

Living with AI – Critical Questions for the Social Sciences and Humanities

2023 WASP-HS Conference

WASP—HS

Contents

Living with AI – Critical Questions for the Social Sciences and Humanities	1
Contents	2
Introduction	3
Reboot: Ethical AI Through a Behavioral Lens.....	4
Exploring AI Futures Through Fictional News Articles.....	10
AI and Heritage Collections: Exploring Challenges and Opportunities.....	15
Workshop on Knowledge Elicitation: Data Collection and Knowledge Modeling for Human-Aware Systems	22

Introduction

The 2023 WASP-HS Conference AI for Humanity and Society (AI4HS) addressed the theme: “Living with AI – Critical Questions for the Social Sciences and Humanities”. The conference theme underscores the importance of critically examining the proliferation of AI systems in everyday practices. The conference provided a space for the WASP-HS community to engage with this theme through three broad areas of inquiry: (1) the critical conceptual approaches provided by the social sciences and humanities; (2) the underlying values, norms, and epistemological/ontological assumptions guiding AI research, which require illumination, scrutiny, development, or debate; and (3) the methodological considerations and decisions that could facilitate interdisciplinary collaboration. The pre-conference workshops, which were held for the first time, provided an additional opportunity to stimulate dialogue and the sharing of ideas animated by the conference theme.

The pre-conference workshops provided a platform for conference participants from WASP-HS and the wider community to meet around common interests, to engage in in-depth discussions and to build connections with individuals and research groups sharing similar concerns and questions. Following an open call and selection process, seven pre-conference workshops took place that welcomed ninety-nine participants from a variety of backgrounds and disciplines. The workshops covered a broad range of topics and themes that addressed the social and ethical issues related to the development and deployment of AI. The workshops engaged with a breadth of themes, from AI and heritage collections to fictional news articles, from knowledge elicitation to living with AI in education. In addition to a diversity of themes, the workshops utilized a range of innovative methods, from more traditional paper-based discussions to more experimental “hands on” empirical and design approaches.

As the contributions to these proceedings make clear, the workshops created a space for informed and critical discussions, fostered collaboration within the vibrant WASP-HS community and furthered a holistic and interdisciplinary approach to the examination of AI systems in our everyday practices. The workshop “Knowledge Elicitation: Data Collection and Knowledge Modeling for Human-Aware Systems” addressed the challenges that arise in the process of extracting structured descriptions of a domain from experts, such as capturing sufficient real-world information, and innovative methods for doing so such as through human-robot interactions studies and feminist frameworks. By contrast, the “Exploring AI Futures Through Fictional News Articles” workshop utilized the method of creating fictional news articles to foster critical discussions and reflections on possible AI futures. Mirroring this innovative methodology, the workshop “AI and Heritage Collections: Exploring Challenges and Opportunities” employed educational live-action role-playing (EduLarp) to investigate the challenges and opportunities provided by technological progress and the integration of AI and machine learning in the cultural heritage sector. Finally, the “Reconfiguring the Mold: Behavior Science and Ethical Tech Culture” workshop addressed how ethical principles can be effectively integrated into the practices, work and behavior of the technology industry. What shines through in these contributions is the value of creating spaces where a diversity of disciplines and views can address the critical questions that arise increasingly prevalence of AI systems in our lives and everyday practices.

Donal Casey and Liane Colonna

Reboot: Ethical AI Through a Behavioral Lens

Workshop Organizers: Annabel Gillard, Caroline Friedman Levy, Luca Nannini

Co-Authors: Nimish Gâtam, Anna King, Beatrice Tylstedt, Nitish Upadhyaya

Introduction to the workshop:

Tech companies' influence on human behavior has arguably eclipsed that of Freud, Piaget, Watson, and every other psychologist combined. Indeed, the industry has ridden its behavioral mastery to a new kind of global dominance. In addition to capturing attention, AI-powered tools help companies gauge, predict, and shape our actions, with remarkable success. Yet one can find scant evidence that tech companies have tried to leverage their behavioral know-how to meet their own commitments to ethical outcomes.

All the companies at the center of AI development – including Google, Microsoft, Anthropic, OpenAI, and Meta – have published principles statements promising their work will be: *fair, accountable, safe, and beneficent*/of benefit to humanity. Yet Google and Microsoft were caught using individuals' biometric data to train AI systems without permission (Hatmaker, 2023); the US Author's Guild has filed a class action suit against OpenAI and Microsoft for training on copyrighted works (Alter & Harris, 2023) and 41 out of 50 US states recently sued Meta for fostering a mental health crisis among children (Martin & Allyn, 2023).

How might behavioral science help tech companies adhere to their ethical promises? Our workshop gathered industry, academic, and policy practitioners to explore this question. Below we gather contributions on: the academic literature, voices from inside tech, parallel efforts within the financial sector, the unstudied risks of human-AI interaction, and a case study on risks accompanying AI-enabled products. We hope this contribution stimulates further conversation to support a reconfigured AI landscape that is genuinely fair, accountable, safe and beneficent.

Lessons from the academic literature

While behavioral science has yet to be factored into ethical AI implementation, decades of research on organizational behavior, behavioral economics, and cognitive and social psychology yield insights for practitioners operationalizing this work.

The well-studied COM-B model provides a useful lens, positing that behavior change hinges on *capability, opportunity* and *motivation* (Michie et al., 2011). Applying this to tech ethics highlights the need for practitioners to meaningfully engage with ethical principles, predicting compliance will improve when ethical charters are thoughtfully aligned with developers' motivations and practicalities (eg. Brey & Dainow, 2023). Studies show that leadership is fundamental in creating organizational change and in setting ethical norms, with supervisors' visible, ethical actions influencing employees (eg. Peng et al., 2021).

Non-compliance with ethical guidelines often stems from unintentional oversight or skewed incentives rather than deliberate neglect (Bazerman & Sezer, 2016). Thus, managers and leaders will need to minimize any friction that impedes ethical outcomes. Indeed, despite recent controversies surrounding "nudge theory", there is substantial evidence that interventions within workplace "choice architecture" can significantly shape behavior (eg. Thaler & Sunstein, 2009). Tech leaders committed to making their

ethical charters a reality will need to consider how component behaviors – for example, documentation allowing for model transparency and accountability – can be integrated into workflows such that they become the default action. Visibility matters, too: when ethical standards are explicitly emphasized, they are more likely to be incorporated into daily practices and decision-making processes (Adam & Rachman-Moore, 2004).

In sum, operationalizing AI ethics will require a thorough understanding of the factors that hinder or encourage the component behaviors for implementation. Interventions rooted in behavioral science will be essential to tech leaders committed to bridging the gap between lip service and enactment.

Voices from inside big AI

A question that haunted our workshop and ensuing research has been: is it naive to hope AI companies are motivated to align their work with their ethical charters? We know companies cannot themselves be “motivated”; their leadership, employees, customers and – where relevant – board of directors will play varying roles in a company’s direction of travel. Still, we wondered how AI insiders would view this topic. We were fortunate to have an insider in our workshop with contacts who have worked in implementing AI models and/or AI policy at Google, Facebook/Meta, Twitter, Amazon, Microsoft, and Alibaba.

Using a qualitative research methodology involving semi-structured interviews, our insider conducted recorded, virtual interviews with six of these AI professionals/colleagues, ensuring anonymity, regarding: challenges in adhering to ethics codes; external regulations and how they were upheld at their companies; and their personal concerns about AI’s influence on behavior and society. Each interview was transcribed and analyzed using thematic analysis to identify common themes. Some themes that emerged:

All respondents acknowledged the profound *tension* between *business incentives* – including profit maximization and/or shareholder value maximization – and *ethical standards* at their companies:

It’s all about the incentives and the metrics you optimize for. Because if these don’t include any outcomes that we are striving to achieve when it comes to civilization, society, human behavior, then none of the activities will be incentivized to optimize for these aspects. It’s easy to sacrifice those when the function you optimize is generally bottom-line revenue.

-Manager, Machine Learning, leading global tech company

Respondents pointed to challenges in adhering to ethical principles when *operationalization* is *unclear, or not clearly incentivized*:

I think having ethical aspects and rules...everybody should have them. But I think there are two levels as well, right? Because there are [the] people who created it and that might not be...[who] needs to stick with it.

-Manager, Machine Learning, leading global tech company

External regulations were seen as significant in shaping companies’ internal processes, with potential long-term benefits for adherence to ethical principles:

We are seeing this kind of trickling down effect of the GDPRs and CCPAs. I mean, companies are obliged to think more about those.... I can see them just becoming more baked-in in the functions you optimize within the company itself, because the incentive is not to get fined. So that's a pretty tangible incentive. You want to optimize for that. But not necessarily for the greater good....

-Head of Technology at a smaller AI-focused company

Lessons from ethical initiatives within the financial industry

The finance sector serves as a useful comparison for tech given parallel imbalances of power and information, and a mutual potential to cause societal harm. When the finance industry misjudged its ethical oversight, the resulting Global Financial Crisis (GFC) in 2007 caused several bank failures, billions of dollars in corporate losses, and bailouts using public funds (Mor, 2024) There are important lessons to be drawn for the tech industry.

Finance is a heavily regulated sector with mature governance structures policed by multiple 'lines of defense', including mandatory regulations and voluntary commitments to high ethical standards. Yet even this multi-faceted approach did not prevent the GFC. Having become 'too big to fail', posing a systemic risk which threatened wider devastation, banks needed rescuing by the public – somehow rewarding bad behavior and punishing the (public) victims. This serves as a stark warning, as the dominance of certain tech companies exceeds that of the large banks, offering potentially comparable systemic risks.

