WASP-HS

Research Projects 2019-2025

WASP-HS

Research Projects
2019-2025

Cover image (https://unsplash.com/license)

Published by:

WASP-HS
Umeå University
901 87 Umeå

ISBN 978-91-7267-507-0 (print)
ISBN 978-91-7267-508-7 (PDF)

Printed in Sweden by Media-Tryck, Lund University,
Lund, 2025

Content

Preface

Wallenberg AI, Autonomous Systems and Software Program – Humanity and Society (WASP-HS) is a national research program stimulating interdisciplinary knowledge about artificial intelligence and autonomous systems and their impact on human and social development.

WASP-HS it is the largest freestanding research program in the social sciences and humanities in Sweden to date, funded by Marianne and Marcus Wallenberg Foundation from 2019 to 2031. Through a national graduate school, international postdoctoral fellowships, tenure track positions, international guest professorships, conferences, and collaborations the program builds research capacity across the humanities and social sciences. Over 250 researchers from many academic disciplines based at more than 15 Swedish universities are currently affiliated to WASP-HS.

The program was set off with open, bottom-up, calls for research projects on the consequences of AI and autonomous systems. The projects funded through these calls ended 2004 and 2025. In many ways these projects and researchers define the foundation upon which WASP-HS continues to build its approach in second half. Taken as a whole, the projects testify to the diversity of research questions that the AI transformation provokes and to the importance and relevance of studying technological change from a multidisciplinary perspective.

This booklet provides a glimpse of the research projects that were funded during the start-up phase of WASP-HS. Projects appear in alphabetic order of reporting PIs, presenting a research result, its implication and relevant publications. We hope that these brief presentations will encourage interaction with WASP-HS research and researchers.

AI as a New Strategic Imperative

Cases and analyses of AI implementation projects revealed significant challenges, but clear effects and consequences of AI implementation remained uncertain. Various technical, organizational and business related challenges hindered companies and public organizations from fully capitalizing on the value of their AI implementations at this stage. While AI can streamline internal operations and reduce costs through e.g. automation, transforming AI analyses and augmented decision making into new revenue streams remained a challenge for many enterprises. Specific findings in an empirical study of AI in public procurement indicated a low level of AI maturity in the investigated governmental agencies. The perceived benefits of AI revolved around improved operational capabilities, potential for certain process efficiencies and the ability to enhance monitoring through AI.

Implications

Implications for research: The exemplified AI in Public Procurement study contributes empirical insights and knowledge regarding the perceived benefits of AI in public procurement and the challenges associated with its implementation. The study contributes to knowledge on how AI can contribute to the creation of public value through public procurement, thus adding knowledge to previous studies that mainly have focused on legal and ethical aspects.

Implications for management: The framework developed can assist practitioners in anticipating potential challenges that may arise during the implementation of AI. Some challenges are common in procurement, while others are specific to public procurement. In what specific procurement process the AI change starts is crucial, affecting other parts of the procurement

operations, and more broadly public administration and public value creation.

References

Andersson, P.E., Arbin, K. and Rosenqvist, C. (2025). Assessing the value of artificial intelligence (AI) in governmental public procurement, *Journal of Public Procurement* 25(1):120-139. doi.org/10.1108/JOPP-05-2024-0057.

Andersson, P., Cramner, I., Nadeem, H. and Rosenqvist, C. (2023). Implementing AI in source to contract operations: how procurement managers in a global organization make sense of AI opportunities and inhibitors, The IPSERA 2023 Annual Conference, Barcelona, Spain.

Research team

Per Andersson, Stockholm School of Economics
Christopher Rosenqvist, Stockholm School of Economics
Hadia Nadeem, Stockholm School of Economics
Katarina Arbin, SSE Institute for Research (SIR)

Contact

per.andersson@hhs.se

Ethics for Autonomous Systems and AI

The project explored the ethical, cognitive, and philosophical challenges of developing autonomous AI systems that act responsibly and align with human values. A key contribution lies in human-robot interaction, showing that emotional signaling and speech can help rebuild trust after a robot makes a mistake. The project demonstrated how ethical AI can be grounded in context-sensitive interaction between robot and human. Such interaction depends on honest signals that accurately reflect the robot 's capabilities and understanding. Another important result is that ethical theories can be computationally embedded and tested. Various ethical principles were evaluated in simulated robots, revealing both the promise and current limits of ethically responsive AI systems. A key finding is that computational complexity influences what ethical theories can practically be implemented in machines

Implications

By integrating philosophical ethics with cognitive science and robotics, the project shifts the AI ethics debate from abstract theory to concrete implementation. It demonstrates that frameworks like virtue ethics, utilitarianism, and deontology can be modeled computationally and tested in robotic systems. This approach lays the groundwork for machines capable of acting ethically in real-world scenarios, while also exposing where current technologies fall short. Crucially, the research challenges assumptions about "universal" AI ethics, stressing the need for context-aware, culturally sensitive design. Overall, the project offers a roadmap for responsible AI development—one that prioritizes human values, ethical reflection, and social engagement. It envisions a future where AI augments rather

than replaces human morality, contributing to systems that are intelligent, fair, and empathetic.

References

Balkenius, C. and Johansson, B. (2022). Almost Alive: Robots and Androids, *Frontiers in Human Dynamics*. doi: 10.3389/fhumd.2022.703879.

Brinck, I. and Balkenius, C. (2020). Mutual Recognition in Human-Robot Interaction: A Deflationary Account. *Philosophy & Technology*. doi.org.10.1007/s13347-018-0339-x.

Stenseke, J. (2024). On the computational complexity of ethics: moral tractability for minds and machines. *Artificial Intelligence Review* 57(4):105. doi.org/10.1007/s10462-024-10732-3.

Research team

Christian Balkenius, Lund University
Amandus Kranz, Lund University
Birger Johansson, Lund University
Frank Zenker, Lund University
Jakob Stenseke, Lund University
Trond A. Tjøstheim, Lund University
Ylva von Gerber, Lund University

Contact

christian.balkenius@lucs.lu.se

AI-Driven Contextual Advertising

We study how ads are assessed when positioned alongside news articles that evoke negative emotions in readers. We find that in general, negative emotion does not influence advertising evaluation. Contrary to industry claims, the perceived source credibility of the news site is not found to moderate the effects of negative content. However, on its own, the credibility of the news site improves ad evaluations. Furthermore, high applicability between article and ad can enhance ad recognition and produce a weak negative effect on attitudes towards ads and brands.

Implications

The findings show that news site credibility strongly influences how ads are received, and that negative news content does not harm ad evaluations on reputable sites. Overly broad blocklists may unnecessarily restrict ad placements, hurting revenue for quality journalism. Advertisers are advised to work with trustworthy publishers rather than avoiding negative content altogether. However, caution is still needed on platforms with less editorial control, like social media, where certain negative contexts can harm brand perception. Emerging AI tools offer more nuanced content understanding, enabling precise targeting. Combining AI with human oversight can improve brand safety without overly limiting ad opportunities.

Reference

Häglund, E. & Björklund, J. 2024. Should advertisers avoid negative news? Advertising effects of negative affect, news site credibility, and applicability between article and ad. In: Jeseo & Parajuli (Eds) *Marketing and AI: Shaping the Future Together*. Springer. doi.org/10.1007/978-3-031-76193-5_2.

Research team

Johanna Björklund, Umeå University
Adam Åbonde, Stockholm School of Economics
Emil Häglund, Umeå University
Igor Ryazanov, Umeå University
Jingwen Cai, Umeå University
Richard Wahlund, Stockholm School of Economics
Sara Leckner, Malmö University

Contact

johanna@cs.umu.se

AI in Motion

Vehicle movement (speed, position, trajectory) is a crucial form of implicit communication on the road. Autonomous Vehicles (AVs) must both understand these cues from others and produce movements that are legible to human road users. Basic movements like gaps, speed, position, indicating, and stopping, allow drivers to communicate. This 'communication through movement' is an important part both of the design of AVs, but also robots that need to collaborate with humans in share social spaces.

Implications

This work shows how self-driving cars (both Waymo and Tesla FSD) struggle with yielding, a fundamental social interaction on the road. This involves not just stopping/slowing, but communicating intent to other road users, who must interpret those signals and respond appropriately. Failures include not yielding when others offer the right-of-way, and not "going" when yielded to. This research demonstrates how we can design the movement of so called 'embodied AIs', and how robots more broadly can collaborate - through motion - with others in shared social spaces.

Reference

Brown, B., Laurier, E., & Vinkhuyzen, E. 2023. Designing Motion: Lessons for Self-driving and Robotic Motion from Human Traffic Interaction. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (CHI '23). ACM, 5:1- 21. doi.org/10.1145/3567555.

Research Team

Barry Brown

Contact

barry@dsv.su.se

Ethical and Legal Challenges with AI-Driven Practices in Higher Education

Taking an interdisciplinary approach, the project investigated ethical and legal challenges to contribute to just and caring AI-driven practices in higher education (HE). Across the studies, we identify that while AI and autonomous experts invest time and effort in developing AI systems, moved by the promises of more efficient workload management and improved assessment practices, the actual use may introduce tensions, leading to breakages disrupting assessment practices. We found that breakages in assessment practice illustrate a disconnect between grand technical visions and the complex nature of mediating assessment practice, as well as friction between professional, pedagogical, and relational values. This means that as AI becomes more integrated in university assessment, there is a growing need to carefully balance the drive for efficiency with the preservation of educational judgment and accountability.

Implications

University stakeholders need to gain agency and collectively discuss and reflect on the raison d'être of universities to actively negotiate the place of AI and autonomous systems in quintessential university practices such as assessment.

AI Regulation, suffers from key weaknesses, including ambiguity around high-risk AI applications in EdTech and insufficient focus on students and teachers. Additionally, challenges such as fragmented governance, limited oversight resources, and the global nature of AI-based EdTech raise concerns about the regulation's effectiveness.

