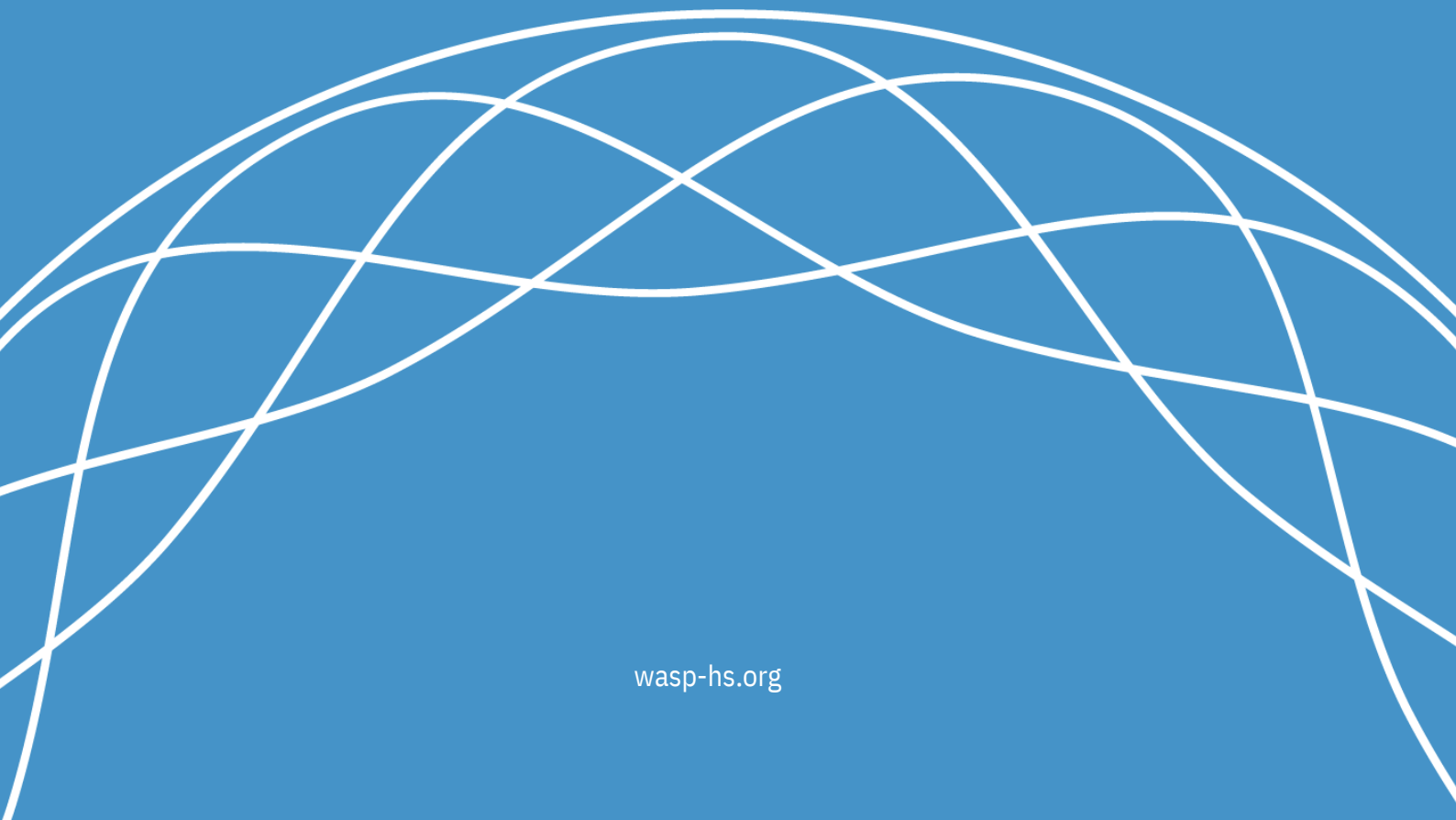
WASP—HS

COMMUNITY REFERENCE MEETING: LIFE IN THE DIGITAL WORLD

REPORT

August 2021



wasp-hs.org

Introduction

Artificial Intelligence (AI) and the online world are connected in ways that exceed any other application of AI. The use of AI is transforming the internet by predicting user actions and preferences by determining what we see and how others see us by enabling, or disabling, our ability to connect, participate and be included.
Already now, we are delegating many tasks to machines, from the mundane, e.g. which film to watch tonight, booking a taxi or ordering a pizza, to those tasks that can deeply affect our lives, such as determining whether you get credit or a subsidy, prioritization of access to services, or supporting diagnostic, or autonomously drive cars.