Notably, the GFC did not result from law-breaking but from overconfident risk assessment and a failure to adequately balance gains with potential harms. The resultant post-mortems have advocated a focus on ethical behavior and decision-making over structural changes. (Mikes & Power, 2023). Specifically, regulators have demanded more focus on conduct and organizational culture, rigorous financial stress testing and assigned personal legal responsibility to individuals. Pre-GFC, the sector focused on training, certification, personal attestation, and requirements for professional qualifications (with ethical codes) to perform particular roles. Post-GFC, forward-thinking firms have adopted behaviorally informed training programmes.

The multi-level oversight that failed to prevent the GFC is far more stringent than current AI governance. While AI regulations and ethical guidelines are emerging, they may yet prove inadequate, as was the case in the finance sector. Overconfidence and misaligned incentives, which contributed to the GFC, are also notable within tech companies. If these companies are committed to a safe and benign future for AI, they will learn from the failures of finance giants, and attend to the urgency of incentivizing ethical outcomes.

Ethical perils accompanying (increasing) human-AI interactions

Economists have long studied human-human interactions to understand how factors like trust, reciprocity and cooperation contribute to economic growth (eg. Algan & Cahuc, 2014). More recently, behavioral economics has furthered our understanding, offering a lens that has enriched and, in some cases, uprooted long-standing paradigms (Wittek et al., 2013). With the exponential increase in AI use leading towards a so-called "robot economy" (Agrawal et al., 2019), attention has turned towards understanding human-AI interactions. New frames of reference are needed, particularly as "AI's influence on human behavior is intricate and often counterintuitive" (Chugunova & Lacetera, 2023).

In the workplace, the integration of AI into traditionally human roles raises important questions about the employee experience and opportunities for humans to conduct

meaningful work (Bankins & Formosa, 2023). Further, the growing role of AI in human resource practices, from hiring to personnel development, raises concerns ranging from surveillance to the reproduction of implicit biases (Giermindl et al., 2022). Viewing such challenges through a behavioral lens can help us anticipate and address the ethical and social implications that will accompany increased human-AI interaction. Firstly, there are a range of potential reactions to interactions with artificial agents – from appreciation, to aversion, to individuals relinquishing control without adequate oversight (known as automation bias) (Dula et al., 2023) – with which the design and implementation of any AI tools in the workplace will have to contend.

The potential for ethical peril extends well beyond the workplace, affecting the political landscape (Bradshaw & Howard, 2017), having an impact on human psychology, not least in the sphere of relationships (Zimmerman et al., 2023), and in healthcare, where the benefits of AI need to be balanced across a range of potential risks (Karimian et al., 2022). While we have seen generalized ethical frameworks for human-AI interaction – for example, IEEE’s Ethically Aligned Design (Chatila & Havens, 2019), these remain incomplete without a human behavioral perspective built-in: from the design of the principles themselves, through to implementation and oversight. Ultimately, “...the rise of AI is not just a technological or economic phenomenon, but a behavioral one.” (Chugunova & Lacetera, 2023). A nuanced understanding of human behavior must guide AI development, oversight, and governance.

Case study: Ethical challenges of FemTech

To illustrate the ethical challenges accompanying AI-enabled technologies, we turn to FemTech – a portmanteau of Female Technology referring to technologies, wearables, and software that address female health and wellness. This rapidly growing industry generated \$1.15 billion in global venture capital investments in 2021 (Knickerbocker, 2022), with around half addressing reproductive health (FemTech Analytics, 2022), and the vast majority sold in the form of self-tracking apps.

Reproductive health apps tracking menstruation, fertility indicators, and pregnancy stages claim to give users control through increased knowledge about their menstrual cycles and reproductive markers. Yet there is tension between the declared benefits of FemTech and the products’ ethical risks. Gathering from public debate and prior research, we here describe two of the salient ethical issues accompanying FemTech:

Surveillance, data commodification and lack of privacy

FemTech companies have been culpable in startling data privacy breaches (eg. Paul, 2022). The revelation that the period app Flo leaked data to Facebook (Schechner & Secada, 2019) led to an FTC complaint in the US, and later settlement. Concerns about FemTech data privacy gained new resonance following the overturning of *Roe v. Wade* in the US in 2022, with the risks to users no longer reflecting simply targeted marketing, but legal prosecution (Kelly & Habib, 2023). After warnings that period tracking apps could reveal if a user had an abortion, many women in the US deleted these apps (Garamvolgyi, 2022). Researchers have also noted the risk of privacy breaches in apps delivered from employers through workplace wellness programs (Brown, 2021), raising concerns about repercussions based on reproductive data. Taken together, this body of critique points to the need for FemTech to be strictly regulated to prohibit data breaches (e.g. McMillan, 2022).

Transparency and accountability

FemTech apps are marketed as helping users gain knowledge about their own cycles and bodies, however, there is no transparency in how their algorithms work, and the ability of these apps to accurately predict ovulation, a key measure, has been questioned. A recent study revealed that apps standardize predictions for ovulation despite that medical research shows ovulation varies both from person to person and from cycle to cycle (Tylstedt et al., 2023). This lack of accountability and transparency in the apps starkly undermines the medical authority they claim.

Conclusion

We hear often about “the alignment problem” in AI development, but underlying this is an alignment problem within the companies building these powerful technologies. To fix this will require tech companies to align their cultures and incentives to their ethical commitments. In the absence of such industry-motivated transformation, we need external regulations to fill an incentivizing gap. Yet regardless of whether interventions are spurred by industry or by law, a nuanced understanding of human behavior will be central to creating an AI future which might truly benefit humanity.

References

- Adam, A. M., & Rachman-Moore, D. (2004). The methods used to implement an ethical code of conduct and employee attitudes. *Journal of Business Ethics*, 54, 225-244. <https://link.springer.com/article/10.1007/s10551-004-1774-4>
- Agrawal, A., Gans, J., & Goldfarb, A. (Eds.). (2019). *The economics of artificial intelligence: An agenda*. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/E/bo35780726.html>
- Algan, Y., & Cahuc, P. (2014). Chapter 2 - Trust, growth, and well-being: New evidence and policy implications. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (Vol. 2, pp. 49–120). Elsevier. <https://doi.org/10.1016/B978-0-444-53538-2.00002-2>
- Alter, A., & Harris, E.A. (2023, Sept 20). Franzen, Grisham and Other Prominent Authors Sue OpenAI. *The New York Times*. <https://www.nytimes.com/2023/09/20/books/authors-openai-lawsuit-chatgpt-copyright.html>
- Bankins, S., & Formosa, P. (2023). The ethical implications of artificial intelligence (AI) for meaningful work. *Journal of Business Ethics*, 185(4), 725–740. <https://doi.org/10.1007/s10551-023-05339-7>
- Bazerman, M. H., & Sezer, O. (2016). Bounded awareness: Implications for ethical decision making. *Organizational Behavior and Human Decision Processes*, 136, 95-105. <https://doi.org/10.1016/j.obhdp.2015.11.004>.
- Bradshaw, S., & Howard, P. (2017). *Troops, trolls and troublemakers: A global inventory of organized social media manipulation*. Computational Propaganda Research Project. <https://ora.ox.ac.uk/objects/uuid:cef7e8d9-27bf-4ea5-9fd6-855209b3e1f6>
- Brey, P., & Dainow, B. (2023). Ethics by design for artificial intelligence. *AI and Ethics*. <https://doi.org/10.1007/s43681-023-00330-4>
- Brown, E. A. (2021). The FemTech paradox: How workplace monitoring threatens women's equity. *Jurimetrics*, 61(3), 289-329. <https://www.proquest.com/docview/2568314630>
- Chatila, R., & Havens, J. C. (2019). The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. In M. I. Aldinhas Ferreira, J. Silva Sequeira, G. Singh Virk, M. O. Tokhi, & E. E. Kadar (Eds.), *Robotics and well-being* (Vol. 95, pp. 11–16). Springer International Publishing. https://doi.org/10.1007/978-3-030-12524-0_2
- Chugunova, M., & Lacetera, N. (2023). Editorial: The ethics and behavioral economics of human-AI interactions. *Frontiers in Behavioral Economics*, 2. <https://doi.org/10.3389/frbhe.2023.1338608>
- Dula, I., Berberena, T., Keplinger, K., & Wirzberger, M. (2023). Hooked on artificial agents: A systems thinking perspective. *Frontiers in Behavioral Economics*, 2. <https://doi.org/10.3389/frbhe.2023.1223281>
- Epstein, D. A., Lee, N. B., Kang, J. H., Agapie, E., Schroeder, J., Pina, L. R., Fogarty, J., Kientz, J. A., & Munson, S. (2017). Examining menstrual tracking to inform the design of personal informatics tools. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 6876–6888. <https://doi.org/10.1145/3025453.3025635>