Interdisciplinary research is essential to investigate the long-term effects of AI and automation on students' learning experiences and outcomes as well as teacher-student relationships in HE.

References

Figueras, C., et al. (2025). Promises and breakages of automated grading systems: a qualitative study in computer science education. *Education Inquiry*. doi.10.1080/20004508.2025.2464996.

Sporrong, E., McGrath, C. and Cerratto Pargman, T. (2025). Situating AI in assessment—an exploration of university teachers' valuing practices. *AI and Ethic*s 5:2381–2394. doi.org/10.1007/s43681-024-00558-8.

Colonna,L. (2022). Addressing the Responsibility Gap in Data Protection by Design: Towards a More Future-oriented, Relational, and Distributed Approach, *Tilburg Law Review* 27(1):1–21. doi.org/10.5334/tilr.274.

Research team

Teresa Cerratto Pargman, Stockholm University
Alexandra Farazouli, Stockholm University
Cecilia Magnussson-Sjöberg, Stockholm University
Cormac McGrath, Stockholm University
Elin Sporrong, Stockholm University
Jaakko Holmén, Stockholm University.
Johanna Velander, Linnaeus University.
Liane Colonna, Stockholm University

Contact

tessy@dsv.su.se

Machine Learning to Study Causality with Big Datasets – Towards Methods Yielding Valid Statistical Conclusions

Neural networks are revolutionizing science, and our research shows their potential in causal inference. Convolutional neural networks (CNNs), known for their success in image analysis, can also help estimate treatment effects in complex data. We use CNNs to model background factors when studying how early retirement affects health. By capturing time-structured covariates, CNNs improve estimation accuracy. We prove that CNNs can achieve optimal convergence rates, making them powerful tools for causal inference. A simulation study confirms their reliability, and we apply this method to Swedish population data, uncovering new insights into retirement and its effect on health.

Implications

Convolutional neural networks efficiently capture time-structured individual characteristics when estimating causal effects.

While naïve analyses seem to indicate that early retirement has negative effects on the risk of hospitalization, no such effects remain when carefully controlling for individual characteristics.

Reference

Ghasempour, M, Moosavi, N, & de Luna, X. 2024. Convolutional neural networks for valid and efficient causal inference. *Journal of Computational and Graphical Statistics* 33(2):714–723. doi.org/10.1080/10618600.2023.2257247.

Research team

Xavier de Luna, Umeå University
Filip Edström, Umeå University
Mohammad Ghasempour, Umeå University
Tetiana Gorbach, Umeå University

Contact

xavier.deluna@umu.se

Socially Aware Dialogue Management

A dialogue management system interprets user messages (i.e., intent recognition), plans for the next move (i.e., planning), and returns an appropriate response (i.e., responding). Socially aware dialogue systems aim to manage the dialogues by taking the social context into account. Thus, enabling more natural and human-like conversations. In this project, I introduce a system I call Budgie that started as a dialogue management system and later evolved into an interaction system that is capable of incorporating social context at all levels. Budgie is built on top of social practices: habitually performed activities that are mediated between the participating actors. We use LLMs to interpret and generate the actual natural language sentences while other techniques needed to be used to create the actual dialogue management. We used Budgie to create a system to train medical students in an anamneses dialogue where they try to create a diagnosis for an artificial patient.

Implications

Based on the developed system we can now train students on many different patients with increasing complex symptoms. The patients can be varied in age (children to retired), varied in cultural background, varied in personality, etc. This facilitates a much broader training of students than what is currently available with life actors. Although the use case was created for medical students, we can use the same basis for a system to train police officers to interview suspects, witnesses and victims in order to establish the best possible diagnosis of a crime that has been committed.

Reference

Yildiz, E., Bensch, S. & Dignum, F. 2022. Incorporating social practices in dialogue systems. In: Følstad et al. *Chatbot Research and Design.* CONVERSATIONS 2021. Lecture Notes in Computer Science 13171. Springer. https://doi.org/10.1007/978-3-030-94890-0_7.

Research Team

Frank Dignum, Umeå University
Eren Yildiz, Umeå University
Fatima Pedrosa-Domellöf, Umeå University

Contact

dignum@cs.umu.se

AI – Destroyer or Enabler of Democracy and Self-Determination?

A key result of the project is the interdisciplinary investigation of how AI systems impact self- determination and democracy. Politically, the concept of automated systems of governing was developed to describe how AI is embedded in governance, challenging transparency and accountability. Philosophically, the project examined how AI reshapes the balance between technology and autonomy, emphasizing the normative ambiguity in AI's influence. Another result of the project is an agent-based simulation, CollAct, to how AI systems influence democratic processes, power dynamics, and self-determination. The platform enables to investigate collective action and inequality, as demonstration that design choices impact trust and democratic legitimacy. The analysis of Voting Advice Applications (VAAs) highlighted transparency gaps and the normative implications of automated electoral guidance.
Implications

Implications

Our results reveal that AI systems are not neutral tools but political and moral systems that reconfigure public administration, citizenship, and human agency. This challenges political science to rethink democratic legitimacy, power distribution, and institutional accountability in the age of automation. For philosophy, our results show how AI challenges established ideas of autonomy, responsibility, privacy, and moral agency, calling for new conceptual frameworks and further reflection on how to navigate tensions between autonomy and algorithmic nudging, and between human and machine agency. AI research itself must move beyond technical

efficiency to include normative reflection, participatory design, and critical social inquiry. Without this interdisciplinary engagement, AI risks deepening inequality, eroding trust, and shifting democratic power to opaque systems and private actors.

References

Öjehag-Pettersson, A., Carlsson, V. and Rönnblom, M. (2023). Political Studies of Automated Governing: A Bird's Eye (re)view. *Regulation & Governance* doi.org/10.1111/rego.12569.

Stockinger, E., Maas, J., Talvitie, C. and Dignum, V. (2024). Trustworthiness of voting advice applications in Europe. *Ethics and Information Technology* 26(55). doi.org/10.1007/s10676-024-09790-6.

Lindgren, H., Lindvall, K. and Richter-Sundberg, L. (2025). Responsible design of an AI system for improving health – an Ethics perspective on a participatory design process of the STAR-C digital coach for behaviour change. *Frontiers in Digital Health* doi.org/10.3389/fdgth.2025.1436347.

Research team

Virginia Dignum, Umeå University
Helena Lindgren, Umeå University
Kalle Grill, Umeå University
Malin Rönnblom, Karlstad University

Contact

virginia@cs.umu.se

Artificial Intelligence and Industrial Transformations

AI has been increasingly experimented in various sectors of society and industry. However, AI experimentation does not mean AI adoption. While some AI experimentations can be promising and may generate value, many others often result in misuse or discontinuance of AI in organizations. Albeit the growing expectations, there are surprisingly very few documented cases of AI adoption where AI led to substantial value for business and society so far.

The disruptive impacts of AI are often induced by the stories of a very few large companies – mostly big tech companies – that have access to big data and operate in highly digitalized contexts. Most small-sized firms however do not have the necessary resources to gather reliable datasets and make new arrangements between domain and algorithm experts, which are key for robust adoption of AI.

Implications

AI is a fuzzy, umbrella concept referring to a large number of technologies and applications. To embrace the implications, we need to discuss specific AI-applications separately. There is a strong need of empirical studies on the implications of various AI-technologies in organizations that have moved from experimentation to adoption.

We need domain experts, who are capable of evaluating the quality of algorithmic decisions in various domains (e.g., consulting, accounting, public authorities, hospitals etc.), complementing algorithm experts. It is also important to have a critical perspective on AI as the overarching solution. Currently, there is a strong AI imperative in society. AI is viewed as a

sought for solution, but the problem is often unclear (and AI might not be the valid solution).

References

Karakaya E. & Engwall M. 2025. Variety and Generality of Artificial Intelligence: Revisiting managerial case studies. *Proceedings of Academy of Management Conference*. https://journals.aom.org/doi/abs/10.5465/AMPROC.2025.23410abstract.

Gerigoorian, A., Kloub M., Dembrower K., Engwall M. & Strand F. 2025. Risk Inventory and Mitigation Actions for AI in Medical Imaging - A Qualitative Study of Implementing Standalone AI for Screening Mammography. *BMC Health Services Research*. 25(998). doi.org/10.1186/s12913-025-13176-9.

Research team

Mats Engwall, KTH Royal Institute of Technology
Emrah Karakaya, KTH Royal Institute of Technology

Contact

mats.engwall@indek.kth.se

Culturally Informed Robots in Learning Interaction

Robots can alter conversations between students of different linguistic levels, making quieter ones more active when given more attention with non-verbal robot cues (humming, gaze, head movements). The strength of this effect partly depends on socio-cultural background.

Culturally-coded robots are regarded as having different levels of competence based on their accent or facial features, but this bias fades in real interaction. Cultural matching (e.g., Swedish accented English for Swedish students) can improve recall, despite lower perceived competence.

Students often trust robots' faulty answer over their own correct knowledge. Trust varies with prior use of Large Language Models (more use leads to less critical thinking), self-image (students with higher perception of their own knowledge are more critical), and the robot's verbal and non-verbal confidence in its delivery.

Implications

The three main results above have the following implications:

First, since students' socio-cultural background affects how effective non-verbal robot attention is, feedback signals and frequency may need to be enforced for groups of students who are less responsive to attention by a robot.

Second, cultural matching between robot and students can be used to potentially improve learning.

Third, educators need to be aware of students' over trust in information provided by AI and developers need to adjust the wording and delivery of LLM generated information in order to

maintain students' critical assessment of facts presented by robots and other AI systems.
In particular, the first and third implication are explored further in the team's new research projects.

