. The invited speakers are selected to bring complementary perspectives in the space of *Data Craft* practices, i.e., Joseph Lindley for his experience with AI media art and wide engagement with the public, and Cristina Zaga, for her critical and systematic inquiry of Data Work and her investigations for integrating alternative data practices in research and design work. In particular, the keynote speakers will provide two radically different views into data craft work, which carry complementary values.

Joseph Lindley, in his talk titled "*A Renaissance of Craft? GenAI, Data and Vibes*", will draw parallels between the Arts and Crafts movement, that emerged as artisans offered a counterpoint to the reduced quality and loss of traditional skills that were brought about by industrial mass production, and our current moment in history, where the mass production of GenAI has democratised our relationship with data, the defining creative and economic material of the 21st century. As Lindley argues: "almost overnight, our relationship with data has transformed, becoming something anyone can massage, manipulate, make beautiful, harness, control, and vibe with. GenAI is precipitating new forms of creativity, a rapid development in culture, and we are living through a new 'Golden Age' of Data Craft. At first glance, this is a wonderful place to be, with democratised creative potential unleashed as never before. But, *if we cast an eye towards the Arts and Crafts movement's ultimate decline, are there cautionary tales we should be mindful to consider?*. This cautionary note is interwoven with the counter-perspective on Data Craft discussed by Cristina Zaga. In her talk titled "*Reframing Reclaming and Re-Imagining Data in the Post-Human Convergence*" will provide an overview of how data is culturally and politically shaped in both knowledge and technological production. She will provide examples that demonstrate how we, as designers working alongside both humans and non-humans, can reframe our approach to crafting data. By employing critical design, we can re-appropriate and re-imagine data to align it more closely with lived and sensory experiences. I will discuss how embodied and political sensemaking can help us explore new avenues for granular, slow, and situated data that support the shift of power, promote emancipation, and help individuals regain agency in healthcare and work settings. She will ultimately stress how data serves as a crucial form of social, cultural, political, and economic power, carrying socio-technical implications that necessitate critical and intentional consideration in HCI and design.

The short talks from selected participants and organizers will showcase examples of data practices that will subsequently be used in the hands-on activity. For instance, Willem wan der Maden will discuss how the project *Deviation Game*, by Tomo Kihara, includes an element of data craft that translate into an experience where participants play against an AI. Juri Sanni will discuss *Italian Brainrots* as an exemplar of internet phenomenon that is deeply intertwined with dataset and genAI appropriation from the general public. Roberta Antognini will speculate on how *forgotten or dismissed personal data* could be used to dialogue with health AI technologies, such as self-management tools, as a way to acknowledge and listen to subjective narratives and ultimately make medical care more respectful, transparent, and useful. More cases will be added.

Acknowledgments

Funded by Italian Ministry of University and Research, FIS 2 funding program (Project reference PARJAI – FIS-2023-00349).

References

[1] Anne Arzberger, Maria Luce Lupetti, and Elisa Giaccardi. 2024. Reflexive data curation: Opportunities and challenges for embracing uncertainty in human–AI collaboration. *ACM Transactions on Computer-Human Interaction* 31, 6 (2024), 1–33.

[2] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the people? Opportunities and challenges for participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–8.

[3] Kevin Brine and Mary Poovey. 2013. *From Measuring Desire to Quantifying Expectations: A Late Nineteenth-Century Effort to Marry Economic Theory and Data*. 61–76. doi:10.7551/mitpress/9302.003.0005

[4] Lukas Budach, Moritz Feuerpfeil, Nina Ihde, Andrea Nathansen, Nele Noack, Hendrik Patzlaff, Felix Naumann, and Hazar Harmouch. 2022. The effects of data quality on machine learning performance. *arXiv preprint arXiv:2207.14529* (2022).

[5] Tiago Sérgio Cabral. 2025. AI Literacy Under the AI Act: An Assessment of its Scope. *EU Law Analysis (https://eulawanalysis. blogspot. com/)* (2025).

[6] Baptiste Caramiaux and Sarah Fdili Alaoui. 2022. " Explorers of Unknown Planets" Practices and Politics of Artificial Intelligence in Visual Arts. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–24.

[7] Kate Crawford. 2021. *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

[8] Catherine D'ignazio and Lauren F Klein. 2023. *Data feminism*. MIT press.