- FemTech Analytics. (2022, December). *FemTech Industry Landscape Q4 2022 Teaser*. <https://www.femtech.health/report-q4-2022>
- Fox, S. E., Menking, A., Eschler, J., & Backonja, U. (2020). Multiples over models: interrogating the past and collectively reimagining the future of menstrual sensemaking. *ACM Transactions on Computer-Human Interaction*, 27, 1–24. <https://doi.org/10.1145/3397178>
- Garamvolgyi, F. (2022, June 28). Why US women are deleting their period tracking apps. *The Guardian*. <https://www.theguardian.com/world/2022/jun/28/why-us-woman-are-deleting-their-period-tracking-apps>
- Giermendl, L. M., Strich, F., Christ, O., Leicht-Deobald, U., & Redzepi, A. (2022). The dark sides of people analytics: Reviewing the perils for organisations and employees. *European Journal of Information Systems*, 31(3), 410–435. <https://doi.org/10.1080/0960085X.2021.1927213>
- Hatmaker, T. (2020, July 15). Lawsuits allege Amazon, Microsoft and Google violated Illinois facial recognition privacy law. *TechCrunch*. <https://tinyurl.com/4nr3yx44>
- Heyder, T., Passlack, N., & Posegga, O. (2023). Ethical management of human-AI interaction: Theory development review. *The Journal of Strategic Information Systems*, 32(3), 101772. <https://doi.org/10.1016/j.isis.2023.101772>
- Karimian, G., Petelos, E., & Evers, S. M. A. A. (2022). The ethical issues of the application of artificial intelligence in healthcare: A systematic scoping review. *AI and Ethics*, 2(4), 539–551. <https://doi.org/10.1007/s43681-021-00131-7>
- Kelly, B. G., & Habib, M. (2023). Missed period? The significance of period-tracking applications in a post-Roe America. *Sexual and Reproductive Health Matters*, 31(4), 2238940. <https://www.tandfonline.com/doi/full/10.1080/26410397.2023.2238940>
- Knickerbocker, K. (2022). What is Femtech? <https://pitchbook.com/blog/what-is-femtech>
- Martin, M., & Allyn, B. (2023, Oct 25). Meta sued by states claiming Instagram and Facebook fueled youth mental health crisis. *NPR: Morning Edition*. <https://www.npr.org/2023/10/25/1208391495/meta-sued-by-states-claiming-instagram-and-facebook-fueled-youth-mental-health-c#:~:text=Forty%2Done%20states%20are%20suing.protection%20and%20child%20safety%20laws.>
- McMillan, C. (2022). Monitoring female fertility through 'FemTech': The need for a whole system approach to regulation. *Medical Law Review*, 30(3), 410-433. <https://doi.org/10.1093/medlaw/fwac006>
- Mikes, A., & Power, M. (2023). How culture displaced structural reform: Problem definition, marketization, and neoliberal myths in bank regulation. *Journal of Business Ethics*. <https://doi.org/10.1007/s10551-023-05530-w>
- Mor, F. (2024). *Bank rescues of 2007-09: Outcomes and cost*. <https://commonslibrary.parliament.uk/research-briefings/sn05748/>
- Paul, K. (2022). How private is your period-tracking app? Not very, study reveals. *The Guardian*. <https://www.theguardian.com/world/2022/aug/17/pregnancy-period-tracking-apps-privacy>
- Peng, J., Li, M., Wang, Z., & Lin, Y. (2021). Transformational Leadership and Employees' Reactions to Organizational Change: Evidence From a Meta-Analysis. *The Journal of Applied Behavioral Science*, 57(3), 369–397. <https://doi.org/10.1177/0021886320920366>
- Schechner, S., & Secada, M. (2019). You give apps sensitive personal information. Then they tell Facebook. *The Wall Street Journal*. <https://www.wsj.com/articles/you-give-apps-sensitive-personal-information-then-they-tell-facebook-11550851636>
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.
- Tylstedt, B., Normark, M., & Eklund, L. (2023). Reimagining the cycle: interaction in self-tracking period apps and menstrual empowerment. *Frontiers in Computer Science*, 5. <https://doi.org/10.3389/fcomp.2023.1166210>
- Wittek, R., Snijders, T. A. B., & Nee, V. (2013). *The handbook of rational choice social research*. Stanford University Press. <https://www.sup.org/books/title/?id=21110>
- Zimmerman, A., Janhonen, J., & Beer, E. (2023). Human/AI relationships: Challenges, downsides, and impacts on human/human relationships. *AI and Ethics*. <https://doi.org/10.1007/s43681-023-00348-8>

Exploring AI Futures Through Fictional News Articles

Authors: Martin Lindstam, Uppsala University; Elin Sporrang, Stockholm University; Camilo Sanchez, Aalto University; Petra Jääskeläinen, KTH Royal Institute of Technology

Abstract

The aim of this workshop was to enable critical discussion on AI futures using fictional news articles and discussion groups. By collaboratively imagining and presenting future scenarios in a journalistic news article format, participants explored the socio-political, ethical and sustainability factors of AI through an accessible narrative form. Participants engaged in further anticipatory work by analyzing the issues raised by the articles in a group discussion, emphasizing the underlying motivations, assumptions and expectations conveyed within the news articles.

Introduction

There has been an increasing attention to discussing ethics, sustainability, and politics of AI technologies. Along with the increasing urgency of discussing the consequences of AI technologies, there has also been a surge of interest in Design Fiction during the last 10 years (Bleecker, 2009; Tanenbaum et al., 2012; Dunne, 2013; Baumer, 2014; Penzenstadler, 2014; Blythe, 2014). This workshop made use of *fictional news articles* to materialize our imaginaries on AI and discuss how it will affect and change society now and in the future. By focusing on the *ethics, sustainability, and politics of AI futures*, we aimed to understand the potential future consequences of AI ten years from now – in 2034 – and how we can engage in the responsible design of AI.

This workshop was based on ICCC '23 workshop on Fictional Abstracts (“Workshops - ICCC 2023”, 2023). As facilitators, we wanted to engage researchers and practitioners who are interested in the ethical reflection of emerging AI technology and practices. To foster a wide-ranging and critical discourse, we encouraged participants from different disciplines and backgrounds to bring their unique perspectives to the table. The workshop sought to offer a contribution by developing sensitivities on how to move forward in the responsible design and use of AI technologies.

The outcomes of the workshops showed diverse ways in which participants engaged with fictional news articles. While one group focused on extrapolating current trends – self-tracking and healthy lifestyles – into a dystopian speculative scenario of an AI-enabled lunar rave, the other presented emergent uses of AI by describing the development of an AI-driven food system in a rural community. Many discussions emerged during the workshop regarding themes, such as “world-building” and how power was distributed in the future scenarios. Another point for discussion was how perspectives are projected in the news articles. While the articles represented the authors’ worldviews, readers themselves bring in their own perspectives to make sense of the speculation.

In the subsequent part of this paper, we will provide an account written by two of the workshop participants, reflecting on their experiences of the workshop, and the method of fictional news articles.

Reflections from two workshop participants

Speculative methods invite people to engage with various temporal frames, what-ifs and counterfactual scenarios, to generate debate and new sociotechnical visions (Meskus & Tikka, 2022). While such creative engagements can spark inspiration and hope, we entered the workshop with a critical perspective drawing from our backgrounds in journalism, foresight and history of ideas.

Working with speculative methods, Elin has incorporated critical perspectives into her research on anticipations of AI in higher education, building on e.g. the discipline of anticipation, which provides tools for unpacking and expanding visions of the future (Miller et al, 2018). By identifying narratives and visions of AI technologies and the automation of teachers' practices, Elin's work aims to provide insights into how teachers' anticipations of AI, their assumptions and their values framing the future are interconnected in the context of higher education.

Coming from a journalistic background, Martin entered the workshop enmeshed in the emergent field of existential media studies (Lagerkvist, 2020), which seeks to recenter the often sidestepped and vulnerable human in visions of AI futures. The field of existential media studies emphasizes the future as a horizon of possibility. Building on this view of the future, Martin is working with Berardi's concept of futurability; denoting the idea that the horizon of possibility exists beyond entrenched contemporary norms and structures (Berardi, 2017). In his research, Martin is interested in techno-existential perspectives on sociotechnical imaginaries (Jasanoff & Kim, 2015), mainly concerning implementations of AI in the workplace.

With these backgrounds, both workshop participants attended the workshop with an understanding of news situated in a Swedish context, in which journalism is informed by a specific set of nationally agreed upon guidelines (Journalistförbundet, 2021) and media is commonly referred to as a third public power (Riksdagen, 2023). By these understandings of journalism, news writing is a narrative practice of compressing large amounts of information into intelligible text (Häggbom, 2020), structured around an understanding of public interests. As such, journalism is "society's conversation with itself", and newsworthiness is normatively informed (Häger, 2021). The prospect of news articles as venues for anticipation is interesting as the news article format is entangled in the web of journalistic moral and ideological practices (Gardeström, 2020), norms and mission statements.