References

Engwall, O., Cumbal R. and Majlesi, A.R. (2023). Socio-cultural perception of robot backchannels. *Frontiers in Robotics and AI* 10. doi.org/10.3389/frobt.2023.988042.

Gonzalez Oliveras P., Engwall O. and Majlesi A.R. (2025). Sense and Sensibility: What makes a social robot convincing to high-school students? *Proceedings of Robotics: Science and Systems*. doi.org/10.48550/arXiv.2506.12507.

Gonzalez Oliveras P., Engwall O. and Wilde A. (2025). Social Educational Robotics and Learning Analytics : A Scoping Review of an Emerging Field. *International Journal of Social Robotics* 17:1113–1128. doi.org/10.1007/s12369-025-01235-4.

Research Team

Olov Engwall, KTH Royal Institute of Technology
Ali Reza Majlesi, Karolinska Institutet
Pablo Gonzalez Oliveras, KTH Royal Institute of Technology
Ronald Cumbal, KTH Royal Institute of Technology
Olga Viberg, KTH Royal Institute of Technology
Iolanda Leite, KTH Royal Institute of Technology

Contact

engwall@kth.se

Quantifying Culture – AI and Heritage Collections

By integrating artificial intelligence and machine learning into the digitization, interpretation, and classification of cultural heritage (CH) objects, the project pioneers both technological innovation and critical examination of AI's societal impact.

AI technologies offer profound opportunities in cultural heritage for Digitization and Restoration – Machine learning mo-dels improve the restoration of artifacts and expand accessibility via predictive maintenance, multimodality and 3D reconstruction, giving new life to historical materials; Interpretation and Classification – Collaborative models now attempt to recognize deep cultural nuances, such as regional differences in traditional attire, rather than perpetuating reductive, monolithic categorizations. Integration of ethical theory within algorithmic design further supports this nuanced approach.

Despite these benefits, the application of AI in CH raises significant concerns about Bias Amplification – AI trained on historically skewed datasets tends to reinforce colonial-era or dominant narratives, leading to artifact misclassification, such as misgendering based on stereotypical assumptions about clothing; Synthetic Inaccuracies – Generative AI tools (e.g., DALL-E) can create spurious visualizations, like anachronistic renderings of Greek kouroi statues or medieval maps, that may seem credible but lack historical accuracy, introducing misinformation; Metadata Limitations – Insufficient or poorly annotated datasets result in oversimplified classifications, erasing vital cultural subtleties and leading to interpretive errors.

Implications

AI can enhance inclusivity and accessibility in CH collections, fostering richer and more representative interpretations of cultural

artifacts. By directly addressing AI bias, the project supports the development of tools and practices that respect and reflect diverse cultural perspectives, potentially transforming digital heritage practices globally. AI's impact on the authenticity and reliability of CH materials requires robust ethical frameworks to balance technological advances with cultural sensitivity. Critical solutions include Human-in-the-loop validation, Community led anno-tation, and Interoperable standards.

References

Foka, A. and Von Bonsdorff, J. (2025). *AI and Image: Critical Perspectives Heritage and Art*. Cambridge: Cambridge University Press.

Foka, A. and Griffin, G. (2024). AI, Cultural Heritage, and Bias: Some Key Queries That Arise from the Use of GenAI. *Heritage* 7: 6125–6136. doi.org/10.3390/heritage7110287.

Foka, A., Griffin, G., Ortiz Pablo, D. Badri, S. and Rajkowska P. (2025). Tracing the bias loop: AI, cultural heritage and bias-mitigating in practice. *AI & Society* doi.org/10.1007/s00146-025-02349-z.

Research Team

Anna Foka, Uppsala University
Sushruth Badri, Uppsala University
Gabriele Griffin, University of the Free State
Nasrin Mostofian, Uppsala University
Paulina Raijowska, Uppsala University
Dalia Ortiz Pablo, Uppsala University
Fredrik Wahlberg, Uppsala University

Contact

anna.foka@abm.uu.se

The Artificial Public Servant

Strides in AI technology have led to the development of systems based on increasingly advanced algorithms. These systems have limited transparency, can process large quantities of data, are flexible, and can to some extent design themselves. In this project, the researchers are examining the issue of accountability from a philosophical and legal perspective in relation to automated or AI-controlled decision making in public administration. When automated or AI-controlled decision-making systems become more common, they might conceivably replace a large proportion of human decision making in public administration. In other situations, decision making processes will still be human centred albeit supported by AI systems. This prospect raises fundamental philosophical issues and legal questions about accountability: who or what should be held accountable for wrong decisions and why?

Implications

We ascertain whether there is a need to find new rules governing accountability for autonomous systems, or modify existing ones. To do so, we analyze both how accountability works under Swedish laws, and how moral accountability and moral agency can be understood by means of philosophical reasoning.
We present proposals for modified categories of accountability and new forms of accountability that can be applied both generally and more specifically in a Swedish context under the law. The philosophical and legal analysis yields a deeper understanding of the ethical, legal and societal implications of introducing AI systems in public administration.

References

Li, O. (2024). Should We Develop AGI? Artificial Suffering and the Moral Development of Humans. *AI and Ethics* 5: 641–651, doi.org/10.1007/s43681-023-00411-4.

Razmetaeva, Y. (2024). Private Algorithms, Public Consequences, Pp. 161-173 in Cascione, Codiglione, Pardolesi (Eds) *Public and Private in Contemporary Societies*. Rome: Roma Tre Press. doi.org/10.13134/979-12-5977-393-7.

Research team

Sandra Friberg, Uppsala University
Johan Eddebo, Uppsala University
Oliver Li, Uppsala University
Yulia, Razmetaeva, Uppsala University

Contact

sandra.friberg@jur.uu.se

Drone Technology and their Presence in Society

First, through ethnographic methods, we have investigated the complexity of managing safe drone practices in organizations such as film industry, forestry, energy, police and more. This has revealed the need for planning and engagement of different stakeholders and the many challenges of ensuring safety, when flying drones. We have also investigated hobby drones from perspectives of commonly excluded stakeholders, such as children and people with disability needs. Second, following a more classical Human-Computer Interaction (HCI) approach with indoor lab experiments, we have examined the feasibility of drone functions, drone appearance (including noise), and investigated sound-masking alternatives. We also studied human experiences with a novel robotic platform, namely a bioinspired flapping-wing drone. Overall, our work has both problematized the use of drones and explored alternative form-factors and stakeholder views in a variety of situations and settings.

Implications

Our research stresses the importance of investigating existing practices and a variety of stakeholders, to understand the impact of drone practices on people. Our research implies that drone technology is fundamentally disruptive, and that avoiding drone presence in certain situations is needed. Safety was found to be a top priority for professional drone pilots and should not be taken for granted in early drone concepts. Our implications also point towards usability and accessibility issues with drones, but also potential opportunities for drones to act as accessible camera devices in inaccessible environments for people with disabilities. As for the perception of drone noise, we have found that sound-masking is difficult and culturally dependent. Overall, drones have

several form, safety and noise issues, and there are many different stakeholders and groups of people that are affected when they are used.

References

Ljungblad, S., et al. (2021). What matters in professional drone pilots' practice? An interview study to understand the complexity of their work and inform human-drone interaction research. In *Proceedings of the 2021 CHI conference on human factors in computing systems* 1-16. doi.org/10.1145/3411764.3445737.

Gamboa, M. (2022). Living with drones, robots, and young children: informing research through design with autoethnography. In *Nordic Human-Computer Interaction Conference* 1-14. doi.org/10.1145/3546155.354665.

Wang, Z., et al. (2023). The effects of natural sounds and proxemic distances on the perception of a noisy domestic flying robot. *ACM Transactions on Human-Robot Interaction* 12(4):1-32. doi.org/10.1145/3579859.

Research team

Morten Fjeld, Chalmers & University of Gothenburg
Sara Ljungblad, Chalmers & University of Gothenburg
Miriam Sturdee, St. Andrews University
Mohammad Obaid, Chalmers & University of Gothenburg
Ziming Wang, Chalmers & University of Gothenburg

Contact

sara.ljungblad@chalmers.se

How to Use Large Language Models in a Sensitive way to Political Data

In our project we detect biases in methods of AI technology studying political materials. A characteristic of political texts is that information can be more or less explicit. In our work, we find that large language models generally benefit from pre-training on a large off-the-shelf dataset, rather than a specialized dataset. The need for large data is thus present also studying political texts in an LLM context. As an extension, we find that word embeddings have a tendency to produce more reliable results for "objects", rather than more value-laden concepts. We find a similar tendency when producing so-called silicon samples that respond to social science experiment questions. This indicates that present large language models are insufficient to detect tacit information, whereas it can be really efficient to provide general tendencies of word use and word combinations.

Implications

Researchers should think twice before engaging LLMs in a task, since its performance will vary heavily on the question and also on the proficiency of the user who poses the question/prompts.

References

Fredén, A., Johansson, M. and D. Saynova. (2024). Word embeddings on issues and ideology from Swedish parliamentarians' motions: A comparative approach. *Journal of Elections, Public Opinion and Parties*. doi.org/10.1080/17457289.2024.2433979.

Bruinsma, B, et al. (2024). Setting the AI-Agenda. Evidence from Sweden in the Chat GPT Era. Paper presented at the AEQUITAS workshop, ECAI, Santiago de Compostela, 20 October 2024. doi.org/10.48550/arXiv.2409.16946.

Saynova, D., et al. (2025). Identifying Non- Replicable Social Science Studies with Language Models. Pre-print doi.org/10.48550/arXiv.2503.10671.