To understand how AI can benefit or harm our interactions and our society, we must first have a clear understanding of what AI is and what it is not. In a sense, AI is neither Artificial nor Intelligence. It is not artificial because it is built by and builds upon human expertise, relations and structures. It reflects not only the bias of those that collect and manage the data, or design the algorithms, but also the bias hidden behind each decision or action we all take. It augments our capabilities and our abilities to deal with complex problems, but it also augments the power structures and the distance between those that own it and those that are owned by it. And it is not intelligent because despite of its pattern identification and prediction, AI lacks the ability to understand the meaning of those patterns and how these relate to the context in which they are created, observed and applied. As intelligent technologies create new venues for virtual social interaction and influence, it is imperative to understand the power relations behind the use and development of AI, and how societal and ethical principles are reflected in the design of technology.

In this Community Reference Meeting, WASP-HS (the Wallenberg Artificial Intelligence Autonomous Systems and Software Program – Humanities and Society) brings together Swedish researchers, practitioners and other interested parties, to discuss how AI is shaping our online experience and our interactions with others. Following an introductory note on AI, Games and Virtual Worlds, by Professor Julian Togelius, New York University, participants discussed how the development and use of AI in the design of virtual spaces and our interactions with and within the online world. The discussions focused on four different areas: education, personalization, policy-making, and interaction. Participants included researchers from Swedish universities, industry, and national and regional governments, as well as the general public, NGOs and international organizations.

A research roadmap to the development and use of online interactions mediated by AI needs to be grounded on the fundamental democratic principles at the core of Swedish society and political tradition, highlighting studies into:

- How power relationships between people inevitably affect the shaping of technology. Analyzing human-computer interaction through the lenses of different altered relationships to technology can reveal more nuanced connections between the designers and users of autonomous systems.
- How to give users control of what data they share and with whom. Users typically engage with several subcultures and collectives at the same time, and their willingness to share identity markers depends on the context.
- Addressing ethical and legal issues of AI systems in practice is central to a down-to-earth understanding of AI's potential and risks in higher education. A continuous dialogue between multiple educational stakeholders remains critical for the responsible configuration of just and caring data-informed practices in higher education.
- Supporting policy makers and helping them understand how technology affects and is affected by social behaviour. Tools that visualize behaviour need to include constructs to model social behaviour, in ways that support understanding of the effects of policies in particular in times of crises, and complex social change.

WASP-HS Community Reference Meetings (CRMs)

CRMs are aimed at helping public and private organizations in Sweden with challenges and questions regarding their interests, as well as developments within WASP-HS. This is done to identify opportunities for collaboration between different sectors.

Marketing in Virtual Worlds

Main Takeaways

- Build and retain relationships in digital and physical spaces - this is a challenge. Understand how different actors define and value relationships, and what implicit obligations exist.
- Understand the effectiveness and impact of digital advertising. The aim is typically to disseminate information, build customer relationships, etc., but in practice we optimize for clicks because it is easy to measure. A balance between automated personalization and human interaction in marketing and customer relationships is needed.
- Assess and manage the risk for discrimination and reinforcement of stereotypes when using personalization and pseudo-segmentation, i.e., the practice of targeting media channels associated with a particular demographic group.

Marketing and communication are central activities for businesses and public bodies. During the last 10 years, digital channels have grown in importance and are now the primary route for many information flows. This trend is expected to continue, and new forms of marketing will arise in online games and other immersive media.