Producing the speculative artifact

Inspired by the format of fictional news articles, we considered how current narratives, values and assumptions regarding AI in society might be brought into our creative process of envisioning future trajectories, and how we could critically engage with our own anticipatory processes as our speculative visions unfolded in the workshop. In particular, we critically considered: i. newsworthy vs ordinary futures; ii. grand narratives of AI and; iii. fallacies in futuring for broadening our own thinking and imagination.

i. Newsworthy vs ordinary futures

We considered that news reportages have the ability to transport media consumers to places they have never been, and to meet people they otherwise never would have met (Häggbom, 2020). Additionally, news reportages note the extraordinary – the newsworthy – from a normative standpoint of the journalist or newsroom, in response to people's interests in and of the surrounding world. As such, news as a speculative format may illuminate people's reflections on the newsworthy and ordinary. To imagine new places and the extraordinary, we first had to consider the familiar and the ordinary.

ii. Grand narratives of AI

We contemplated domains of ordinary life and society within our lifeworlds. Being both teachers and students in higher education, we turned to the context of education that currently reflects the international political competition in technological progression, enacted through various AI strategies and policies (Bareis and Katzenbach, 2021). Situated in Western academia in 2023, our discursive understanding of ordinary education and future was discussed by critically noting current dominant narratives of AI in education, understood as an inevitable and necessary solution in education (Rahm and Kaun, 2022; Facer and Selwyn, 2021), echoed in our everyday contexts. Such narratives may marginalize alternative educational visions that draw on emerging perspectives in and on the world (Sporrong, 2024). Building on this understanding of the role of narratives, speculative methods may be more or less open to emerging perspectives by their capacity to both critically unpack and reproduce narratives that displace emerging thoughts and ideas (Ross, 2022).

iii. Fallacies in futuring for broadening our own thinking and imagination

Speculation may draw on common errors in reasoning about the future that can hinder critical speculation (Dorr, 2017). Building on Dorr (2017), we considered ways to engage with the complexity of change and to imagine the future beyond the horizon set through an extension of the past. Thus, to critically consider dominant narratives and disentangle ourselves from fallacious prospective thinking, we tried to de-center Western tech utopias in our speculative artifacts and to re-imagine what education might mean in the future.

In this process, we situated ourselves in an imagined future in which education focused on sociality and wellbeing, rather than efficiency. Considering the environmental crises as an intertwined factor with future wellbeing, and thereby with educational development, we discussed how promises of “more efficient” education and environmental costs of technology stand in friction with environmental and human existential needs and the survival of living beings. As such, we problematized current projections of better education by AI and considered what AI could bring to education should the environment be a central and intertwined part of everyday school. Through this critical stance, we moved toward ideas of futures articulated by Berardi (2017), who underscores the importance of disentangling from entrenched norms and structures in order to perceive the horizon of possibility.

Producing the speculative artifact

In our speculative project we set out to construct a rich and nuanced narrative as encouraged by the workshop facilitators. We began by situating education within a fictive rural municipality in which the community, learning, people’s wellbeing and the environment were tightly connected aspects of ordinary life. In this vision, the educational and sustainability goals were interrelated for example, by schools working towards self-sufficiency through farming, situating schools within the broader context of environmental sustainability. By framing schools as central actors in common efforts to live more sustainably, we envisioned EarthTrack, a monitoring technology for streamlining agriculture that had been attuned to the need in private and public schools and their roles in the community. Two stakeholder perspectives were presented in the fictional news article to illustrate a tension and power dynamic in the vision: the perspective taken by a tech-positive local politician and that of a more cautious school principal. Supplying the final lines of the article, the principal is quoted: “I think we are seeing, particularly in the West, the many problems that schools have encountered because they were too quick to rush into this. That we have the possibility to do the same does not mean that we should”.

By engaging in non-dominant visions of slow and alternative AI-implementation in education within the news article format, we were, to some extent, able to consider the complexity of the speculative scenario we imagined.

Discussion and conclusion

The fictional news article method as outlined in the workshop is expressly reflective by viewing the future as an open space where multiple possible futures may be enacted. Berardi (2017) emphasizes the importance of this reflective approach not just in thinking about the future, but also humans' place in it. In line with Dorr (2017), Berardi (2017) describes and laments a reduction of the possible to the probable, and the probable to the necessary. Through our discussions, it became clear that we sought to generate expansive visions of the future by taking an awareness of the fallacious and reductive envisioning of the future.

Being prompted with multiple reflective questions, such as: “Are there any elements, contexts or situations from the article that we can already notice today?”, and asked to reflect on the social, cultural, and environmental values present in our article, we found that many assumptions about education and currently guiding values were embedded in our vision. Despite our attempts to disentangle ourselves from dominant narratives, we identified extrapolative tendencies in our speculative artifact, reflected in the imagined future education system and ways of living in the rural municipality. For instance, physical schools were envisioned as a constant in the future, situated in or in conflict with a political discourse. Additionally, our community-oriented ideas that were central to our vision of sustainability drew on current ideas of communality - municipal private and public physical schools situated within a rural area where people lived together. Such ideas excluded alternative visions of schooling, nomad lifestyles and environmental tipping points that could completely alter our current lifestyles.

Considering such excluded visions, we note that AI visions may be speculative by various degrees and dependent on the materiality and normativity of the speculative method applied (Cerratto Pargman et al., 2023). For example, fictional news articles require multiple instances or sequences of speculation whereby we imagine not only the newsworthy event itself, but the actors, institutions and values that configure it as such – a complex web of moving parts, mediating our speculation. The materiality of the news article genre applied in the workshop was instantiated by specific stylistics and news-formats such as opinion pieces, interviews etc. By constructing visions by such material choices, we identified a further exclusion of alternative visions of journalism. Journalism became a lens for speculation rather than a practice, somewhat distanced from the logics and structures that commonly shape the news. As a lens, we found that journalism both challenged us and presented us with opportunities for speculation.

Understanding journalism as society's conversation with itself, we reflected on the nature of such a conversation in a fictive world. Journalism, ideally, focuses on human stories that motivate the newsworthiness of events by their importance in people's lives (Häger, 2021). Re-centering the human is a core concern of existential media studies, and we note that the fictional news article workshop method lends itself well to discussions of what it means to be human in relation to technology. As a closing session in the workshop, we discussed our intent with the construction of the speculative news article, noting that news carries with it normative assumptions of what is important information for the public. As such we found that in our efforts to adhere to the news article format, we were inadvertently making normative statements about what is important public information.

Finally, we consider that any process of disentanglement of dominant narratives may further substantiate, or reinforce them. For instance, we identify that efforts to disentangle from dominant narratives may be “reactionary”, by arranging and structuring visions in response to visions under critique. In this workshop context of speculation, we identify that the collective and reflective closing discussion sparked new ideas and critical perspectives on anticipatory processes taking place, and what participants “bring into the future”.

References

- Baumer, E.P., Ahn, J., Bie, M., Bonsignore, B., Borutecene, A., Buruk, O.T., Clegg, T., Druin, A., Echtler, F., Gruen, D., Guha, M. L., Hordatt, C., Kruger, A., Maidenbaum, S., Malu, M., McNally, B., Muller, M., Norooz, L., Norton, J., Özcan, O., Patterson, D., Riener, A., Ross, S., Rust, K., Schoening, J., Silberman, M. S., Tomlinson, B., and Yip, J. (2014). “CHI 2039: Speculative Research Visions”. In Proc CHI ’14. ACM, 761-770.
- Berardi, F. (2017). *Futurability: The Age of Impotence and the Horizon of Possibility*. Verso books.
- Bleecker, J. (2009). *Design Fiction: A short essay on design, science, fact and fiction*. Near Future Laboratory, 29.
- Blythe, M. (2014). Research through design fiction: narrative in real and imaginary abstracts. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 703-712). ACM.
- Cerratto Pargman, T., Lindberg, Y. & Buch, A. Automation Is Coming! Exploring Future(s)-Oriented Methods in Education. *Postdigit Sci Educ* 5, 171–194 (2023). <https://doi.org/10.1007/s42438-022-00349-6>
- Dorr, A. (2017). Common Errors in Reasoning about the Future: Three informal fallacies. *Technological Forecasting & Social Change*, 116. 322–330. <https://doi.org/10.1016/j.techfore.2016.06.018>.
- Dunne, A., & Raby, F. (2013). *Speculative everything: design, fiction, and social dreaming*. MIT Press.
- Facer, K., & Selwyn, N. (2021). Digital Technology and the Futures of Education: Towards ‘Non-stupid’ Optimism. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000377071>.
- Gardeström, E. (2020). Vad är en journalist? In Gardeström, E., & Rehnberg, H. (Ed.) *Vad är journalistik?* 17-16.
- Häger, B. (2021). *Reporter: En grundbok i journalistik* (3rd ed.). Studentlitteratur.
- Häggbom, M. (2020). Journalistik är att formge berättelser. In Gardeström, E., & Rehnberg, H. (Ed.) *Vad är journalistik?* 61-65.
- Workshops - ICCC 2023. (2023). *Computational Creativity*. <https://computationalcreativity.net/iccc23/workshops/>
- Jasanoff, S., & Kim, S-H. (2015). *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. The University of Chicago Press
- Journalistförbundet. (2021). Publicitetsregler. <https://www.sjf.se/yrkesfragor/yrkesetik/spelregler-press-radio-och-tv/publicitetsregler>
- Lagerkvist, A. (2020). *Existential Media: A Media Theory of the Limit Situation*. Oxford University Press.
- Meskus, M., & Tikka, E. (2024). Speculative approaches in social science and design research: Methodological implications of working in ‘the gap’ of uncertainty. *Qualitative Research*, 24(2), 209-228. <https://doi.org/10.1177/14687941221129808>
- Miller, R., Poli, R. Rossel, P. (2018) The Discipline of Anticipation: Foundations for Futures Literacy In: Miller, R (Ed.) *Transforming the Future: Anticipation in the 21st Century*, (pp 48–60). Routledge
- Penzenstadler, B., Tomlinson, B., Baumer, E., Pufal, M., Raturi, A., Richardson, D., Cakici, B., Chitchyan, R., Da Costa, B., Dombrowski, L., Picha Edwardsson, M., Eriksson, E., Franch, X., Hayes, G.R., Herzog, C., Lohmann, W., Mahaux, M., Mavin, A., Mazmanian, M., Nayebaziz, S., Norton, J., Pargman, D., Patterson, D.J., Pierson, J., Roher, K., Silberman, M.S., Simonson, K., Torrance, A.W., and van der Hoek, A. (2014). *ICT4S 2029: What will be the systems supporting sustainability in 15 years*. In Proceedings of the 2014 conference ICT for Sustainability (pp. 30-39). Atlantis Press.
- Rahm, L., & Kaun, A. (2022). Imagining Mundane Automation: Historical Trajectories of Meaning Making Around Technological Change. In S. Pink, M. Berg, D. Lupton, & M. Ruckenstein (Eds.), *Everyday Automation: Experiencing and Anticipating Emerging Technologies* (pp. 23–43). London: Routledge. <https://doi.org/10.4324/9781003170884-3>.
- Riksdagen. (2023). The Riksdag in Swedish society. <https://www.riksdagen.se/en/how-the-riksdag-works/democracy/the-riksdag-in-swedish-society/>
- Ross, J. (2022). *Digital Futures for Learning: Speculative Methods and Pedagogies* (1st ed.). New York: Routledge. <https://doi.org/10.4324/9781003202134>.
- Sporrong, E. (2024). Anticipation: Understanding the Role of Anticipation in Configuring Visions of Education. In Buch, A., Lindberg, Y., Cerratto Pargman, T. (eds). *Framing Futures in Postdigital Education: Critical Concepts for Data-driven Practices*. Springer
- Tanenbaum, J., Tanenbaum, K., & Wakkary, R. (2012). Design fictions. In Proc. TEI’12. ACM, 347-350.