Research team

Annika Fredén, Lund University
Bastiaan Bruinsma, Chalmers University of Technology
Denitsa Saynova, Chalmers University of Technology
Kajsa Hansson, Lund University
Moa Johansson, Chalmers University of Technology
Pasko Kisic-Merino, Karlstad University

Contact

annika.freden@svet.lu.se

Trusting AI – When Children Seek Answers from Machines over Parents

AI might change how children seek knowledge. Our latest research examines when children prefer AI over their parents for information. When engaging with human-like AI voice interaction, 8-year-olds selectively seek different types of information from AI versus parents, recognizing that AI provides more extensive formal knowledge but lacks awareness of the child's personal history. However, strategies for obtaining social information vary: Some children rely almost entirely on AI, while others exclusively turn to their parents. Although strategies remain unchanged throughout the experiment, many children ask a few control questions to verify their assumptions about the system's capabilities and limitations. When alone with AI, behaviors diverge: some challenge norms with unconventional questions, whereas others follow conventional conversational patterns.

Implications

The results indicate that children understand what AI excels at and actively test their assumptions about its capabilities. The findings further suggest that AI's role in children's learning environments will vary based on individual differences in trust and curiosity. Since asking questions is fundamental to learning, AI might reshape early learning dynamics, particularly the parent's role as an "oracle" of knowledge.

As children gain access to highly responsive and seemingly all-knowing AI systems, their reliance on parents for information may shift, with potential implications for pedagogy, child-parent relationships, and social development. Next, we will implement AI in interactive robots and assess these

dynamics across different ages. Understanding these processes is essential for guiding the future relationship between children, technology, and those who support their development.

Research team

Gustaf Gredebäck, Uppsala University
Kim Astor, Uppsala University

Contact

gustaf.gredeback@psyk.uu.se

The Ethics and Social Consequences of AI and Caring Robots

Social robots are imagined and being developed as providers of care to vulnerable groups such as the elderly or children. In these contexts, social robots are primarily considered as providing social support and companionship. This narrative is widely disseminated and accepted both amongst care managers and the HRI research community. The reality is that social robots are mostly used in controlled laboratory settings due to technical limitations. Our work revealed a pressing need for more "in the wild" studies, but also a willingness within the HRI community for interdisciplinary collaboration around ethical questions of how power, bias and normativity shape development of social robots. We also found a great deal of resonance with and interest in feminist work by the HRI community.

Implication

We've changed the discourse around research ethics and normative practices for (a few) robotics researchers through engagement with HRI researchers, hands-on experimental collaborations and co-authorships. The team worked to sensitise the field to ethical questions understood within the frame of feminist ethics, emphasizing the importance of (legally dictated) research ethics structures as well as the value of thinking about ethics as the power dynamics of research practices.

References

Winkle, K et al. (2023). 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. doi.org/10.1145/3568162.3576973.

Velázquez, I.G. (2023). The Making of Gendered Bodies in Human-Robot Interactions. *International Journal of Social Robotics* 15(11):1891-1901. doi.org/10.1007/s12369- 023-00979-1.

Perugia, G. and Lisy, D. (2023). Robot's gendering trouble: a scoping review of gendering humanoid robots and its effects on HRI. *International Journal of Social Robotics* 15(11): 1725-1753. doi.org/10.1007/s12369-023-01061-6.

Research team

Katherine Harrison, Linköping University
Ericka Johnson, Linköping University
Ginevra Castellano, Uppsala University
Amy Loutfi, Örebro University

Contact

katherine.harrison@liu.se

AI and the Artistic Imaginary – Socio-Cultural Consequences and Challenges of Creative-AI Technology

The environmental footprint of Creative AI (AI used in the context of creative practices, such as image or music generators) and other recent machine learning technologies is increasing. To understand the scale of the problem in machine learning research, more knowledge is needed of the current energy cost of the undertaken research. In a recent study, we provide an inquiry of how research concerning automatic music generation and computing-heavy music analysis currently discloses information related to environmental impact. Our study demonstrates a lack of transparency in model training documentation. It provides a careful first estimate of energy consumption related to model training at the main music informatics conference (ISMIR), amounting to 7.5 tons of carbon dioxide. A majority of this environmental impact is related to papers with industry affiliation.

Implications

We have previously analysed the entanglement between corporations and environmental impact, arguing that such a perspective of political ecology is important to grasp the complexity of sustainability. While it may be argued that the environmental impact of Creative AI is small compared to other industries, we raise the question of how much (green) energy we can use to generate images and music: should we use the energy for this purpose while some people lack it for basic needs such as food production or heating? The need to guide behaviour change and to consider planetary boundaries is

something we argue for. Creative AI technologies will be needed that respect local cultural practices and that are small and flexible in their design and application, implications in stark contrast with current technology innovation.

References

Holzapfel, A., Kaila, A.K. and Jääskeläinen, P. (2024). Green MIR?: Investigating computational cost of recent music-Ai research in ISMIR. In *International Society for Music Information Retrieval Conference (ISMIR)*. ismir2024program.ismir.net/poster_113.html.

Holzapfel, A. (2023). Introducing political ecology of creative-AI, In Lindgren (Ed.) *Handbook of critical studies of artificial intelligence*. Edward Elgar Publishing, pp. 691-703. doi.org/10.4337/9781803928562.00070.

Jääskeläinen, P. and Biørn-Hansen, A. (2024). Critical Questions for Sustainability Research in Computational Creativity. In *ICCC'24 15th International Conference on Computational Creativity*. https://computationalcreativity.net/iccc24/proceedings/.

Research team

Andre Holzapfel, KTH Royal Institute of Technology
Anna-Kaisa Kaila, KTH Royal Institute of Technology
Bob L. T. Sturm, KTH Royal Institute of Technology
Cecilia Åsberg, Linköping University
Petra Jääskeläinen, KTH Royal Institute of Technology

Contact

holzap@kth.se

Cultivating Ethical Sensibility in Design Practice

In this project, researchers from KTH and Stockholm University explored the felt and affective dimensions of ethics in design practice and technology production. Our approach to autonomous systems—such as robots and drones—is grounded in felt ethics, a framework developed within this project to cultivate ethical sensibility in design. Felt ethics emphasizes:

- Processual cultivation through analytical, pragmatic, and hands-on engagement
- Critical attentiveness to the limits of our own bodies and lived experiences
- Recognizing ethical practice as a matter of care

By attending to the felt dimensions of ethics, including vulnerability and discomfort, and connecting them with reflections on the power structures within which novel technologies are created, we seek to reframe how ethics is embedded in design and technology development and articulate pathways to better practice and design.

Implications

Our interaction design work engages with autonomous technologies through creative practices – such as repurposing in-dustrial robots via dance or designing drones to explore human-technology relations. We use felt ethics to deconstruct these interactions somatically, analytically, and critically, exposing the ethics embedded in their design. By reimagining the aesthetics, ethics, and politics of such technologies, we then foster somatic freedoms. In our sociologically oriented

research, we examine the felt and affective dimensions of ethics in technology production, focusing on practitioners' experiences. We are particularly interested in discomfort—arising from power imbalances or a misalignment between perceived responsibility and institutional authority to act. The work highlights the pivotal role emotions play in recognizing ethical issues and embracing responsibility.

Reference

Popova, K., et al . 2022. Vulnerability as an ethical stance in soma design processes, *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* pp. 1–13. doi.org/10.1145/3637434.

Research Team

Kristina Höök, KTH Royal Institute of Technology
Airi Lampinen, Stockholm University
Kristina Popova, KTH Royal Institute of Technology
Rachael Garrett, KTH Royal Institute of Technology

Contact

khook@kth.se

Professional Trust and Autonomous Systems

Trust in AI and autonomous systems parallels the conditions governing interpersonal trust. It is not an optional stance but rather the default in routine activities. Professionals evaluate AI-generated actions and outcomes based on their assumptions about standard operations. While algorithmic processes can remain opaque—black-boxed and hidden— this is not inherently problematic. However, any observable components or outputs must align with the professionally expected norms. For AI systems to be trusted, they must confirm these expectations; failure to do so results in a loss of trust, ultimately disrupting the activity.

Implication

If trust is tied to the perceived normality of events, as our research demonstrates, this poses a challenge for information that deviates from historical data. Any incongruous information risks being dismissed as irrelevant or seen as a disruption to the expected order. While AI can identify novel patterns and insights, the emphasis should not solely be on designing systems that inspire trust. Instead, organizations must critically assess their routines and strategies for integrating unexpected findings. New insights hold little value if they are disregarded as mere anomalies.

Reference

Ivarsson J. 2023. Dealing with Daemons. Trust in Autonomous Systems. Pp.163-180 in Sormani P and vom Lehn D (Eds) *The Anthem Companion to Harold Garfinkel*. Anthem Press. doi.org/10.2307/jj.4418210.13.

Research Team

Jonas Ivarsson, University of Gothenburg
Alan Said, University of Gothenburg
Jabbar Hussain, University of Gothenburg
Magnus Båth, University of Gothenburg
Rolf Heckmann, University of Gothenburg
Shuren Yu, University of Gothenburg
Thomas Hillman, University of Gothenburg
Åsa Mäkitalo, University of Gothenburg
Åse Johnsson, University of Gothenburg

Contact

jonas.ivarsson@gu.se

The Quantum Law Project

Quantum computers are legal things which are going to affect our lives in a tangible manner. As such, their operation and development must be regulated. While the transformational potential of quantum computing is remarkable, their development will also significantly impact social and legal power-relations. The Quantum Law Project seeks to identify legal principles that can guide regulatory action to hedge the risks associated with quantum computing. One way in which The Quantum Law Project contributed to the development of such principles is by proposing the "quantum imperative". The quantum imperative provides that regulators must ensure that the development of quantum computers: (1) does not create or exacerbate inequalities, (2) does not undermine individual autonomy, and that it (3) does not occur without consulting those whose interests they affect.

Implications

The findings of The Quantum Law Project reveal that, contrary to popular belief, the ethical, legal, and social challenges posed by quantum computing are not unique but in many ways similar to those raised by artificial intelligence (AI).

