

EDITORS

David Budtz Pedersen, Josephine Bøgen Theilgaard

Pathways to Research Impact

A Collection of Seven Impact Case Studies
from the Algorithms Data & Democracy Project



AALBORG
UNIVERSITY

add algorithms
data &
democracy

EDITORS

David Budtz Pedersen, Josephine Bøgen Theilgaard

CONTRIBUTORS

Sine Nørholm Just, Prins Marcus Valiant Lantz, Laura Anna Kocksch, Torben Elgaard Jensen, Theresia Veronika Rampisela, Maria Maistro, Christina Lioma, Johann Ole Willers, Ida Schrøder, Carl Stefan Roth-Kirkegaard, Rolf Hvidtfeldt and Ulrik Gensby.

Cover image: Jan Antonin Kolar / Unsplash

Portraits: Tobias Scavenius

Illustrations: Julia Massow

Layout: Design 8

Print: Printdivision

FUNDING

This publication is made possible with support from The Velux Foundation and Villum Foundation (grant no. 00037102)

SUGGESTED CITATION

Pedersen, D.B., & Theilgaard, J.B. (eds.) (2026). ADD Impact Reporting: A collection of seven impact case studies. Aalborg University: Aalborg.

This work is published under a [Creative Commons Open Access License CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) which permits reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator.

DOI: 10.54337/aau.pri.add

ISBN: 978-87-89383-61-3

Executive summary	5
Introduction	8
Impact Case Studies	
1. Stimulating a new kind of cybersecurity awareness: The development of a dilemma board game	12
2. Diagnostic AI. The case of the Danish mammography screening program	18
3. A qualitative evaluation of algorithmification in a Scandinavian NGO	24
4. Beyond control? Exposing and shaping the global surveillance industry	30
5. Fairness in recommender systems	36
6. How to lead AI adoption as a manager	44
7. Responsible use of AI in scientific advice	50



Executive summary

This report presents seven impact case studies from the first phase of Algorithms, Data & Democracy (ADD). Taken together, the cases showcase an interdisciplinary research programme already generating societal value across cybersecurity, healthcare, civil society, management, machine learning, international policy, and science advice.

ADD is producing various types of impact. The research teams involved in the project have contributed to a diverse and rich portfolio of pathways through which research informs practice, shapes debate and equips external actors to respond more thoughtfully to digitalisation and AI.

Across the seven impact cases it is clear that impact emerges through sustained collaboration with stakeholders rather than one-way dissemination.

In the first case, ethnographic research into cybersecurity was conducted together with SMEs and the Danish Business Authority, leading to a more constructive understanding of everyday cyber security practices. The ensuing 'dilemma board game' changed how stakeholders discuss security challenges.

In the second case, collaboration with Centre for Clinical Artificial Intelligence (CAI-X) and a broader set of professionals and patients produced practical recommendations for diagnostic AI in mammography screening centred on human oversight, context sensitivity, and co-creation.

In the third case, ADD researchers' engagement with the Danish Children's Rights organisation helped develop an ethical framework for AI-assisted counselling which informed later changes to how the AI-tool is used.

A second pattern is ADD's ability to translate complex research into usable formats. Several cases resulted in practical instruments such as games, workshops, policy briefs, managerial recommendations, and guidelines for evaluating fairness in recommender systems:

In the fourth case, the global surveillance industry's efforts to move beyond control are exposed – and guidance for how to reign in the industry are offered using a policy-facing platform to map global markets, support briefings for stakeholders, and inform debates on cyber proliferation and democratic governance.

In the fifth case, fairness in recommender systems is improved by identifying flaws in existing fairness metrics, proposing better measures, and offering guidance that can be used by researchers, developers, companies, and potentially regulators.

In the sixth case on managerial AI adoption, a collaboration with DJØF produced concrete recommendations for leaders responsible for implementing AI in organisations.

In the seventh case, research on AI-generated policy advice was translated into actionable solutions, benchmarks, and recommendations for decision-makers to ensure consistency, credibility, access, and human oversight.

In sum, the impact cases presented in the report indicate a number of pathways to impact – across a spectrum of outputs and outcomes covering strategic dialogues, agenda-shaping workshops and field-building platforms, artifacts, tools, and games.