AI and Heritage Collections: Exploring Challenges and Opportunities

Authors: Paulina Rajkowska, Anna Foka, Dalia Ortiz Pablo, Sushruth Badri, Larissa von Bychelberg, Nasrin Mostofian, Pille Pruulmann Vengerfeldt

Introduction

The purpose of the "AI and Heritage Collections: Exploring Challenges and Opportunities" workshop, held on November 14th 2023, was to explore the current state of knowledge about artificial intelligence (AI) and machine learning (ML) within the Cultural Heritage Sector (CHS) and how introduction of these solutions may affect the Swedish organizational context. The workshop brought together a group of eight participants including researchers and practitioners active within CHS. The event consisted of two parts: a roleplay scenario meant to stir up discussion and a follow up brainstorming session that was meant to channel the points lifted in the first part and organize them in a research agenda that could be applicable within the different academic fields represented by the participants.

Through both exercises the workshop participants investigated opportunities and challenges created by the rapid technological progress and increasing integration of AI/ML technologies in our professional lives. By exploring the perspectives and experiences of Swedish CHS, this paper aims to provide insights that can inform the development of strategic approaches to address the infrastructure and qualification needs arising from the growing influence of AI/ML in the cultural heritage sector. The paper highlights that the growing ability to digitally organize and render heritage collections has opened up many new possibilities for Cultural Heritage Institutions (CHIs). As Geismar (2018) notes, this digitalization process has created numerous opportunities for these institutions. However, this paper also acknowledges that the digitalization process requires careful consideration of the nuances of past heritage contexts to ensure appropriate meaning-making, as emphasized by Risam (2018) and Foka, Eklund et al (2023). This is crucial to preserve the integrity and significance of the cultural heritage being digitalized. To contribute to this discussion, we first introduce the methods employed during the workshop to then proceed to elaborating on the points lifted by the various participants. Finally, we synthesize the thoughts that drove the workshop discussion to articulate the most urgent areas of interest for academic research concerning AI/ML technology in CHS.

The EduLarp

Educational live-action role-playing (EduLarp) has been increasingly recognized as a valuable tool for fostering empathy and understanding among diverse audiences (Bowman, 2014). By engaging in roleplay activities, participants have the opportunity to step into the shoes of others, gaining insight into perspectives different from their own and developing a deeper emotional connection to the subject matter under discussion (Waern et al., 2020). To tap into this potential, the workshop's first activity was an EduLarp scenario written by the organizers and titled "The AI Project". The setting for the scenario

was a project meeting at a museum institution. Each participant was assigned a role and was provided with relevant materials outlining the scenario. They were given a few minutes to name their characters and understand how they would behave in the given situation, relying on their prior experience with museum institutions and processes of digital innovation. They were then briefed on the task at hand: to consider different approaches to implementing AI in CHS. Two AI projects were presented, each with its own set of advantages and drawbacks. As the discussion unfolded, participants engaged in lively debate, weighing the potential risks and benefits of each project. Some expressed concerns about the ethical implications of AI in cultural heritage, while others highlighted the potential for innovation and preservation. To wrap up the scenario, the facilitator guided the group through a debriefing session. Participants reflected on which project was ultimately chosen and the reasoning behind the decision. They also discussed the various issues that were raised during the discussion, such as data privacy, bias in algorithms, and the role of AI in cultural preservation.

Brainstorming

Following a brief break, participants engaged in a rapid-fire brainstorming session, transitioning from the roleplay session to a more structured exercise. Their task: to translate the insights gained from the roleplay into tangible ideas by jotting them down on sticky notes. These notes were then organized into overarching themes, capturing both potential points of contention and avenues for productive exploration. Once that was done, the participants discussed the outcomes and formulated possible areas of research or research questions that could become relevant starting points for future academic work in this domain.



Fig. 1 Photography of the results of the brainstorming including themes

Outcomes

The brainstorming resulted in the following categories: data, audience, training, money, labor, purpose, product, balance, ethics, time, and future proofing. In this section we briefly summarize the essence of these labels and how it connected to questions and conflicts observed during the roleplay session.

Audience, purpose, balance

One of the central inquiries revolved around the target audience for the implementation of AI within the cultural heritage domain. While AI technology holds promise for various groups such as visitors, practitioners, and researchers, each of these stakeholders possesses distinct and divergent needs. For practitioners and researchers, AI implementations offer the potential to streamline data organization and processing, thereby enhancing efficiency and productivity in professional contexts (Griffin et al., 2023). However, the appeal of such solutions may be limited to a select group of relevant parties within these fields. In contrast, visitors stand to benefit from AI-driven solutions that enhance their on-site experiences, placing greater emphasis on public engagement and user-centric design (Gaia et al., 2019). This necessitates a focus on user experience and usability, with AI applications tailored to cater to a diverse audience base with varying ages, educational backgrounds, and levels of technological proficiency.

Moreover, the ultimate purpose of implementing AI solutions varies significantly across these stakeholder groups. For practitioners, AI initiatives may prioritize improving accessibility, facilitating preservation efforts, or optimizing internal organizational processes. Similarly, AI-driven visitor experiences could encompass initiatives aimed at enhancing accessibility, introducing innovative ways of interacting with cultural heritage, developing educational programs, or increasing visitor engagement and attendance through novel offerings. However, navigating these diverse needs and objectives is fraught with challenges, particularly within the complex landscape of cultural politics at both local and state levels. Practical constraints such as limited time and resources further complicate matters, often leading to oversimplification and homogenization of stakeholder needs within each category. Of particular concern is the risk of marginalizing traditional visitor populations, who form a significant segment of existing museum audiences. Overemphasizing AI solutions without considering the preferences and expectations of these audiences could lead to their alienation and disengagement from cultural heritage institutions.

Addressing these multifaceted considerations requires a delicate balancing act, one that the cultural heritage sector is currently grappling with. Workshop participants identified decision-making support in organizational contexts as a promising avenue for academic intervention, highlighting the potential for scholarly input to inform strategic approaches that reconcile the diverse needs and priorities of stakeholders while fostering innovation and inclusivity.

Labor, training and money

All the possible activities, while enticing, require an extensive amount of work. Tasks such as data scanning, data selection, forming the potential solutions, involving relevant stakeholders, all the way through actual development and implementation are time consuming activities which would have to be incorporated into existing work practices of institutions working with them. How could the increased workload be tackled and ultimately who would be responsible for the quality of the final product? As existing scholarship on digitalization processes shows, museum employees are already

overwhelmed with the amount of work required by increasing technological demands. For example, Foka et al. (in press 2023) write that a lack of knowledge about digitalization is a general problem in many museums. Additionally, in Not & Petrelli's (2019) study, cultural heritage workers expressed concern with digital technology, although they seemed positive about its implementation.

Moreover, the specialized skills required to develop AI solutions are not evenly distributed among cultural heritage practitioners, further exacerbating the challenge. With limited training opportunities and a rushed, unstructured approach to task execution, practitioners may struggle to adapt to the new demands placed on them. This is compounded by chronic underfunding and understaffing within museums, where the introduction of additional tasks without corresponding increases in resources could prove detrimental to practitioners' well-being and the quality of their work. Cushing & Osti (2023), who have interviewed archivists about the transition to working with AI in archives, have encountered worries about how AI – just as the more general digitalization process that happened earlier – would be considered a panacea by management that takes funding away from more traditional archival work such as appraisal.

Considering these challenges, addressing the increased workload associated with AI implementation requires a multifaceted approach. This may include investing in employee training and professional development, securing adequate resources, and fostering a culture of innovation and collaboration within cultural heritage institutions. By prioritizing the well-being of practitioners and promoting sustainable practices, the sector can navigate the complexities of AI integration while preserving its core mission of safeguarding and promoting cultural heritage for future generations.

Data

The quality of existing data within the cultural heritage sector emerged as a significant focal point of discussion among workshop participants, underscoring the critical importance of data integrity and consistency in driving effective AI applications. While digitalizing cultural heritage collections has become increasingly widespread, the absence or scarcity of standardized data quality standards poses a formidable challenge (Ioannides 2005). Metadata, essential for contextualizing collection objects such as photographs, paintings, and sculptures, exhibits significant variability in quality, influenced by factors such as intended use, responsible parties, funding availability, and more.