AI-driven communication in immersive media environments is a new but rapidly growing field of research. A central challenge is the lack of diversity in data sets. Since available classifiers have been developed for modern media, they underperform on historic materials [1]. Maximizing societal benefit while minimizing risks requires one to tackle a diverse set of complex tasks, which in turn calls for a multidisciplinary approach that includes humanities and social sciences. A central challenge is that there are subtle but important differences in the terminology used in different fields, which can act as stumbling blocks for collaborative work.

Every day, there are new technical possibilities to influence user attitudes and behaviours, for example, through personalized and contextual content. Some benefits of this are that the user interface can be adapted to meet the user's needs and to filter out irrelevant information. On the negative side, the technology can have a negative impact on the user's privacy, and interfaces can be outright misleading, for instance, to make it more difficult to cancel a subscription.

From an ethical perspective, the interlinks between personalization and commercial interests need to be considered and it is also important to consider whether government organizations and other public bodies should use targeted and personalised communication. There can be valid reasons for official authorities to personalise services, for example, to simplify application processes or provide relevant information. The law allows this but there is a general reluctance to the approach, as there are serious concerns linked to privacy and fairness. Another major issue is bias in data. Research is needed to analyse datasets with respect to bias, and develop tools to address imbalances, for example, resampling or synthesizing data. The idea of personalization builds on the perceived co-produced relationships between organizations and consumers, but actuality it is designed to downplay the relationship between the two. What would happen if organisations that used chatbots and similar avatars had to provide a designated human point of contact on demand that is responsible for the organisation's side of the relationship? We can agree that the need for human interaction varies greatly with the situation. Whether or not the system is managed by a bot or a real human being, it is important that the user knows which is the case, and can set their expectations of the system accordingly.

Another topic is pseudo-segmentation. This is similar to personalisation, but rather than adapting the messaging based on personal markers, it is adapted to context markers that are predictive of a particular demographic group. For example, the possibility to direct advertisements away from certain groups has led to some house-for-sale advertisements not being shown to accounts which have searched for hair straighteners. Hair straightener products here become a stand-in for a racial category, and this technology then allows for a version of digital 'redlining', a practice which is illegal.

In a more positive vein, some media companies use personalisation to deliver relevant news to their audiences based on the geographic location of their readers. However, this involves editorial decisions, and some media brands within the same media company take a restrictive stance to personalisation of news. It is felt that it is important to inform everyone in the same way, with marketing being an exception where personalisation is sometimes permissible.

References:

[1] See the related WASP project on AI and cultural heritage collections
https://wasp-hs.org/projects/att-kvantifiera-kultur-en-studie-av-ai-och-kulturarvssamlingar/

From Social Simulations to Interactive Games for Policy Makers

Main Takeaways

- Tools that visualize social behaviour are currently very limited. They do not include constructs to model social rules (like norms) themselves.
- Models that properly simulate social behaviour need both violating and socially compliant interactions to be calibrated.
- Social simulations are not part of the tool box of policy makers and decision makers.
- Models based on existing data are not suitable to predict how people react in a crisis when the environment changes radically. This can make it hard for stakeholders to find out how to make decisions that secure the desired outcome.
- Interactivity of the simulation platforms is very limited. Being able to use a simulation as an interactive game-like environment gives stakeholders the opportunity to actually immerse in a situation and see the effect of their policies and interactions.

The collaboration of all parties involved in a crisis situation such as the COVID crisis or the refugee problems is crucial to get optimal results for those who are primarily affected. However, often parties only act from their own perspective and thus actions of different agencies might oppose each other without anyone noticing. E.g. stimulating refugees to do some sport, while organizing medical assistance at the same time. The 'orchestration' of the multiple factors involved makes for a complex puzzle, which can be studied with social simulations that can show the stakeholders how their policies affect other stakeholders and can sometimes be much better coordinated.

We take as examples the refugee and COVID crisis as the cases of how simulations may have a very tangible impact on society.

In a simulation of logistics for refugees, factors like housing and daily activities can be very differently perceived by the refugees and the different NGO's involved in refugee support. In the Netherlands, more than 40 organizations have been managing the challenging issues around refugee logistics. Some social simulations were used in which stakeholders could experiment with different policies for their own work, and explore the dependencies and the effects of these on the refugees.