The cases also indicate that fostering societal impact requires an active effort to build bridges between science and society, using knowledge mobilisation and knowledge exchange strategies, and thinking carefully about design, leadership and research culture.

In this capacity, the seven case studies show a variety of different forms of impact, which can be summarized by the following three overall categories:

1. Sense-making. Research that makes a difference to society contributes to conceptual and discursive changes; it moves along non-linear pathways and is exchanged dynamically between researchers and practitioners in the cases. Making sense of new research is more than dissemination: it requires active engagement and exchange of knowledge.

2. Productive interactions. Research that generates changes in organisations, people, and the economy often arises from well-functioning collaborations and partnerships. The selected cases show how impact is relational and interactional. By finding the right partners and creating collaborative 'shared' spaces, pathways researchers and stakeholders are established and consolidated.

3. Absorptive capacity. Research that leads to new and improved understanding, organisational practices, policy influence, or industrial services requires not only strong research capacities but also the presence of enabling conditions among lead users and stakeholders. Translation of knowledge and expertise is not a one-way street but requires an iterative process of building relationships, clarifying expectations, and allocating resources among partners to absorb and implement knowledge.

Seen through the lens of sense-making, productive interactions, and enhanced absorptive capacity, the seven impact cases presented in

the report indicate various forms of *intermediate* outcomes – rather than final 'hard' impacts. In some instances, there is clear uptake through collaboration, co-creation and dissemination; in others, the researchers have established foundations, informed partners, and created links and networks whose full effects are not yet seen.

The emphasis on *impact pathways* is a precondition for a research programme that is still very much evolving; it is also a more credible and realistic way of describing progress. Researchers rarely make a significant impact on practice or policy all by themselves. Rather, they collaborate, disseminate, and interact with various partners in society, policy, and industry with the objective of translating findings and knowledge into valuable organisational practices. On that basis, the ADD programme is well-positioned to deepen and widen its societal impact in the coming years.

Introduction

Algorithms, Data & Democracy (ADD) is a 10-year inter- and transdisciplinary research and outreach programme, running from 2021 to 2031. The programme aims to advance digital democracy, envisioning a society in which flourishing organisations, businesses, and democratic institutions are supported by digital technologies and in which technological developments are attuned to creating desirable outcomes for people and the planet. The ADD project aims to strengthen the capacity of society and citizens to evaluate technological developments and assess their influence on everyday life through participation in democratic discourse.

To realize this vision, the project focuses on the interrelations of sociotechnical developments and re-frames research on algorithms to include datafication, public perception, democratic governance and system-wide concerns about controversies and dilemmas involved in the development, deployment and regulation of AI. The

ADD project addresses one central concern: How may the algorithmic organisation of data engage, enlighten, and empower individuals, organisations, and institutions to act democratically?

In order to examine this research agenda, the six ADD project teams have approached digitalization and datafication as an essentially controversial force of change. Digitalization may lead to increased loss of accountability and trust; technology may accelerate populism, polarization, misinformation, and social conflicts; and AI may be used to enhance human endeavours, businesses, and social advancement. For these reasons, the project has worked to address the question of controversial algorithms from a multitude of interdisciplinary perspectives, including environmental, social, political, and economic perspectives.



The six subprojects

Digital organisation of datafied health led by PI Sine Nørholm Just, Roskilde University

Cybersecurity, privacy, and controversy mapping led by Co-PI Torben Elgaard Jensen, Aalborg University

Public administration and prediction led by Co-PI Helene Friis Ratner, Aarhus University

Finance and transparency led by Co-PI Leonard Seabrooke, Copenhagen Business School

Machine Learning and bias led by Co-PI Christina Lioma, University of Copenhagen

Innovation and democracy led by Co-PI Alf Rehn, University of Southern Denmark

Knowledge Broker Unit led by Prof. David Budtz Pedersen, Aalborg University

Outreach partner led by Strategic Director Lisbeth Knudsen, Think Tank Monday Morning

The ADD project has explored the underpinning drivers of controversies and the possibility of turning them into public matters of concern by establishing a number of subprojects and research questions, e.g., how can data and algorithms be used to enlighten and engage citizens to strengthen democracy? And how can research offer explanations of – and solutions to – societal problems by studying controversies as these play out on digital platforms and are shaped by digital technologies?