AI applications necessitate vast quantities of data prepared in a specific format, highlighting the pressing need to develop a deeper understanding of data requirements and standards within the cultural heritage context. The workshop's roleplay exercise shed light on the divergent perceptions of data's function among different stakeholder groups and underscored the varying degrees of comprehension regarding the prerequisites for functional AI solutions. The concept of data agency emerged as a recurring theme, emphasizing the dynamic relationship between data and its users across diverse contexts. The process of data gathering and preparation encompasses a range of activities, from documentation and digitalization to curation and the training of AI models, culminating in the assessment of their usability. The determination of necessary professional skills is intricately linked to the intended purpose of the technology and its end-users. Consequently, the quality of the final solution is heavily contingent upon the time invested, resources allocated, and the expertise of practitioners involved in the process.

Addressing these challenges necessitates a holistic approach that encompasses capacity-building initiatives, the establishment of robust data governance frameworks, and the cultivation of interdisciplinary collaboration. By investing in the development of standardized data quality protocols, promoting knowledge exchange, and fostering a

culture of continuous learning and innovation, the cultural heritage sector can harness the transformative potential of AI technologies while safeguarding the integrity and accessibility of its rich cultural heritage for future generations.

Future proofing/sustainability

All the above considerations led to questions of how viable and sustainable AI solutions are for the cultural heritage sector. Despite the widespread excitement surrounding the latest advancements in AI technology, its practical applicability within the sector remains uncertain. A critical issue that emerged is the compatibility of older digitalized data with the requirements of new AI technologies. In many instances, legacy datasets do not align seamlessly with the capabilities of AI systems, necessitating significant labor to adapt or redo previous work. This poses challenges not only in terms of resource allocation but also in maintaining the integrity and accuracy of historical records and artifacts. Without a comprehensive vision of how AI technology might evolve and integrate into the sector in the future, there's a risk that hastily incorporating it into existing institutional frameworks may lead to unsustainable work practices and suboptimal outcomes.

However, for example Korke et al. (2022) have proposed a model demonstrating how AI techniques can help safeguard cultural heritage sites, providing evidence that AI can be an asset to the sector regarding long-term sustainability of cultural artifacts. Also, Bastian (2023, Introduction) seems positive about the possibilities that digitalization offers to archives, claiming that it will lead to an expansion of the possibilities for recording people, and a transformation of traditional archival work. Additionally, Villaespesa & Crider (2021) and Cieccko (2017) predict potential for classifying collections in new ways using AI. As the subject is still very new, and the opinions of experts vary greatly, the group identified a definite need for further scholarship in this area.

Ethics

Another issue that weaved its way through the other topics discussed at the workshop was the question of ethics. While there was a significant emphasis placed on the latest technological developments and how they can potentially transform a certain industry or practice, the participants felt that we often lost track of the human aspect of this issue. The introduction of AI solutions holds profound implications for how individuals engage with cultural heritage and connect with their historical narratives (Manžuch, 2017). The very essence of cultural heritage lies in its significance to human experience, memory, and identity. As such, any technological intervention must be approached with sensitivity to its impact on these deeply personal and communal connections. One of the key concerns raised was the reliance on large datasets favored by AI algorithms. While these datasets offer unparalleled opportunities for analysis and insight generation, they also present inherent biases and limitations. Participants highlighted the potential consequences of such biases on the construction of historical narratives. In an era where inclusivity and representation are paramount, there's a growing recognition of the need to amplify marginalized voices and narratives within cultural heritage discourse.

In recent years, cultural heritage institutions have increasingly embraced narrative approaches that center the experiences and perspectives of minority and previously excluded groups. These efforts aim to rectify historical injustices and present a more nuanced understanding of the past. However, the adoption of big data approaches in AI may inadvertently dilute these efforts by prioritizing the narratives of well-documented and prominent historical figures or events. This is well exemplified by Villaespesa & Crider (2021) who mention the proliferation of bias in commercial AI models, such as facial

recognition models that recognize white males more easily than non-white females (Buolamwini & Gebru, 2018).

Agenda for future research

The potential of AI in various industries is widely acknowledged, yet its applicability within the cultural heritage sector remains a subject of skepticism and ambiguity. Workshop participants delved into the intricacies of this topic, drawing upon their individual experiences and insights gleaned from existing scholarship. Through discussion, the group identified several key challenges and opportunities, paving the way for the identification of research areas where the academic community can make significant contributions.

Balancing institutional stakeholders emerged as a critical consideration, with museums bearing the responsibility of safeguarding cultural heritage for posterity. Traditional practices rooted in manual curation and conservation techniques form the cornerstone of preservation efforts spanning generations. Introducing AI technologies into this landscape necessitates a delicate balance to ensure that they complement rather than undermine existing preservation methodologies. Understanding the agency of data and heritage workers emerged as another salient theme. These individuals play a pivotal role in collecting, curating, and maintaining the datasets used to train AI models. Recognizing their agency is paramount, as it ensures that their ethical concerns, biases, and values are duly considered throughout the AI development process. By incorporating their perspectives, institutions can mitigate biases and promote fairness in AI systems. The workshop also delved into the complexities of defining audiences, beneficiaries, and potential ethical risks inherent in AI implementation within the cultural heritage sector. Identifying the primary stakeholders and understanding their needs and concerns is essential for developing AI solutions that resonate with diverse user groups while upholding ethical standards and values. A crucial aspect that garnered attention was the labor dynamics associated with AI adoption. While AI technologies have the potential to automate repetitive tasks and streamline processes, they also raise concerns about job displacement and the need for workforce reskilling. The workshop participants recognized the importance of addressing these labor implications and exploring how they differ from prior technical revolutions.

In summary, the workshop provided a platform for deepening understanding and exploring avenues for academic research and collaboration in the realm of AI and cultural heritage. By grappling with these complex issues and identifying research priorities, the academic community can play a pivotal role in shaping the responsible and equitable integration of AI technologies into cultural heritage practices.

Bibliography

- Bastian, J. (2023). *Archiving Cultures: Heritage, community and the making of records and memory*. Routledge. <https://doi.org/10.4324/9781003091813>
- Bowman, S. L. (2014). Educational live action role-playing games: A secondary literature review. *The Wyrld Con Companion Book*, 3, 112-131.
- Buolamwini, J., & Gebru, T. (2018, January). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Conference on Fairness, Accountability, and Transparency. Proceedings of Machine Learning Research*, 81, 1-15. PMLR.
- Ciecko, B. (2017). Examining the Impact of Artificial Intelligence on Museums. *MW17: MW 2017*.
- [Cushing, A.L.](#), & [Osti, G.](#) (2023). "So how do we balance all of these needs?": how the concept of AI technology impacts digital archival expertise. *Journal of Documentation*, 79(7), 12-29. <https://doi.org/10.1108/JD-08-2022-0170>

- Foka, A., Eklund, L., Løvlie, A. S., & Griffin, G. (2023). Critically assessing AI/ML for cultural heritage: potentials and challenges. *Handbook of Critical Studies of Artificial Intelligence*, 815-825.
- Gaia, G., Boiano, S., & Borda, A. (2019). Engaging Museum Visitors with AI: The Case of Chatbots. In T. Giannini & J. P. Bowen (Eds.), *Museums and Digital Culture* (pp. 309–329). Springer International Publishing. https://doi.org/10.1007/978-3-319-97457-6_15
- Griffin, G., Wennerström, E., & Foka, A. (2023). AI and Swedish Heritage Organisations: Challenges and opportunities. *AI & SOCIETY*, 1–14.
- Ioannides, M., Georgopoulos, A., & Scherer, M. (2005). Standards in cultural heritage: The missing grammar for the digital documentation of the past. *International Conference for Documentation of Cultural Heritage*. <https://ktisis.cut.ac.cy/handle/20.500.14279/18707>
- Korka, E., Emmanouloudis, D., Kravari, K., Kokkinos, N., & Dimitriadi, K. (2022). The Protection of Natural and Cultural Heritage Monuments, Museums and Archives from Risks: Bridging Artificial Intelligence, Risk Assessment and Stakeholders. In A. Moropoulou, A. Georgopoulos, A. Doulamis, M. Ioannides, & A. Ronchi (Eds.), *Transdisciplinary Multispectral Modelling and Cooperation for the Preservation of Cultural Heritage*. TMM_CH 2021. *Communications in Computer and Information Science*, vol 1574, pp. 17-28. Springer, Cham. https://doi.org/10.1007/978-3-031-20253-7_2
- Manžuch, Z. (2017). Ethical issues in digitization of cultural heritage. *Journal of Contemporary Archival Studies*, 4(2), 4.
- Not, E., & Petrelli, D. (2019). Empowering cultural heritage professionals with tools for authoring and deploying personalised visitor experiences. *User Modeling and User-Adapted Interaction*, 29, 67-120. <https://doi.org/10.1007/s11257-019-09224-9>
- Villaespesa, E., & Crider, S. (2021). A critical comparison analysis between human and machine-generated tags for the Metropolitan Museum of Art's collection. *Journal of Documentation*, 77(4), 946-964.
- Waern, A., Rajkowska, P., Johansson, K. B., Bac, J., Spence, J., & Løvlie, A. S. (2020). Sensitizing scenarios: Sensitizing designer teams to theory. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13.