Even though resulting insights are already useful, the social simulation platforms have very little interactivity so far. In many cases some actual board games are used to provide more interactivity. However, it would be very helpful if these board games could be played virtually and be connected to the simulations.

Social simulations of the COVID crisis made with the ASSOCC framework showed (see Figure 1) that most of the infections occurred at home, which was somewhat unexpected but later confirmed by health authorities in different countries. Simulations made with the ASSOCC framework also support the analysis of the ineffectiveness of curfews [1]. Curfews were assumed to lead to fewer private contacts at homes of family and friends. However, simulations that modelled the needs of people showed that many people just meet earlier in the day or stayed over to circumvent the restrictions. Again, these insights were later on confirmed by observations in practice. The main advantage of behaviour-based simulations is that alternative behaviours can be modelled and will appear when restrictions are introduced. This can give good insights in what changes a policy actually creates.

One of the most heard criticisms of using social simulations for policy making is that it is very hard to create good behavioural models for situations where the environment changes due to the introduction of a new policy. The advantage of using agent-based models for social simulations is that each scenario can be explained in terms of the types of decisions the individuals make and how they affect each other.

In general, people have different needs, and thus basing models on one combination of needs is not very appropriate. However, the use of massive amounts of data that is available nowadays can provide a good insight in the different types of needs of different parts of a population.

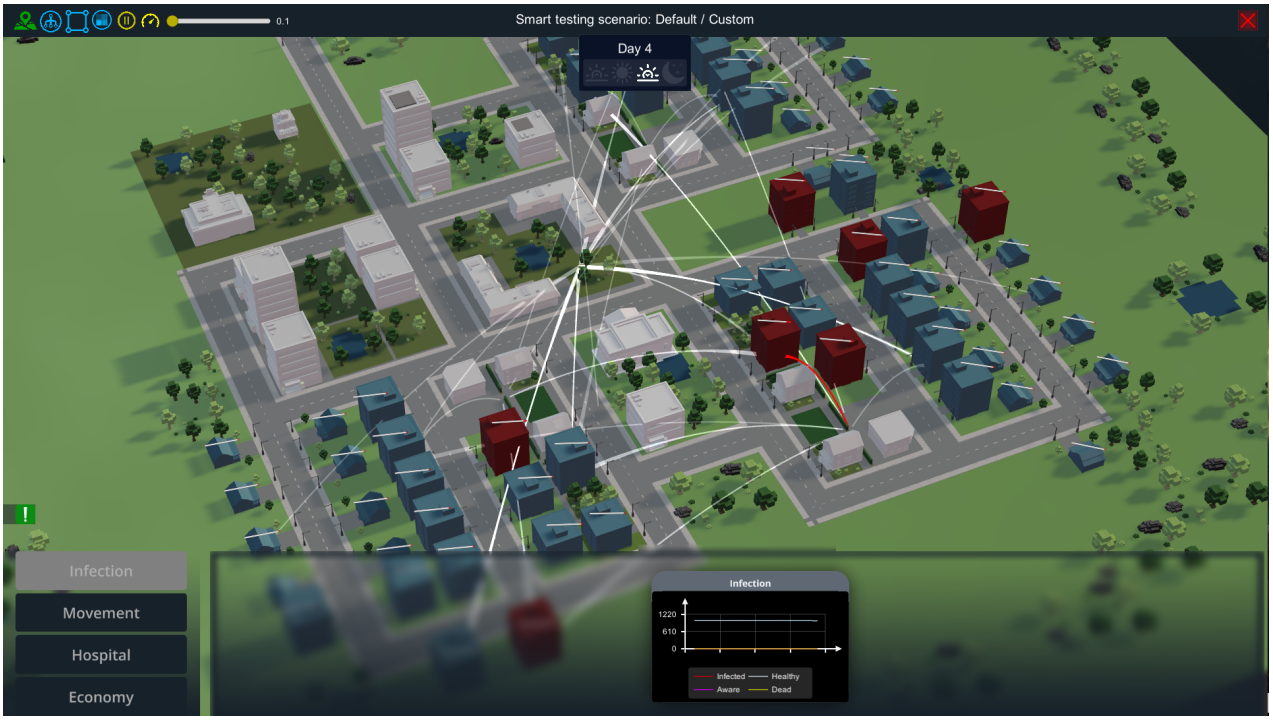


Figure 1. Screenshot of social simulations of the COVID crisis made with the ASSOCC framework

Social simulations are not meant to provide detailed numerical predictions for the future but to provide insights on what might happen and give an indication of how people might react.