While quantum computing promises to revolutionize fields such as cryptography and optimization in completely novel ways, the concerns related to these processes overlap significantly with those associated with AI (e.g. biases, accountability, predictability). Both technologies challenge existing legal frameworks and societal norms, demanding a re-evaluation of laws, ethics, and governance to address potential risks and ensure equitable benefits for all. This convergence calls for unified, forward-thinking approaches to regulation and

policy development. We must think of quantum-governance as being one part of AI-governance in general.

References

Jeutner, V. The Quantum Imperative: Addressing the Legal Dimension of Quantum Computers, *Moral & Machines* 52-59. doi.org/10.5771/2747-5174-2021-1-52.

Atik, J. and Jeutner, V. (2021). Quantum Computing and Computational Law. *Law, Innovation and Technology* 13(2):302-324. doi: 10.1080/17579961.2021.1977216.

Research Team

Valentin Jeutner, Lund University
Aurelija Lukoseviciene, Lund University
Jeffery Atik, Lund University
Seyedehoda Hosseiny, Lund University
Timo Minssen, Lund University

Contact

valentin.jeutner@jur.lu.se

Complexity and Fairness in Synthetic Structured Data

AI-generated synthetic structured datasets are made using machine learning techniques to reproduce the essential elements of an existing dataset. This might be done to ensure privacy or to expand, enhance or substitute for real-world datasets or to create a portable or shareable dataset that is considered safe for open access, for example to share via a data repository. While synthetic structured data may reproduce some of the essential elements of an original dataset, it will also inevitably introduce “intersectional hallucinations” - anomalous inter-attribute relations. AI generated synthetic data also have a known tendency to minimize minority elements and amplify majority elements. Thus, knowing in what ways a synthetic dataset demonstrates fidelities and in what ways it is ‘different’ from the original data is essential for successful and responsible re-use of synthetic data.

Implications

Theoretically, understanding and showing the complexity of synthetic structured data is essential for explainable AI. This also has implications for the development of theories about world-data relations. Practically, synthetic structured data should be examined for reliability and representation (and labeled with these) before data is shared, used, or added to existing datasets. Given that the goal of many data repositories is to provide access to data that is replicable and/or reusable, there is a clear need to establish protocols for documenting synthetic data. Domain experts should be prompted to document the context and motivation for generating synthetic data, in addition to its specific reliability. Understanding and

controlling for intersectional fidelities and hallucinations will be important when using synthetic data as output or when producing it to amplify training data.

References

Johnson, E and S. Hajisharif. (2024). The Intersectional Hallucinations of Synthetic Data. Curmudgeon's Corner. *AI & Society* 40:1575–1577. doi.org/10.1007/s00146-024-02017-8.

Lee, F., Hajisharif, S., and Johnson, E. (2025). The ontological politics of synthetic data: normalities, outliers, categories, and intersectional hallucinations. *Big Data & Society*. doi.org/10.1177/20539517251318289.

Research team

Ericka Johnson, Tema, Linköping University
Francis Lee, Södertörn University
Tahereh Dehdarirad, Linköping University
Gabriel Eilertsen, Linköping University
Saghi Hajisharif, Linköping University

Contact

ericka.johnson@liu.se

Existential Challenges and Ethical Imperatives of Biometric AI in Everyday Lifeworlds

The project has focused on how the rapid developments of biometric AI (e.g. facial and voice recognition) co-shape, bring about, and transform human existence. It has led to vital insights about biometrics through existential, ethical and embodied perspectives. Key results demonstrate the sometimes violent potency of the biometric imaginary in public discourse and administrative practice, alongside the ambiguities of biometrics as lived experience, and the leaps of faith on the side of AI professionals. Regarding biometric technologies as relational technologies, the project has highlighted how they variously forge, violate, and renegotiate relations between human-machine, state-citizen, self-other, and body-subject. Our findings show that they come with relational expectancies; highlight bodily processes as relational events; and have an antagonistic relationship with the living subject.

Implication

BioMe has furthered the field of existential media studies through transdisciplinary competencies and practice-based, artistic research. During the course of the project the approach has spread internationally and has featured in courses at distinguished universities. Collaborative research with engineers and societal actors about ethically and existentially sustainable futures with technologies has resulted in public activities at various cultural institutions (e.g. The Museum for Science and Technology, Liljevalchs Gallery, The Nobel Prize

Museum) as well as a popular science book communicating our results to a broader audience.

References

Lagerkvist, A. (Ed.) (2024). *AI och samtalet om de stora frågorna. Möten mellan existentiella och teknologiska perspektiv*. Göteborg: Makadam

Lagerkvist, A. & J. Smolicki (Eds.). forthcoming. Relational Technologies: In *Search of the Self Across Datafied Lifeworlds*, London: Bloomsbury Press.

Lagerkvist, A., Tudor, M., Smolicki, J. et al. 2024 Body stakes: an existential ethics of care in living with biometrics and AI. *AI & Society* 39:169–181. doi.org/ 10.1007/s00146-022-01550-8

Research team

Amanda Lagerkvist, Uppsala University
Charles M. Ess, Unversity of Oslo
Jacek Smolicki, Uppsala University
Jenny Eriksson Lundström, Uppsala University
Maria Rogg, Uppsala University
Matilda Tudor, Uppsala University

Contact

amanda.lagerkvist@crs.uu.se

Shaping a Future with Complex Intelligent Systems

The implementation of AI beyond stand-alone solutions emerges as a long journey, taking place during an extensive period, affecting organizations, business and society. Shaping a future with AI in a responsible way implies considering advantages and risks with AI related to the context AI is implemented within. Our results show new relationships between organization and system design, beyond classic mirroring approaches. This builds on multifacetted dialectic solutions that address many aspects and the system's dynamic context. For instance, our results overturn traditional perspectives on system autonomy as weakening human authority by showing that system autonomy potentially strengthens human authority when management, engineering, organization and societal aspects are simultaneously addressed.

Implications

AI is increasingly seen as a contributor to complex systems in the fields of for instance energy, healthcare, and transporation. With AI, they are increasingly capable of analyzing a dynamic context and taking actions towards specific goals with some degree of autonomy. The design and operation of such complex intelligent systems needs to be shaped so that human intelligence, organization, and system autonomy coexist. Future management needs to build on a renewed understanding of the fluid boundaries between humans, organizations and technical systems, their interaction, and the balance between human authority and system autonomy. This represents a context beyond the AI application itself, affecting the role of existing

system integrating firms and requiring future engineers, managers, and stakeholders to integrate a wider spectrum of aspects in system development.

References

Lakemond, N., Holmberg, G. and Pettersson, A. (2024). Digital transformation in complex systems. *IEEE Transactions on Engineering Management* 71:92-204. doi.org/10.1109/TEM.2021.3118203.

Yu, Youshan, Lakemond, N. and Holmberg, G. (2024). AI in the Context of Complex Intelligent Systems: Engineering Management Consequences. *IEEE Transactions on Engineering Management* 71:6512-25. doi.org/10.1109/TEM.2023.3268340.

Balachandran, A., Holmberg, G. and Lakemond, N. (2024). Understanding the development of emerging complex intelligent systems. *Journal of Engineering and Technology Management* 72:101815. doi.org/10.1016/j.jengtecman.2024.101815.

Research team

Nicolette Lakemond, Linköping University
Appu Balachandran, Linköping University
Bijona Troqe, Linköping University
Elinor Särner, Linköping University
Gouthanan Pushpananthan, Linköping University
Gunnar Holmberg, Linköping University
Youshan Yu, Linköping University
Yunchen Sun, Linköping University

Contact

nicolette.lakemond@liu.se

AI Transparency and Consumer Trust

We research how transparency in applied AI can strengthen consumer trust, and promote fair and accountable AI. Most striking is to uncover the multifaceted meanings of AI transparency and their relevance for peoples' understanding, trust and interaction with AI-assisted tools and agents, and as a governance tool in the EU. Transparency serves as an entry point for the versatile, multi-methodological and interdisciplinary study of chatbot and human-robot interaction, consumer attitudes, major EU regulation, and normative implications of the embedding of societal structures in AI training data.

Implications

There are several implications of the results from this project, including the clarification of regulatory challenges brought about by the EU's bet on transparency as a governance tool. For example, the link between transparency and trust is far more complicated than what at times is argued in narrow disciplinary takes. Not only does this require bridging the technical, social and humanistic sciences, but from an implementation perspective there is a need for institutionalised validation of AI-systems that ensure safe and fair use irrespective of consumer and other groups' ability to scrutinise their details. In addition, as AI-services and agents become more varied and individualised, the structural aspects of fairness become increasingly important to understand – for whom (and not) and by whom (and not) they are developed – just as how to push for accountable design of such individualised and adaptive ("agentic") AI when deployed for human interaction.

References

Larsson, S., Liinason, M., Tanqueray, L. and Castellano, G. (2023). Towards a Socio-Legal Robotics: A Theoretical Framework on Norms and Adaptive Technologies. *International Journal for Social Robotics* 15(11):1755-1768. doi.org/10.1007/s12369-023-01042-9.

Haresamudram, K. (2025). *Interactions with Pseudo-Sapiens: User perception of Anthropomorphism, Mind, and Trust in Humanlike Social Agents*. Lund University: Doctoral dissertation.

Söderlund, K. (2025). *AI Transparency in Trustworthy AI: From Metaphor to Governance Tool in EU Technology Regulations*. Lund University: Doctoral dissertation.