Workshop on Knowledge Elicitation: Data Collection and Knowledge Modeling for Human-Aware Systems

Authors: Andreas Brännström, Umeå University; Harko Verhagen, Stockholm University; Joel Wester, Aalborg University; Katie Winkle, Uppsala University; Niels van Berkel, Aalborg University; Irene Kaklopoulou, Umeå University; Derya Akbaba, Linköping University

Introduction

Knowledge elicitation comprises methods and techniques for extracting knowledge from domain experts to create structured descriptions of a domain. Knowledge elicitation typically involves analyzing qualitative data, such as interview transcripts, to identify concepts and relations of the domain, aiming to develop common vocabulary and semantics to enable representing, sharing and re-using the elicited knowledge across human-system boundaries. In the field of Artificial Intelligence (AI), such knowledge models, typically implemented in the form of ontologies, can be integrated into knowledge-based intelligent systems to enable context-aware reasoning and decision-making capabilities. A main challenge in this field regards capturing the important information to sufficiently represent the real-world domain. The social sciences and humanities often deal with complex and nuanced information of human domains, enabling insights that capture multi-dimensional views on, e.g., human, societal, socio-economic and political considerations of the world. Incorporating knowledge elicitation tools and methods in social sciences and humanities, which have well-established methods for qualitative data collection and analysis, can support the development of knowledge formats for understanding and sharing knowledge of, e.g., social structures, cultural practices, and human behaviors across disciplines and human-system boundaries to develop human-aware intelligent systems.

When developing knowledge models for intelligent systems aimed at understanding and interacting with humans, key questions arise:

- How, where, and from whom do we gather relevant information?
- When is theoretical saturation reached and when can data collection end?
- When is a theory about a human domain complete?
- How do we evaluate if a theory sufficiently represents the human domain?

These questions were the focal points of the Workshop on Knowledge elicitation: Data Collection and Knowledge Modeling for Human-Aware Systems at AI4HS 2023, Malmö, Sweden. Throughout the workshop, we discussed and presented various approaches to capturing and modeling knowledge, such as developing ontologies through qualitative data analysis; using human-robot teams to learn person-adapted robot behaviors; employing participant voting techniques for consensus-building in data; and exploring a feminist framework for knowledge elicitation. Our aspiration is for this workshop to serve as a starting point for forthcoming events dedicated to exploring knowledge elicitation methods and techniques for human-aware systems.

Understanding Social Challenges in Autism: Human Knowledge to Machine Descriptions

Andreas Brännström, Umeå University

This presentation serves as an introductory exploration into the process of knowledge elicitation, particularly when working with qualitative methods to understand and structure human knowledge into machine readable vocabularies. Throughout the presentation, we follow a case study focusing on understanding the social challenges faced by individuals with autism. We walk through each step of the knowledge elicitation process, starting with the collection of data through interviews with autism experts, followed by the intermediate steps of qualitative coding and modeling, where the collected insights are organized and structured into a framework consisting of concepts from the domain and their interrelations. This systematic approach leads us to the development of computational models capable of reasoning about social challenges in autism. Implemented using computational methods, such as Web Ontology Language (OWL) and Semantic Web Rule Language (SWRL), the elicited knowledge is reflected in a vocabulary readable by computer systems. The final model equips a system with the ability to reason about context specific social challenges to provide behavioral support, tailored to the unique needs of an individual (Brännström et al., 2024). By presenting practical methodologies and theoretical insights, this presentation aims to offer an overview of the process from data collection to computational modeling, highlighting the potential of capturing human knowledge in a precise format that can be shared across human and system boundaries. We further look at some general challenges in knowledge elicitation, such as the concepts of theoretical saturation, referring to the point in data collection and analysis where no new concepts are emerging from the data, and minimal commitment, an essential principle in knowledge representation and engineering for capturing the relevant elements of a domain. In illustrating how knowledge elicitation unfolds by employing social science methodologies, such as Grounded Theory, we highlight the vast potential for interdisciplinary collaborations across various stages of the process aimed at capturing domain knowledge within computational models.

References

Brännström, A., & Nieves, J. C. (2024) Towards Control in Agents for Human Behavior Change: An Autism Case. *Journal of Intelligent & Fuzzy Systems* (forthcoming).

Data Collection and Knowledge Modeling for Human-Modeling

Harko Verhagen, Stockholm University

As part of the AI4HS 2023 workshop on Knowledge Elicitation: Data Collection and Knowledge Modeling for Human-Aware Systems, I focused on a slightly different but highly related topic as can be seen from the title as there is quite some overlap concerning how data can be collected to better understand humans in general. This includes the following topics from the workshop description:

- analyzing qualitative data, such as interview transcripts,
- development knowledge-based intelligent systems to enable context-aware reasoning and decision-making capabilities,
- capturing the important information that sufficiently represents the real-world domain,
- human, societal, socio-economic and political considerations of the world,
- qualitative data collection and analysis, and
- knowledge formats for understanding.

One simulation study of interest concerns decision-making in the context of common-pool resource problems – situations where the use or non-use of a resource by a large group of people may have long term effects for the resource that are difficult to foresee given the small contribution of individual use to the availability of the resource in the short run. This includes the so-called “tragedy of the commons”, which predicts the depletion of resources as being unavoidable in the long-run as individuals act out of self-interest and thus maximize their personal consumption, and the systems studied by Elinor Ostrom that show that in relatively small communities, the emergence of norms and trust counteracts the short-term thinking on the individual level by premiering the system-level thinking which individuals trust is shared within the community. How behaviour may switch from one mode to the other is amongst others studied in behavioural economics experiments. Typical experiments consist of a shared resource from which individual participants in the experiment may consume at most a certain number in one decision round (with the number multiplied by the number of participants will deplete the resource more in one round of choices than the restocking of the resource). By varying experimental settings, such as allowing communication or not or even deceiving, the effect of such variation on the overall outcome of the experiments can be studied – in essence looking at the appearance or disappearance of said norms and trust. Of course, the outcome also depends on individual variation of traits e.g., altruistic versus egoistic as a baseline makes a difference. Varying these in experimental settings is not easy, which creates a need for artificial participants in which there is control even at the level of traits to investigate in more detail.

To develop artificial experiment participants, i.e., AI agents, we need to have some models on decision-making in the experiments. These models then combine the outcomes of human-only experiments and theories on decision-making and traits to investigate the possible combinations of these. Apart from collecting data in lab settings such as in behavioural economics experiments, one can also use observation studies of behaviour in natural settings. Both analyse the choice from the outside – behaviouristic. On the other side of the spectrum, we can find data collection that tries to look on the inside (e.g. interviews and questionnaires). Triangulating the two general types would allow for a more complete picture of thought and action. Figuring out what data is needed

to model humans for human aware AI-systems needs to start from the analysis of what is at play in the situation that is of interest, what factors may play a role etc.

Concluding these general remarks based on social and behavioural science research, some recent developments reverse the relationship between AI and the study of humans as described above. For instance, some social scientists are testing the use of LLM models as artificial (or synthetic if you wish) participants in a survey study. For instance, by prompting an LLM to “act as” a right- or left-wing oriented person, the LLM would make different choices with regards to policy items that matter such as the choice to wear a facemask or not in a pandemic situation. Another example is putting together a set of agents, each with their own LLM connection, in a small society to see how collaboration between these agents would emerge, with all agents prompted to have certain demographics and personal traits. Even if the base for the LLM is a large dataset, one may wonder if studying human choice by connecting these models to hypothetical or simulated situations is a step forward or backward.

A Qualitative Approach to Modeling the Critical Friend

Joel Wester, Aalborg University

In people's interactions with interactive AI systems (e.g., Alexa and Siri), these systems often present overly conforming behaviour as they lack any understanding of what behaving like a 'good friend' means and looks like. This lack of understanding directly relate to the focus of UNESCO's report titled "I'd Blush If I Could", spinning on how Siri, a female-gendered voice assistant, responds to users by saying "I don't know how to respond to that" when they say, "Hey Siri, you're a bi****" (West et al., 2019). We draw inspiration from this work, as well as prior work on human-AI friendship (e.g., Brandtzaeg et al., 2022), exploring how a 'good AI friend' can respond to users more appropriately. This talk centers around our exploration of this idea by using qualitative methods to collect, analyze and model data, consequently establishing a formal understanding of the Critical Friend conceptualized as a chatbot (Wester et al., 2023). By using qualitative methods, we can break down the behavioural cues that constitute Critical Friend communication. Ultimately, this process of conceptually breaking down and rearranging the Critical Friend constituents allowed us to use formal methods to transform these constituents into a machine-readable format, theoretically enabling an implementation of the Critical Friend agent (e.g., as a chatbot).

While our understanding of the Critical Friend is formally precise, an implementation of the Critical Friend as a chatbot requires careful considerations of the intended context. More specifically, relevant contextual factors to consider include end-user (e.g., young versus adults), the domain (e.g., health and well-being), and the relation of the model to other social concepts such as empathy (Brännström et al., 2024). This requires careful adjustments of the Critical Friend. While AI systems are concurrently and increasingly deployed in real-life situations, our theoretical understanding of the Critical Friend is in its infancy. However, the Critical Friend is a promising step towards a more careful and innovative approach to how interactive AI systems can interact with and appropriately respond to people in real-life situations.