Often, policies are meant to have effects in the long term, for instance, policies to prevent climate change, or create sustainable neighbourhoods, etc. One of the problems for making policies for these situations is that many effects are only visible years afterwards. For example, when designing a new town, planners have to decide where the main route or the bus stops should be put. Should the neighbourhood contain environmentally friendly buildings that are expensive or cheaper ones that will ensure people with a low income can afford to live there as well. What will the effect be on the demographics of the neighbourhood and the need for schools and medical facilities? Social simulations can be run to cover periods of dozens of years to study potential long-term effects.

Alternatives can be checked and again plausible scenarios can lead to new insights into what the best choices are. This process is interactive, and leads to better policies, than if this interactive interface is not available.

There are several questions that remain open. To begin with, what tools are required to model social interaction. Especially since the related theory is scarce, there is considerable room for development. The more interactive one wants to make the simulation, the easier it should be for policy makers to create scenarios and experiment with variations. For this to be possible the tools need an easy to use interface that is also understandable for domain experts. In order to create social simulations quickly there is a need for simulation patterns and standardized ways to combine these. That will speed up the creation of new applications considerably and also allow for early input of domain experts.

References:

[1] Dignum, F. (ed.), Social Simulation for a Crisis: Results and Lessons from Simulating the COVID-19 Crisis, Springer, 2021.

Ethics in AI-Driven Educational Practices

Main Takeaways

- A wide range of stakeholders needs to be involved in discussing AI in higher education. Starting with students, we need to include teachers, administration, IT department, university management, trade unions, and the EdTech industry to understand better how relations constituting AI-driven educational practices are configured and shaped.
- An ethically and socially responsible approach to AI-driven educational practices operating at the intersection of matters of justice and matters of care is most needed.
- In the present political and economic context, there is an urgency to contribute to inclusive human-centric AI-driven practices, value human autonomy, and cultivate trustworthy relations in higher education.
- The task ahead is one of contributing an AI Agenda for Swedish Education with a focus on ethics in practice and local legal considerations.

AI systems are fast-developing and intriguing technologies that promise to benefit student learning and facilitate educational processes. However, using student data and analytics raises a series of issues that require serious ethical and legal considerations.

With the remarkably increased digitalization and datafication of educational practices due to the forced distance education imposed by the ongoing pandemic, the risks associated with AI-driven practices are acutely highlighted. In Europe, the use of online proctoring systems enabling students to write exams from home has raised concerns about ensuring the integrity of online assessments while respecting students' fundamental rights to privacy, data protection, and non-discrimination [1]. Another example is the mobilization of thousands of students in the UK last year to protest against the final grades allocated by the Office of Qualifications and Examinations Regulation's (Ofqual) algorithm in the so-called Advanced-level [2] scandal [3]. These students expressed their discontent and anger against the predicted grades publicly. More specifically, the Ofqual's grading algorithm that downgraded nearly 40% of A-level assessments done by the teachers reflected systemic bias [4].

While the use of machine learning and deep-learning algorithms in educational learning management platforms can potentially bring efficiency and quality to our current educational practices, the above-highlighted examples remind us of the dilemmas entrenched in current socio-technical educational practices. Such instances also call for more ethically and socially responsible ways to design and deploy complex socio-technical arrangements.

The slow development of AI regulation for the education sector in Europe also adds another layer of complexity. It has so far provoked a plethora of ethical guidelines, frameworks, and institutional codes of practice that, while necessary, contribute to a cacophonic understanding of ethics in AI-driven practices. Moreover, such discourses often focus on what is ethical and unethical instead of providing designers, researchers, and practitioners with guidance on negotiating, co-constructing, and cultivating ethical practices and commitment.