Research team

Stefan Larsson, Lund University
Fredrik Heintz, Linköping University
James M White, Lund University
Kashyap Haresamudram, Lund University
Kasia Söderlund, Lund University
Laetitia Tanqueray, Lund University

Contact

stefan.larsson@lth.lu.se

This study reveals how data systems create what we accept as reality, using pandemic maps as a powerful example. By observing health experts during the Zika crisis, we discovered patterns that apply far beyond disease tracking. Just as Zika maps combined spotty surveillance data with computer predictions to create an authoritative picture of global risk, similar processes shape what we know about climate change, economic trends, election forecasts, and social media content. We found that data gaps were routinely filled with assumptions, different data sources were treated with uneven levels of trust, and uncertain predictions were presented as definitive facts. Most importantly, the final visualizations and reports hid these messy realities, creating an illusion of complete knowledge while masking the human judgments and flawed data that underpinned the entire system.

Implication

These findings matter for anyone who encounters data-driven knowledge in daily life. The same problems we identified in pandemic tracking—missing data treated as "zero cases," algorithms that give different weight to different sources, and simplified visualizations that hide uncertainty—appear in weather forecasting, crime prediction, economic indicators, and social media recommendation systems. Understanding these patterns helps us become more critical consumers of all data-driven knowledge. Instead of accepting colorful maps, compelling graphs, or algorithmic recommendations at face value, we can ask better questions: What data is missing? Who decided which sources to trust? Where is uncertainty being

hidden? By revealing how human values and judgments are embedded in seemingly objective systems, this research provides essential tools for navigating our increasingly data-driven world with a critical eye.

References

Lee, F. (2024). Ontological overflows and the politics of absence: Zika, disease surveillance, and mosquitos. *Science as Culture* 33(3): 417-442. doi.org/10.1080/09505431.2023.2291046.

Lee, F. (2021). Enacting the Pandemic: Analyzing Agency, Opacity, and Power in Algorithmic Assemblages. *Science & Technology Studies* 34(1): 65-90. doi.org/10.23987/sts.75323.

Research team

Francis Lee, Södertörn University
Alicja Ostrowska, Chalmers University of Technology
Shai Mulinari, Lund University

Contact

francis.lee@sh.se

Artificial Intelligence, Democracy and Human Dignity

This project takes a holistic perspective on the institutional, legal and philosophical effects of AI developments. We wanted to contribute to the joint research task to find reasonable paths in a complex and rapidly changing situation. This gave us insights on how to design an effective multidisciplinary research project in a field that is underexplored. We have created new methodological tools in close dialogue within the group. It is important to cherish the different research traditions and experiences so that they are seen as an asset and that these are kept alive. This helped us understand the fields of our inquiry in a new and constructive way.

Implications

Analysing the interplay between technology, law and political processes have revealed unique risks that are related to AI technologies and given us understanding for how future legal regulation and political processes can be dealt with. The broad AI development is in itself a complex historical and cultural phenomenon that demands thorough in-depth research. Our project has contributed to new forms of analysis relating to how technologies, world views, politics, culture and law interact in processes of change. This is in itself an important basis for continuous critical research in the field. The philosophical work done in the different disciplines have contributed to shed new light on the arguments in the debate on artificial subjectivity and a critical approach on the presumptions of strong AI and genuine artificial subjectivity.

Reference

Eddebo, J, Hultin Rosenberg, J, Lind, A-S & Wejryd, J. 2025. *Artificiell intelligens, demokrati och mänsklig värdighet*, CRS rapporter nr 4, 2025

Research team

Anna-Sara Lind, Uppsala University
Johan Eddebo, Uppsala University
Johan Wejryd, Uppsala university
Jonas Hultin Rosenberg, Mälardalen University
Lars Karlander, Uppsala University
Oliver Li, Uppsala University
Silvia Carretta, Uppsala University
Yulia Razmetaeva, Uppsala university

Contact

anna-sara.lind@jur.uu.se

Digital Companions as Social Actors Promoting Health

A digital companion may take on different roles when collaborating with a person. Teams with human team members develop through a transition between phases in the framework by Tuckman. The transition between the phases forming, storming and norming was studied in a team of a human and a digital companion in a joint planning task. It was observed that only when the digital companion elicits disagreement or difference in opinion during the joint planning activity, the person perceives that they come to agreement in the end, meaning that there is a transition from the storming phase to the norming phase in terms of Tuckman's theory on teamwork development in a team of humans. When the digital companion did not explicitly point to the differences, the person maintained the perception that they were in the storming phase although the actions where not indicating that they were.

Implications

The study on teamwork development implies that the digital companion should be capable of identifying disagreements and elicit these to the human actor in collaborative tasks for human-AI teamwork (HAT) to develop.

Further studies are needed to explore to what extent and under what circumstances this observed condition for transitioning from the storming to the norming phase is influencing the transition.

A conclusion made was also that the theoretical framework on human-human teamwork development by Tuckman is applicable also for evaluating HAT. How the framework can be

applied as framework for evaluating HAT will be further explored in future studies.

References

Kaelin, C.K., Tewari, M., Benouar, S. and Lindgren, H. (2024). Developing teamwork: transitioning between stages in human-agent collaboration. *Frontiers in Computer Science*. doi.org/10.3389/fcomp.2024.1455903.

Lindgren, H. (2024). Emerging Roles and Relationships Among Humans and Interactive AI Systems. *International Journal of Human-Computer Interaction* pp. 1–23. doi.org/10.1080/10447318.2024.2435693

Kilic, K., Weck, S., Kampik, T. and Lindgren, H. (2023). Argument-based human-AI collaboration for supporting behavior change to improve health. *Frontiers in Artificial Intelligence* 16(6):1069455. doi.org/10.3389/frai.2023.1069455

Research team

Helena Lindgren, Umeå University
Anna Stigsdotter-Neely, Luleå Technical University
Hanna Malmberg-Gavelin, Umeå University
Julian Mendez, Umeå University
Kaan Kilic, Umeå University
Patrik Björnfot, Umeå University
Vera C. Kaelin, Umeå University
Victor Kaptelinin, Umeå University

Contact

helena.lindgren@umu.se

Cyborg Politics – Non-Human Agency in Democratic Deliberation

The project, currently extended and not yet finalised, works to develop a sociological and critically informed approach to studying non-human agency in digital political settings. Drawing on critical theory and science and technology studies (see also the PI's book Data Theory, Polity, 2020), it investigates how bots, algorithms, and other automated agents shape online deliberation and political discourse. The contributions are, on the one hand, methodological combining computational techniques, such as network analysis, topic modeling, and platform ethnographies, with qualitative discourse analysis. The contribution is also theoretical, entering into much-needed dialogue with tech-centred understandings of AI, the project repositions democratic deliberation as a contested sociotechnical process shaped by asymmetries of power.

Implications

The project has been carried out in a rapidly changing terrain, and has helped position sociology and critical theory as central to the emerging field of critical AI studies, providing conceptual and methodological tools for analysing how automated agency contributes in reshaping political communication, media ecosystems, and public life. It contributes to rethinking democratic agency under conditions of algorithmic governance. Findings overall inform broader debates on misinformation, participation, and accountability in digital society.

References

Lindgren, S. (2023). *Critical Theory of AI*. Cambridge: Polity Press.

Lindgren, S (Ed). (2023). *Handbook of Critical Studies of Artificial Intelligence*. Cheltenhamn: Edward Elgar Publishing.

Research Team

Simon Lindgren, Umeå University
Felicia Lundstedt, Umeå University
Henrik Sigurdh, Umeå University

Contact

simon.lindgren@umu.se

The Missing Teacher in AI

A key result from the project is that teachers, when actively involved in participatory design workshops using "provotypes" (provocative prototypes), were able to identify and articulate critical tensions in how AI tools operate in schools—such as balancing personalized learning with group cohesion, or efficiency with student privacy. These tensions were not treated as problems to eliminate but as starting points for creative design. Teachers used these tensions to co-develop adaptive AI prototype interfaces that supported fairness, transparency, and teacher agency, showing that practitioners can meaningfully shape complex technologies when given the right methods and space.

Implication

This result implies that fairness and ethical use of AI in schools cannot be predefined by developers alone. Instead, adaptive AI systems must be designed for *in-use adaptability*, enabling teachers to modify system behavior based on local classroom needs and values. Standardized AI risks ignoring diverse school contexts and marginalizing vulnerable learners. By involving educators early and meaningfully, schools can develop AI-supported practices that are contextually relevant, equitable, and trusted by those who use them. Participatory design should be a foundational element in educational AI development.

Reference

Utterberg Modén, M., et.al. 2024. When fairness is an abstraction: equity and AI in Swedish compulsory education. *Scandinavian Journal of Educational Research*, 1–15. doi.org/10.1080/00313831.2024.2349908.

Lundin, J. et.al. (2024) A Remedy to the Unfair Use of AI in Educational Settings. *Interaction Design & Architecture(s) Journal* 59:62-78. doi.org/10.55612/s-5002-059-002.

Utterberg Modén, M., et.al. 2024. The Challenge of Incorporating End-User Values into Design: A Methodological Perspective of Using Provotypes, in Barricelli, B.R. et al (eds) *CEUR Workshop Proceedings*. ceur-ws.org/Vol-3685/short11.pdf.

Research team

Johan Lundin, University of Gothenburg
Erik Winerö, University of Gothenburg
Marie Utterberg Modén, University of Gothenburg
Marisa Ponti, University of Gothenburg
Martin Tallvid, University of Gothenburg
Sofia Serholt, University of Gothenburg
Tiina Leino Lindell, University of Gothenburg & Chalmers University of Technology

Contact

johan.lundin@ait.gu.se

Dynamics of AI Use, AI Governance, and Organizational Renewal

More and more decisions are made by algorithms, also in the public sector. But algorithms sometimes take the wrong decisions, which can lead to injustice. So: When algorithms do things wrong, will public institutions put things right?

We analyzed a case of algorithmic decision-making (ADM) in Gothenburg's public school administration, where 1,400 out of 12,000 pupils were placed in the wrong schools. In spite of protests, the school administration failed to address the errors effectively. Furthermore, the administrative court system failed to provide justice, blocking system-level recourse. The study uncovers how organizational ignoring practices can shield an algorithms from scrutiny, creating multiple layers of blackboxing and thus engendering and sustaining both social and legal injustice.