References

- Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). "My AI Friend: How Users of a Social Chatbot Understand Their Human-AI Friendship". *Human Communication Research*, 48(3), 404-429.
- Brännström, A., Wester, J., & Nieves, J. C. (2024). A Formal Understanding of Computational Empathy in Interactive Agents. *Cognitive Systems Research*, 85, 101203.
- West, M., Kraut, R., & Ei Chew, H. (2019) "'I'd Blush If I Could': Closing Gender Divides in Digital Skills Through Education". <https://unesdoc.unesco.org/ark:/48223/pf0000367416>
- Wester, J., Brännström, A., Nieves, J. C., & van Berkel, N. (2023). "You've Got a Friend in Me": A Formal Understanding of the Critical Friend Agent. In *Proceedings of the 11th International Conference on Human-Agent Interaction* (pp. 443-445).

Participatory Design and Automation as Knowledge Elicitation

Katie Winkle, Uppsala University

Participatory design is fundamentally concerned with involving the people who will use and/or will be affected by a technology in the design of that technology, with a focus on mutual learning between participants who typically represent either domain experts (users) or technology experts (designers) (Simonsen & Robertson, 2013). For the design and application of robots, it has been argued that participatory design offers a way to (i) tackle technological determinism, better recognising the sociotechnical nature of robot deployment (Šabanović, 2010) and (ii) increase diversity in who gets to shape robot design and application, aligned to ethical and feminist robot design approaches (Winkle et al., 2023). Where participatory design is often limited to high level and abstract ideas about what e.g. a robot could or should do, participatory automation looks to go further and include domain experts in the “low level” programming/implementation of (autonomous) robot behaviours, e.g. by having them “teach” the robot how to behave, in-situ (Winkle et al., 2021). This has been posited as a particularly pertinent approach for programming social and/or socially assistive robot behaviours. Consider e.g. a robot fitness coach; knowing when to tell the user they need to work harder, versus when to praise them for working so hard, or when to simultaneously sympathise-with-but-further-encourage them because they’re struggling with an exercise.

Whilst not described as so in the original article, I argue that such approaches represent knowledge elicitation, as they attempt to make the capture of tacit, situational, and otherwise generally intangible expertise that drives human-human social interaction. To date, such approaches have been of interest mostly for the purposes of developing social robots capable of imitating such social intelligence. What new insights might we derive if we stop and reconsider the knowledge elicitation process itself as being valuable, rather than a means to end?

References

- Šabanović, S. (2010). Robots in Society, Society in Robots: Mutual Shaping of Society and Technology as a Framework for Social Robot Design. *International Journal of Social Robotics*, 2(4), 439-450.
- Simonsen, J., & Robertson, T. (Eds.). (2013). *Routledge International Handbook of Participatory Design* (Vol. 711). New York: Routledge.
- Winkle, K., McMillan, D., Arnelid, M., Harrison, K., Balaam, M., Johnson, E., & Leite, I. (2023, March). Feminist Human-Robot Interaction: Disentangling Power, Principles and Practice for Better, More Ethical HRI. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 72-82).
- Winkle, K., Senft, E., & Lemaignan, S. (2021). LEADOR: A Method for End-to-End Participatory Design of Autonomous Social Robots. *Frontiers in Robotics and AI*, 8, 704119.

Longitudinal Knowledge Elicitation Using Experience Sampling

Niels van Berkel, Aalborg University

Knowledge elicitation is the process of collecting structured insights into a domain or topic. Such structured insights can be used to build ontologies that support contextual reasoning and decision-making. While a myriad of knowledge elicitation techniques exists, including surveys and observations (Burge, 2001), these methods are not without their limitations. In surveys, for example, study participants are asked to reflect on multiple past events. This can introduce recall bias, negatively affecting data quality, and subsequently reducing the quality of outcomes produced by support systems trained on the elicited data.

This talk outlines the use of the experience sampling method, also known as ecological momentary assessment, for the purpose of knowledge elicitation. Using experience sampling, study participants are instructed to answer short surveys over a longer period of time, reflecting on the current moment rather than recalling past events (Csikszentmihalyi and Larson, 1987). As such, data consists of repeated and in-the-moment reflections. While experience sampling was originally introduced to obtain structured insights into how people spend their time, the method has since been used to answer a wide variety of research questions. The use of smartphones for data collection further provides the additional opportunity to collect contextual insights, such as location data or recent smartphone activity, enhancing the detail of collected insights (van Berkel et al., 2017). This method's disadvantages include the relatively high participant burden compared to traditional surveys, which often results in missing data. Due to the survey's high frequency, questionnaires are also limited in the number of questions that can be asked at a time.

Given the importance of data quality and quantity in knowledge elicitation, experience sampling provides an opportunity to collect contextually richer and more frequent data. Contextual data, collected in conjunction with manually annotated data can enhance the knowledge elicitation process. Simultaneously, participant fatigue and dropout can prove challenging to maintain consistent data quality. Interdisciplinary collaboration can support the collection and annotation of high-quality data necessary for reliable knowledge elicitation.

References

- Burge, J. E. (2001). Knowledge Elicitation Tool Classification. Artificial Intelligence Research Group, Worcester Polytechnic Institute.
- Csikszentmihalyi, M. and Larson, R. (1987). Validity and Reliability of the Experience-Sampling Method. *The Journal of Nervous and Mental Disease*, 175(9):526–536.
- van Berkel, N., Ferreira, D., and Kostakos, V. (2017). The Experience Sampling Method on Mobile Devices. *ACM Computing Surveys*, 50(6):1–40.

A Speculative Feminist Knowledge Elicitation System

Irene Kaklopoulou, Umeå University; Derya Akbaba, Linköping University

Echoing the CFP to the workshop, knowledge elicitation systems are important for utilizing expert domain knowledge in human-machine systems. These systems acknowledge the role that expert knowledge plays in structuring and utilizing information effectively. And yet, there remains the issue of how to “captur[e] the important information that sufficiently represents the real-world domain.” Central to this issue is the characterization of knowledge beyond the systems of data collection and knowledge modeling. We turn to feminist epistemology to approach the notion of knowledge from a critical perspective that challenges normative representational ways of knowing. Feminist theories support generative methods that influence the futuring of knowledge-driven AI systems.

Epistemology is the study of how we know what we know in the world, or how knowledge is created and represented. Stated differently, by understanding epistemology, we can better understand knowledge. In this talk, we put forward a speculative feminist knowledge elicitation system to challenge normative assumptions about knowledge. Through feminist epistemology, we consider knowledge from a feminist lens, adding a new, context-sensitive perspective to knowledge elicitation systems: a re-formulation of content, context, and ethics.

We speculate on a feminist knowledge elicitation system using our collective experiences working with feminist epistemology in human-computer interaction (HCI), science and technology studies (STS), and data visualization. Core to feminist epistemology is the ways in which power inequities affect how knowledge is expressed, can be known, analyzed, categorized, and represented. We present feminist epistemology through three overarching ideas: entanglements; situatedness; and embodiment.

In ubiquitous computing, Dourish (2004) argues that context is entangled, inseparable from activity, and thus, demonstrates how reconceptualizing context could affect the design of ubiquitous systems. In data visualization, Lin et al. (2022) apply these concepts to visual analysis platforms of domain experts, establishing design guidelines that focus on visually and textually recording the entanglements of knowledge production through eliciting both what the experts know along with how they know it. Lin et al. (2022) also tie together the entanglements with situatedness, discussing how identity and experience will result in different partial perspectives of knowledge being externalized. Knowledge is also embodied and contains felt aspects. We speculate what it may mean to value multiple forms of knowledge that derive from living, feeling bodies in the world. In the field of HCI, the emergence of body-sensitive, somaesthetic qualities in design research places the transformative becoming of the body at the center of knowledge production methodology, influencing ethical sensibilities among designers and practitioners (Ståhl et al., 2022; Popova et al., 2022; Garrett et al., 2023). We only offer a handful of examples in this abstract, but there is a rich history that challenges the representational perspective of knowledge in HCI and STS that could only enrich the ways in which we design and build knowledge elicitation systems.

During our talk, we cover core feminist principles, as mentioned above, and engage workshop participants with a short series of reflective prompts. These prompts follow the tradition of an implosion workshop (Dumit, 2014), which employs consecutive questions to tease and show up the technological, epistemological, material, historical, social, political, organic, ethical and numerous other threads that make up our worldviews. Some example questions are: Who is producing the knowledge? What aspects of their identity may or may not impact what they know? What type of knowledge counts; what knowledge

is excluded and whose knowledge is excluded? What is the temporality of the knowledge, or how will time change the meaning of what has been captured?

References

- Dourish, P. (2004). What We Talk About When We Talk About Context. *Personal and Ubiquitous Computing*, 8, 19-30.
- Dumit, J. (2014). Writing the Implosion: Teaching the World One Thing at a Time. *Cultural Anthropology*, 29(2), 344-362.
- Garrett, R., Popova, K., Núñez-Pacheco, C., Asgeirsdottir, T., Lampinen, A., & Höök, K. (2023, April). Felt Ethics: Cultivating Ethical Sensibility in Design Practice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-15).
- Lin, H., Akbaba, D., Meyer, M., & Lex, A. (2022). Data Hunches: Incorporating Personal Knowledge Into Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 29(1), 504-514.
- Popova, K., Garrett, R., Núñez-Pacheco, C., Lampinen, A., & Höök, K. (2022, April). Vulnerability as an Ethical Stance in Soma Design Processes. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).
- Ståhl, A., Balaam, M., Comber, R., Sanches, P., & Höök, K. (2022, April). Making New Worlds—Transformative Becomings with Soma Design. In *Proceedings of the 2022 Chi Conference on Human Factors in Computing Systems* (pp. 1-17).