[9] Martin Dodge and Rob Kitchin. 2005. Codes of Life: Identification Codes and the Machine-Readable World. *Environment and Planning D: Society and Space* 23, 6 (Dec. 2005), 851–881. doi:10.1068/d378t Publisher: SAGE Publications Ltd STM.

[10] Jake Elwes. 2019. Zizi. https://www.jakeelwes.com/project-zizi-2019.html

[11] Brooks Hanson, Shelley Stall, Joel Cutcher-Gershenfeld, Kristina Vrouwenvelder, Christopher Wirz, Yuhan Rao, and Ge Peng. 2023. Garbage in, garbage out: mitigating risks and maximizing benefits of AI in research. *Nature* 623, 7985 (2023), 28–31.

[12] Drew Hemment, Dave Murray-Rust, Vaishak Belle, Ruth Aylett, Matjaz Vidmar, and Frank Broz. 2024. Experiential AI: Between Arts and Explainable AI. *Leonardo* (April 2024), 298–306. doi:10.1162/leon_a_02524

[13] Sharona Hoffman and Andy Podgurski. 2019. Artificial intelligence and discrimination in health care. *Yale J. Health Pol'y L. & Ethics* 19 (2019), 1.

[14] Marius Hoggenmueller, Maria Luce Lupetti, Willem Van Der Maden, and Kazjon Grace. 2023. Creative AI for HRI design explorations. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 40–50.

[15] Amnesty International. 2021. Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal. (2021). https://www.amnesty.org/en/documents/eur35/4686/2021/en/

[16] Karman, B. 2023. Paragraphica. https://bjoernkarmann.dk/project/paragraphica

[17] Hannah Rose Kirk, Alexander Whitefield, Paul Rottger, Andrew M Bean, Katerina Margatina, Rafael Mosquera-Gomez, Juan Ciro, Max Bartolo, Adina Williams, He He, et al. 2025. The PRISM alignment dataset: What participatory, representative and individualised human feedback reveals about the subjective and multicultural alignment of large language models. *Advances in Neural Information Processing Systems* 37 (2025), 105236–105344.

[18] Michèle Lamont and Virág Molnár. 2002. The study of boundaries in the social sciences. *Annual review of sociology* 28, 1 (2002), 167–195.

[19] Weixin Liang, Girmaw Abebe Tadesse, Daniel Ho, Li Fei-Fei, Matei Zaharia, Ce Zhang, and James Zou. 2022. Advances, challenges and opportunities in creating data for trustworthy AI. *Nature Machine Intelligence* 4, 8 (2022), 669–677.

[20] Joseph Lindley and Roger Whitham. 2024. From Prompt Engineering to Prompt Craft. *arXiv preprint arXiv:2411.13422* (2024).

[21] Naja Holten Moller, Claus Bossen, Kathleen H. Pine, Trine Rask Nielsen, and Gina Neff. 2020. Who does the work of data? *Interactions* 27, 3 (April 2020), 52–55. doi:10.1145/3386389

[22] Richard Mortier, Hamed Haddadi, Tristan Henderson, Derek McAuley, and Jon Crowcroft. 2014. *Human-Data Interaction: The Human Face of the Data-Driven Society*. SSRN Scholarly Paper ID 2508051. Social Science Research Network, Rochester, NY. doi:10.2139/ssrn.2508051

[23] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[24] Caroline Sinders. 2020. Feminist data set. *Clinic for Open Source Arts. https://carolinesinders. com/wp-con tent/uploads/2020/05/Feminist-Data-Set-Final-Draft-2020-0517. pd f* (2020).

[25] Chris Speed and Jon Oberlander. 2016. Designing from, with and by Data: Introducing the ablative framework. doi:10.21606/drs.2016.433

[26] Grace Leonora Turtle, Roy Bendor, Elisa Giaccardi, and Blazej Kotowski. 2024. Queering AI: Undoing the Self in the Algorithmic Borderlands. *arXiv preprint arXiv:2410.03713* (2024).

[27] Vera van der Burg, Gijs de Boer, Alkim Almila Akdag Salah, Senthil Chandrasegaran, and Peter Lloyd. 2023. Objective Portrait: A practice-based inquiry to explore AI as a reflective design partner. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 387–400.

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