In this vein, it becomes essential for higher-education institutions to discuss what we can learn from other countries already using AI in their practices to ensure informed conversations on students' integrity, equity, and accessibility. This also requires a better understanding of which kind of AI we refer to when discussing ethics, algorithmic bias, and legal considerations of AI in higher education. The ethical issues that arise with AI cannot be generalized for all AI systems because each system may carry specific ethical issues. In the same way, it is also essential to situate and distinguish better the groups of students addressed by AI in education. Far from being a homogenized stakeholder category, students represent diversity and groups of people with different needs and backgrounds. Likewise, the specific context in which AI-driven practices are deployed (e.g., the student assignment, the exam, the course module, the program, or the institution) also plays a significant role in understanding AI's risks and potential harm in higher education.

It is crucial in conversations with stakeholders to increase awareness of ethical practices in responsible ways as ethics is a moving target. Here, the university leadership and the management staff purchasing cutting-edge digital platforms are key to continuously ensuring conversations about legal aspects and responsibility, that is, before, during, and after the deployment of technologies. Much more knowledge is needed about the legal frameworks regarding personal data management and the intricacies of the whole data supply chain that significantly impact the deployment and use of AI systems, to the point of banning some of them in Sweden because of third-country data transfers.

To summarize, when we talk about AI ethics in education, we talk about involving multiple stakeholders at institutional/organizational levels, from the public and private sector, to nurture a down-to-earth dialogue on professional educational practices and student learning in Sweden. We also speak about specifying which AI services we refer to, which groups of people we have in mind or are omitting, and in which contexts are the practices targeted situated. Moreover, we need to talk more about values we care about in Swedish education. These values should be reflected in a competitive but sustainable, just, and caring education system.

References:

[1] Colonna, L. (Forthcoming). Legal Implications of Using AI as an Exam Invigilator, In 2020-2021 Nordic Yearbook - Law in the Era of Artificial Intelligence (eds. Liane Colonna and Stanley Greenstein) (Stockholm, The Swedish Law and Informatics Research Institute (IRI)).

[2] Advanced or A-level refers to the second of three standardized British examinations in a secondary school subject used as a qualification for university entrance (Merrian-Webster dictionary).

[3] Amoore, L. (2020). Why'Ditch the algorithm is the future of political protest. The Guardian, Wed, 19. https://www.theguardian.com/commentisfree/2020/aug/19/ditch-the-algorithm-generation-students-a-levels-politics

[4] Hao, K. (2020). The UK exam debacle reminds us that algorithms can't fix broken systems. MIT Tech Review. Retrieved September 1, 2020. https://www.technologyreview.com/2020/08/20/1007502/uk-exam-algorithm-cant-fix-broken-system/

Altered Relationships to Autonomous Systems: 'Downgrading' AI / 'Upgrading' People

Main Takeaways

- AI is often characterized as either an assistant to a human or a superior entity that is going to take over from humans in decision making. Yet, there is a much broader spectrum of human-AI relationships we that can think of, including correspondence, symbiosis, alterity, background, and a monster.
- No system is ever completely autonomous. Autonomy requires distribution of agency between humans and technology. With technologies such as autonomous cars and smart homes, the aim often seems to be to skew the balance toward technology, but that is not always desirable. We should work at bringing the balance of control to the human side of the equation.
- AI is diverse. It is not one thing, but many. AI may be a part of a technology an individual can "use" or a part of technological infrastructure that is much broader than an individual human being. We need to be specific and define what exactly are we talking about when we are talking about AI.

The design of new types of autonomous systems and making sense of already existing technologies can benefit from a better understanding of the relationships between humans and AI (or digital technologies more broadly). These are relationships of correspondence, symbiosis, alterity, technology as/in the background, and even encountering technology as a monster – part human, part other.