The work conducted by the ADD teams has shown how developments in the ever-changing technological, political, and economic landscape call for increased interdisciplinary collaboration and societal engagement. The first five years of the project have led to significant advances and new research findings across a wide spectrum of activities, reports, events, and publications.

The research teams have demonstrated a significant engagement with the transformative effects of algorithmic governance in a range of social domains, engaging with different communities, stakeholders and audiences. Research outputs have been characterized by strong quality, with publications appearing in high-ranking, specialized, and cutting-edge journals. Partner organisations, stakeholders, and the public have been actively engaged through a variety of outreach efforts and using a range of formats. The ensuing pathways to impact and social contributions are the subject of this report.

Mapping the societal impact of ADD

Beyond the academic outputs of the ADD project, this report presents a series of examples of societal engagements and impacts originating from programme activities. Since its inception in 2021,

the ADD programme has operated with a number of Impact Goals distributed across the six different subprojects and the Knowledge Broker Unit. This publication presents the first five years of impact reporting with one selected case from each subproject plus one case from the Knowledge Broker Unit. Evidently, other stories could have been told, and other cases could have been selected. By selecting these cases, we have chosen to highlight leading examples rather than focusing on the steady stream of micro-impacts emerging from the subprojects.

Research impact is created as the result of various pathways and enabling conditions.¹ Importantly, the case studies presented in this report do not constitute *final* impacts but are based on a limited time frame and should be considered as *intermediary* steps towards establishing longer-term impacts. The methodology of impact case studies is based on the assumption that creating and evidencing impact should be guided by an open, experimental, explorative and creative approach.

The case studies focus on research impact as *contributions* to broader societal processes of change, including the integration and infusion of research, knowledge and expertise in policymaking, civil society, industry, and practice.² Impact occurs when research teams generate contributions that non-academic stakeholders use to shape their practice, business, attitudes, or behaviour.³ Impact is reached through interactions and connections where information, expertise or ideas are exchanged between a research group and an external stakeholder, beneficiary or co-producer. This involves both direct and indirect interactions that lead to conceptual, relational, instrumental or material changes co-shaped by research contributions.⁴

1 Pedersen, D.B. & Hvidtfeldt, R. (2024). The Missing Links of Research Impact. *Research Evaluation*, 33, Artikel rvad011 <https://doi.org/10.1093/reseval/rvad011>

2 Phipps, D., Poetz, A. & Johnny, M. (2022), "Demonstrating Impact – Considerations for Collecting and Communicating the Evidence of Impact", Kelly, W. (Ed.) *The Impactful Academic (Surviving and Thriving in Academia)*, Emerald Publishing Limited, Leeds, pp. 107-135.

3 Lima, G. & Bowman, S. (2022). *Researcher Impact Framework: Building Audience-Focused Evidence-Based Impact Narratives*. Trinity College Dublin. [pdf] Dublin. <https://doi.org/10.25546/98474>

4 van den Akker W., Spaapen J., Maes K. (2017) *Productive Interactions: Societal Impact of Academic Research in the Knowledge Society*. LERU Position Paper.

The collected case studies include changes to attitudes, awareness, behaviours, capacities, opportunities, performances, policies, practices, processes and the improved understanding of audiences, beneficiaries, and organisations. Examples of impact pathways include research-based products, activities (such as talks, debates, discussions), and participation in workshops, meetings, events, advice, gamification, and collaborative problem-solving, etc.⁵

Pathways to impact

Societal impact is the result of consecutive contributions and iterative interactions that are part of the complex set of causes that collectively result in change and benefits to society. Research impact can be traced through mapping a larger web of interactions, collaborations, partnerships etc. by which research leads to impact. And impact can be captured more locally by using qualitative impact narratives that acknowledges the different factors influencing the dissemination and use of knowledge. We have chosen this qualitative approach for this report.