Implications

The study paves the way for interdisciplinary research on the multilayered blackboxing of ADM, extends algorithmic injustice research to include a legal dimension and provides practical implications in the form of a legal framework for ADM in the public sector. The study points to the need to go beyond prevention of algorithmic errors to also build institutional capabilities to identify and address errors and their detrimental consequences. This will be crucial for building societal resilience in the AI age.The study provides a framework for addressing ADM errors in public organizations, focusing on a) transparency and availability of software code, b) reversal of burden of proof from citizens to public agencies, c) mechanisms for contestation and systemic reversal of erroneous algorithmic

decisions, and d) the creation of ombudsperson roles to support the rights of citizens.

References

Kronblad, C., Essén, A. and Mähring, M. (2024). When justice is blind to algorithms: Multilayered blackboxing of algorithmic decision making in the public sector, *MIS Quarterly*, 48(4):1637-1662. doi.org/10.25300/MISQ/2024/18251.

Kronblad, C. (2024). Algoritmisk orättvisa – Göteborgs felkodade algoritm för skolplaceringar. In Fjaestad, M. and Vinge, S. (eds.) AI & makten över besluten. Stockholm: Volante

Research team

Magnus Mähring, Stockholm School of Economics
Anna Essén, Stockholm School of Economics
Charlotta Kronblad, University of Gothenburg

Contact

magnus.mahring@hhs.se

AI and Automated Systems and the Right to Health

AIcare recent outputs includes a study that explores legal issues concerning connected objects used for health or health-related purposes and their corresponding usage of health and health-related data. It uses a healthcare-user-centred perspective researching the EU legal framework for health and health-related data, focusing on data quality, acceptability of connected objects, availability and accessibility of data, as well as the overarching topic of privacy and data protection. It argues that the legal framework, as recently complemented with the European Health Data Space (EHDS) Act, is plagued by complex intersections, while containing several areas of legal uncertainty concerning the interpretation and applicability of existing norms, and areas potentially left untouched. Its main conclusion is that, examined in conjunction, existing regulatory safeguards and certification mechanisms do not offer sufficient protection and simultaneously result in an excessively complex, cumbersome and opaque regulatory framework that has underestimated the specific needs of users in the health and health-related sectors.

Implication

Connected objects, defined as physical objects characterised by being capable of sensing or acting on their environment (directly or indirectly) and able to communicate with each other and other machines or computers, by themselves or enabled by software, are ubiquitous in daily life. Many can be used for health or health-related purposes, or process data related to determinants of health. Their integration with formal healthcare structures is acknowledged by the EU legislator as key for

personalised, preventive and decentralised care, increasing individual autonomy and participation. However, recent legislative initiatives still fall short of creating a framework that is simultaneously agile, innovation-friendly, future-proof and user-centred.

Reference

Nordberg, A., Eskenazy, D. & Holmberg, P. 2025. Health and Health-Related Connected Objects: Regulatory Intersections, Grey Zones and Blind Spots. *European Journal of Health Law*. 32(3):308-333. doi.org/10.1163/15718093-bja10149.

Research team

Ana Nordberg, Lund University
Jennifer Viberg Johansson, Uppsala University
Michele Fariso, Uppsala University
Santa Slokenberga, Uppsala University
Sarah de Heer, Lund University

Contact

ana.nordberg@jur.lu.se

Skill-Intensive Jobs are More Resilient to Automation

Jobs that were resilient to automation in the past decades, and those that are predicted to be harder to replace by AI, tend to require more cognitive and non-cognitive skills than other jobs, at least when comparing across jobs with similar wage levels. Result resonate surprisingly close when comparing automation patterns during the past decades of transformed labor demand to projections over the future. Most existing projections also suggest a similar relationship to very specific skill components as previous waves of automation. In particular, jobs that tend to employ workers with unusually high levels of social maturity are projected to continue to be more resilient than other jobs.

Implications

It is well-known that previous waves or automation have replaced jobs in the middle of the wage distribution. These results are often interpreted as a replacement of middle-skilled jobs. But our article uses very detailed direct skill assessments in cognitive and non-cognitive dimensions to show that this interpretation is misleading. Instead, replaced jobs are middle paid, but low skilled. Growing jobs in the bottom of the wage distribution are instead surprisingly skill intensive. Furthermore, the article shows that existing projections suggest that future trends have a similar relationship both to overall skills, and to detailed skill-components. Remarkably, this stability is particularly pronounced when using predictions from articles that explicitly emphasize that future automation will be very different. The results highlight the need for more direct evidence, rather than occupational projections, in order to understand how AI may alter the demand for labor.

Reference

Hensvik, L, & Skans, O.N. 2023. The Skill-Specific Impact of Past and Projected Occupational Decline. *Labour Economics* 81: 102326. doi.org/10.1016/j.labeco.2023.102326.

Research team

Oskar Nordström Skans, Uppsala University
Andrei Gorskov, IFAU
Christoph Hedtrich, Uppsala University
Georg Graetz, Uppsala University
Lena Hensvik, Uppsala University
Peter Fredriksson, Uppsala University

Contact

oskar.nordstrom_skans@nek.uu.se

Detecting Political Dogwhistles with AI and Linguistics

A dogwhistle is a communicative act intended to broadcast a message only understood by a select in-group while going unnoticed by others (the out-group). In political communication, dogwhistles are used to signal membership in an in-group with socially unacceptable political beliefs without offending a majority of citizens. For example, “inner city” used to be a dogwhistle in the USA that referred to negative stereotypes about African Americans. Our project, the Gothenburg Research Initiative for

Politically Emergent Systems (GRIPES), was able to employ a linguistic research tool, a lexical replacement task to members of the Swedish Citizens Panel to identify Swedish-language dogwhistles and their in- group and out-group meanings. We then used AI techniques to explore the development of these dogwhistles by comparing Swedish online forums: Flashback and Familjeliv, the former associated with radical discourses about immigration and its effect on society.

Implications

AI-based measures of lexical semantic change—quantification of the “drift” in dogwhistle meaning—show that the Swedish dogwhistles that we identified, such as återvandring (remigration) and berika (enrich, as in cultural enrichment), begin to experience more accelerated change in the “in-group” forum, Flashback, than in Familjeliv, in accordance with their appearance as dogwhistles. Additional work, in preparation for publication, shows a causal relationship between the appearance of dogwhistles in Flashback and its eventual appearance in mainstream discourse found in Familjeliv. This

work contributes to validating the project's novel theoretical model of dogwhistle development, the "life cycle" theory, a game-theoretic utility tradeoff between widespread recognition and in-group exclusivity. This is an important step towards building an AI-based automatic dogwhistle detector.

References

Sayeed, A., et.al. (2025). The utility of (political) dogwhistles–a life cycle perspective. *Journal of Language and Politics* 24(2):214-234. doi.org/10.1075/jlp.23047.say.

Lindgren, E., et al. (2024). Coded appeals and political gains: Exploring the impact of racial dogwhistles on political support. *Journalism & Mass Communication Quarterly*. doi.org/10.1177/10776990241280373.

Boholm, M., et al. (2024). Can political dogwhistles be predicted by distributional methods for analysis of lexical semantic change?. In *Proceedings of the 5th Workshop on Computational Approaches to Historical Language Change* pp. 144-157. doi.org/10.18653/v1/2024.lchange-1.14

Research team

Asad Sayeed, University of Gothenburg
Björn Rönnerstrand, University of Gothenburg
Elina Lindgren, University of Gothenburg
Ellen Breitholtz, University of Gothenburg
Gregor Rettenegger, University of Gothenburg
Max Boholm, University of Gothenburg
Robin Cooper, University of Gothenburg

Contact

asad.sayeed@gu.se

AI-based Language Models for Improving Diagnostic, Monitoring, and Outcomes of Depression and Anxiety

This project investigates whether depression and anxiety can be assessed with open-ended language responses analyzed with AI, and how the validity of these measures compares to rating scales. In particular, we have developed the Question-based Computational Language Assessment (QCLA) approach, where participants are asked to answer prompted open-ended questions about their mental health. The responses are quantified by large language models and machine learning is used to train to assess depression and anxiety. These results have shown that QCLA can assess depression and anxiety with high validity, possibly better than dedicated rating scales. We let participants read narratives related to depression, anxiety, harmony, and satisfaction emotions, and either summarize the emotions in word response or rating scales. The results showed that QCLA outperformed rating scales in the categorization accuracy of these emotions.

Implications

Mental health accounts for almost half of the sick leaves in Sweden, where anxiety and depression are the dominating disorders. Our project indicates that prompted language responses analyzed by AI can show higher accuracy in assessing depression and anxiety compared to state-of-the-art rating scales. This indicates that QCLA method can significantly improve the assessment of mental health, where early and more accurate assessment is known to lead to better outcomes. Language is also the dominating clinical way of assessing mental health, however, such assessment is currently made by human evaluators who are subject to biases and variability in the experience of clinicians.

These concerns can be overcome by the QCLA approach. The project has also laid down the foundation to use generative AI in the assessment and treatment of mental health and secured funding for future developments of AI-supported clinical tools.

References

Varadarajan, V. et al. (2024). ALBA: Adaptive Language-Based Assessments for Mental Health. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* Vol 1 Long Papers: 2466-2478, Mexico City, Mexico. Association for Computational Linguistics.

Sikström, S., et al. (2024). Question-based computational language approach outperforms rating scales in quantifying emotional states. *Communications Psychology* 2(1):45. doi.org/10.1038/s44271-024-00097-2

Kjell, O.N.E., et al. (2022). Natural language analyzed with AI-based transformers predict traditional subjective well-being measures approaching the theoretical upper limits in accuracy. *Scientific Reports* 12, 3918. doi.org/10.1038/s41598-022-07520-w.