Inspired by the works of Don Ihde [1,2], Tim Ingold [3], Peter-Paul Verbeek [4], Jeffrey Bardzell [5], and others, these relationship categories define how technologies transform the bodies of their users. Correspondence relationships point to our bodies being adapted to and formed by technology, for instance, the *Pelvic Floor Chair* developed by Anna Ståhl & Madeline Balaam [6] (see Figure 2a,b).

a *b*

Figure 2a,b. The Pelvic Floor Chair by Anna Ståhl and Madeline Balaam

Symbiotic relationships refer to feelings of unity between the user and technology, while the relationships of alterity speak about technology appearing as 'the other', such as the *Aerial Robotic Choir* by Åsa & Carl Unander-Scharin [7] (see Figure 3).

Background relations reside in the periphery of human attention, such as the heating system of our home or a 'smart home' system. Monster relationships appear when technology appears to be alien to its user, such as the Singing Corset designed by Kelsey Cotton, Özgun Kilic, and colleagues.

The next question is whether novel views on technology can help us generate both better designs and new insights into autonomous systems – and how. We agree that, although different relationships with AI are constantly being constructed, there is a lack of explanatory models to conceptualize AI in its current shape, rather than in a shape that it could take in 30 years from now. Widespread narratives portraying AI as a superior being that can suppress humans in decision making does not correlate with the realities of today's AI.

Another important step in reconceptualizing AI requires understanding that it is not a single entity but a variety of different systems, which rely, in many cases, on nothing more than databases and statistics. "Downgrading" AI is important for abandoning the view on AI as a centre of the cultural order. One can say that it is more accurate to see AI as something that may occasionally become a useful part of a human enterprise, rather than an unconditionally praised technological achievement. Sustaining a realistic view on AI can help to dissolve the discursive inevitability surrounding AI and help us ask whether we actually need all the AI systems that are currently under development.

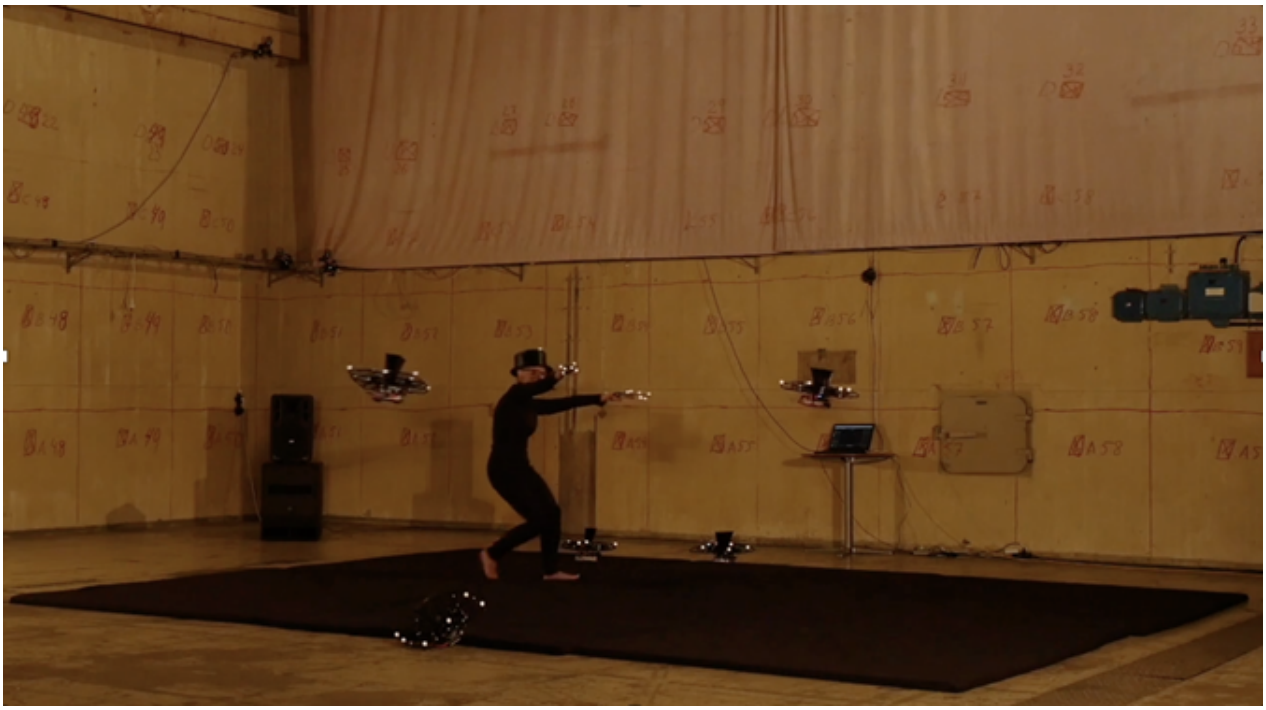


Figure 3. The Aerial Robotic Choir by Åsa and Carl Unander-Scharin

AI often exists as a part of broader technological infrastructures. As AI feeds on data, there is a risk that (most) people stop being the masters-of-technology and turn into the bodies to-be-harvested. AI is produced by humans, so the production and use of autonomous systems are inevitably accompanied by power relationships. Conceptualizing relations between humans and tech as altered, correspondence or symbiotic will reveal or rather hide power relationships between people embedded in technology. Combining these lenses will be productive for seeing the nuances of power relationships.

We need to replace our infatuation with technology with an existential approach to AI. That approach would be based on understanding that the practices of data collection existing for the sake of developing AI have become mundane parts of our lives. The everyday enmeshment of AI has to be explored in order to see how the technicity intersects with our originally human vulnerability and relationality.

References:

[1] Ihde, D. (1990). Technology and the Lifeworld: From Garden to Earth. Indiana University Press.

[2] Ihde, D. (1995). Postphenomenology: Essays in the Postmodern Context. Northwestern University Press.

[3] Ingold, T. (2021). Correspondences. Polity Press.

[4] Verbeek, P.-P. (2011). Moralizing Technology: Understanding and Designing the Morality of Things. University of Chicago Press.

[5] Bardzell, J. (2021). FFF Seminar: Chimeras of Boundless Glamour, Realities of Little Worth. Lecture at KTH. 02-05-2021.

[6] Balaam, M., Woytuk, N., Felice, M., Afsar O., Ståhl A., Søndergaard, M. (2020). INTIMATE TOUCH. Interactions, Vol. 27, No. 6

[7] Eriksson, S., Höök, K., Shusterman, R., Svanaes, D., Unander-Scharin, C. et al. (2020). Ethics in Movement - Shaping and Being Shaped in Human-Drone Interaction In: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing System

WASP—HS

The vision of the Wallenberg Artificial Intelligence, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) is to realize excellent research and develop competence on the opportunities and challenges of artificial intelligence and autonomous systems with a strong investment in research in humanities and social science.

The WASP-HS program is planned to run 2019 – 2028 and will form an independent and parallel program to WASP, The Wallenberg Artificial Intelligence, Autonomous Systems and Software Program.

Request or more details: contact@wasp-hs.org

Authors

Johanna Björklund, Department of Computing Science, Umeå University
Teresa Cerratto Pargman, Department of Computer and Systems Sciences, Stockholm University
Frank Dignum, Department of Computing Science, Umeå University
Virginia Dignum, Department of Computing Science, Umeå University
Clàudia Figueras, Department of Computer and Systems Sciences, Stockholm University
Emil Häglund, Department of Computer and Systems Sciences, Stockholm University
Kristina Höök, Division of Media Technology and Interaction Design, KTH Royal Institute of Technology
Airi Lampinen, Department of Computer and Systems Sciences, Stockholm University
Sara Leckner, Department of Computer Science and Media Technology, Malmö University
Cormac McGrath, Department of Education, Stockholm University
Julian Alfredo Mendez, Department of Computing Science, Umeå University
Cezara Pastrav, Department of Computing Science, Umeå University
Kristina Popova, Division of Media Technology and Interaction Design, KTH Royal Institute of Technology

How to cite this report
WASP-HS. Community Reference Meeting: Life in the Digital World. Report. August 2021.

Design and typesetting
Tatyana Sarayeva