Research team

Sverker Sikström, Lund university
Gergő Győri, Lund University
Mariam Mirström, Lund University
Rebecca Boehme, Lund University
Thibaud Agbotsoka, Lund University

Contact

sverker.sikstrom@psy.lu.se

Using Economic Games to Study the effects of Anthropomorphism in Robots and Chatbots

Do people behave more morally or generously when they know someone is watching? Does it matter if the observer is human or a robot? These questions lie at the heart of our first study, which investigated the audience effect—how the presence of an observer influences human behavior. Through economic games designed to simulate real-life financial decisions, we assessed participants' fairness, honesty, and generosity under three conditions: being observed by a human, a robot, or playing alone at a computer. Surprisingly, two of the games (the Ultimatum Game and the Mind Game) revealed no audience effect at all. For instance, participants were equally dishonest regardless of whether they were alone or observed by another person. However, in one game involving donations to charities (the Dictator Game), an audience effect did emerge: participants donated more generously when observed by a human.

Implications

The goal of our study was to identify effective methods for examining whether human-like AI can influence people's behavior. Since two of the games (the Dictator Game and the Mind Game) failed to demonstrate an audience effect, even with a human observer, these tasks may lack sufficient sensitivity to detect social influences from robots or chatbots as well. However, the Dictator Game might still be valuable for investigating how varying levels of anthropomorphism influence human generosity. Interestingly, participants reported feeling unaffected by the observer's presence, despite their behavior

clearly indicating otherwise. This underscores the value of behavioral measures like economic games, which capture effects overlooked by self- reports. Additionally, the humanoid robot's failure to elicit human-like responses indicates that anthropomorphism alone doesn't significantly alter behavior.

References

Stinkeste, C., et al. (2025). Comparing the Audience Effect of Anthropomorphic Robots and Humans in Economic Games. Under review. Available at dx.doi.org/10.2139/ssrn.5162896.

Stinkeste, C., et al. (2025). Manners Matter: How Robot Politeness Influences Human Risk-Taking and Social Perception. Under review.

Research team

Gabriel Skantze, KTH Royal Institute of Technology
Anna Dreber Almenberg, Stockholm School of Economics
Charlotte Stinkeste, KTH Royal Institute of Technology
Jonas Olofsson, Stockholm University

Contact

skantze@kth.se

AI and the Financial Markets

AI technology challenges well-established assumptions about how to organize and govern organizations, and more specifically the design and use of risk management frameworks and governance systems. When control systems become AI-based, the human social analysis that underlies a mutual understanding of the different parts of the business risks being lost. If part of the business ends up in a 'black box', this has negative consequences for accountability structures both internally in the business and externally. In addition, AI technology challenges human control over operations and decision-making. If a business moves to fully automated decision-making in any of its processes, this will most likely lead to an increased need for more control over AI-based decision-making. Increased control with AI thus leads to increased control of AI.

Implications

AI technology challenges must be embraced, not only because the EU AI Act requires it in relation to high-risk systems, but more generally so that we can achieve a balance between human-centric and algorithm-centric decision-making. Organizations, such as financial businesses, need to implement a framework for risk management in relation to the AI systems that are to be implemented and integrated into the business. The introduction of AI into an organization can have significant consequences for risk management and governance in terms of structure (roles, responsibilities, decision-making powers), processes (risk and performance management) and people (knowledge requirements, professional identity), which must be addressed.

Reference

Mahi, E., Crawford, J. & Strand, M. forthcoming. Regulation, Governance, and AI Systems. Iveroth et al (Eds). *Leading Digital Transformation*. Abingdon: Routledge.

Research team

Magnus Strand, Uppsala University
Andreas Kotsios, Uppsala University
Annina H Persson, KTH Royal Institute of Technology
Ensieh Mahi, Uppsala University
Jason Crawford, Uppsala University
Johanna Chamberlain, Uppsala University
Malou Larsson Klevhill, Uppsala University

Contact

magnus.strand@fek.uu.se

Predicting the Diffusion of AI-Applications

Our research shows that the diffusion of AI among consumers is different from that of earlier general-purpose technologies. A defining feature of the fastest spreading AI diffusion is that it is infused, i.e that it is integrated into existing products—such as YouTube's recommendation algorithm—rather than adoption as a standalone technology. As a result, end-users often don't make a conscious choice to use AI, but continue to use familiar services without needing to make an explicit decision to adopt AI. Instead, choices regarding infused AI are left to the organizations behind the apps. Moreover, using survey data, we find that the diffusion of AI applications (apps) can be predicted based on a limited set of factors. These can be distilled into two core dimensions: quality, defined as the perceived benefit of the apps, and allure, which captures the ease of use and tryability.

Implications

That all infused AI apps are optimized for certain goals but lack consumer-choice creates a lack of transparency with two implications: there is a need for regulations, and we need to understand the people behind the apps. Our research into the values of tech workers shows that they often harbor a strong sense of ideology with a mix of liberal and anti-establishment attitudes. As a result, it can be challenging for companies to develop AI-infused apps unless the apps' features align with these beliefs. Finally, by showing what AI apps will diffuse the fastest, we provide a valuable tool for prioritizing regulatory efforts. A first step for regulators would be to require transparency around the optimization goals of AI algorithms. By mandating disclosure, policymakers can ensure greater accountability and informed use among consumers.

References

Engström, E. and Strimling, P. (2020). Deep learning diffusion by infusion into preexisting technologies – Implications for users and society at large. *Technology in Society* 63:101396. doi.org/10.1016/j.techsoc.2020.101396.

Engström, E. et al. (2024). Comparing and modeling the use of online recommender systems. *Computers in Human Behavior Reports* 15:100449. doi.org/10.1016/j.chbr.2024.100449.

Selling, N. and Strimling, P. (2023). Liberal and anti-establishment: An exploration of the political ideologies of American tech workers. *The Sociological Review* 71(6):1467-1497. doi.org/10.1177/00380261231182522.

Research team

Pontus Strimling, Institute for Futures Studies
Emma Engström, Institute for Futures Studies
Irina Vartanova, Institute for Futures Studies
Jennifer Viberg Johansson, Linköping University
Kimmo Eriksson, Institute for Futures Studies
Niels Selling, Linköping University

Contact

pontus.strimling@iffs.se

The Global Governance of Artificial Intelligence

The US, China, and Europe approach the regulation of AI based on competing preferences, centered around three alternative models: the American market-driven model, the Chinese state-driven model, and the European rights-driven model. The US model places importance on protecting the innovation potential of its world-leading AI industry. This goal leads the US to pursue a lenient regulatory approach domestically and internationally. The Chinese model emphasizes the role of AI as a tool for economic growth and social control. This goal leads China to seek greater influence in global AI governance, while resisting any regulation that could constrain its use of AI for economic and repressive purposes. The EU model focuses on protecting individual rights, democratic values, and human-centric development of AI. This goal leads the EU to advocate global regulation of AI, building on the ambitions of the EU AI Act.

Implication

These competing preferences of the major powers have shaped emerging patterns of global AI governance. First, the different approaches of the US, China, and the EU have prevented convergence on one focal institution for global AI governance. Instead, the AI governance landscape is fragmented, containing a myriad of partly competing and overlapping initiatives. Second, the major power most interested in advancing AI regulation—the EU—has taken the lead in fostering global AI governance. Overall, the EU participates in a significantly larger number of global AI institutions than either China or the US. Third, the competing interests of the major powers tend to result in regulatory outcomes that reflect the lowest common

denominator. As a result of US and Chinese resistance to binding international regulation, almost all existing AI policies consist of non-binding guidelines.

Reference

Geith, J., Lundgren, M. & Tallberg, J. The Emerging Regime Complex for Artificial Intelligence. Under review for publication in *Global Studies Quarterly*.

Research Team

Jonas Tallberg, Stockholm University
Johannes Geith, Stockholm University
Magnus Lundgren, University of Gothenburg
Sonia Bastigkeit Ericstam, Stockholm University
Eva Erman, Stockholm University
Markus Furendal, Stockholm University
Mark Klamberg, Stockholm University

Contact

jonas.tallberg@statsvet.su.se

Wallenberg AI, Autonomous Systems and Software Program – Humanity and Society, WASP-HS

WASP-HS management team

Christofer Edling, Program director
Ericka Johnson, Graduate school and Deputy program director
Helena Lindgren, Deputy program director
Ingar Brinck, Management team member
Catherine Jo, Program coordinator
Francis Lee, Management team member

Virginia Dignum, Program director (2019-2023)
Christian Balkenius, Graduate school director (2019-2023)
Anna-Sara Lind, Management team member (2019-2025)

WASP-HS board

Kerstin Sahlin, Chair
Tora Holmberg, Vice chair
Kjell Asplund
Carolina Brånby
Pernilla Jonsson
Per Mickwitz
Lars Nielsen
Sven Strömqvist
Elisabeth Wåghäll Nivre

Hans Adolfsson, Vice chair (2019-2024)
Göran Blomqvist (2019-2021)
Christofer Edling (2019-2023)

wasp-hs.org
contact@wasp-hs.org

Wallenberg AI, Autonomous Systems and Software Program – Humanity and Society (WASP-HS) is the largest freestanding research program in the social sciences and humanities in Sweden to date, running from 2019 to 2031. Through a national graduate school, international postdoctoral fellowships, tenure track positions, international guest professorships, conferences, and collaborations the program builds research capacity across the humanities and social sciences.

This booklet presents research projects that were funded during the first half of the program, and that define the foundation upon which WASP-HS continues to build its approach in second half. The projects testify to the diversity of research questions that the AI transformation provokes and to the importance and relevance of studying technological change from a multidisciplinary perspective.

WASP—HS