Importantly, impact pathways differ from discipline to discipline and are often specific to the context of application and field of study. What constitutes impact in one domain, may be seen as irrelevant in another domain. Hosting an exhibition or creating a digital artifact may lead to impact in one domain, while it would make little sense in other domains. The same with policy advice, learning tools, participatory experiments, etc. There is no single process or checklist for how to link research and knowledge translation to wider outcomes, which is why a framework for impact cases must be adaptive to the unique impact pathways and contexts associated with the research.

To capture the contextual and variable nature of impacts across the ADD project, the impact cases have been crafted by formulating impact narratives. This allows for each subproject to showcase and explain how their research outcomes have contributed to societal impact. An impact narrative should clearly state the connection of research to societal change by identifying impact pathways, and by presenting underlying evidence, for example, of the use of materials, sources, references, etc. by external stakeholders and lead-users.

The cases have been produced with input and materials from each of the ADD subproject teams, following a general case template suitable for presenting activities and outputs within the broader context of the ADD research. Inspiration has been drawn from e.g. Research Impact Framework;⁶ Research Excellence Framework;⁷ Impact and Engagement Case Study Guidelines;⁸ and the Research Contribution Framework.⁹

As there is no one-size-fits-all, we have provided guidance to each subgroup to write up the cases and establish connections between the specific context of their research and the objectives and goals of the programme. The overall aim has been to support the creation of informative narratives that allow for showcasing the variety and complexity of impacts.

In sum, the impact case show a portfolio of contributions through which research helps to better understand problems, develop responses, and improve decision-making. That is an important finding in itself. The report suggests that the ADD project is making a meaningful contribution to society, while also underlining the need for continued attention to how impact is supported, evidenced, and cultivated.

5 Pedersen, D. B., & Israelsson, P. A. (2021). Ontologies and Tools for Mapping Impact: Data Collection, Analysis and Visualization. (pp. 1-31). DEFF OPERA Project Report Series.

6 Lima, G. & Bowman, S. (2022). Researcher Impact Framework: Building Audience-Focused Evidence-Based Impact Narratives. Trinity College Dublin. Dublin.

7 Reichard, B., Reed, M.S., Chubb, J. et al. (2020) Writing impact case studies: a comparative study of high-scoring and low-scoring case studies from REF2014. *Palgrave Commun* 6, 31.

8 Phipps, D., Poetz, A. & Johnny, M. (2022), "Demonstrating Impact – Considerations for Collecting and Communicating the Evidence of Impact", Kelly, W. (Ed.) *The Impactful Academic (Surviving and Thriving in Academia)*, Emerald Publishing Limited, Leeds, pp. 107-135.

9 Morton, S. (2015). Progressing research impact assessment: A 'contributions' approach. *Research Evaluation*, 24(4), 405-419. doi:10.1093/reseval/rvv016

Acknowledgements

We thank the contributors from the ADD Project who provided thoughtful feedback that helped improve the clarity, practicality and quality of this collection of case studies. Their insights, activities, and results have been instrumental to this collective product. We are also grateful to the reviewers beyond the working group for their constructive suggestions: Ulrik Gensby, Rolf Hvidtfeldt, and Olivia Karlsson Hansen. Their engagement broadened perspectives and strengthened the final publication.

1.



*Impact leads:
Laura Kocksch and
Torben Elgaard Jensen,
Aalborg University*

**Stimulating a new
kind of cybersecurity
awareness:**

**The development of a
dilemma board game**

Subproject: Critical infrastructures,
privacy, and cybersecurity

Context

Cybersecurity poses a technological, socio-technical and organisational challenge. Prior research has investigated tactics for practising cybersecurity in larger organisations, leaving cybersecurity practices of small- and medium-sized enterprises (SMEs) less explored despite their central role in the global economy and local communities.¹⁰ According to the Danish Cyber and Information Security Strategy 2022-2024, only 40% of Danish SMEs have an inadequate level of digital security, and SMEs are identified as a key vulnerability in the digitised Danish society.¹¹ The Danish Business Authority has acknowledged a repeated difficulty in reaching a large and diverse audience of SMEs through traditional channels and information campaigns, even though SMEs constitute almost 99 per cent of all Danish companies.¹²

Several surveys and interview studies have indicated a worrying lack of basic cybersecurity measures in many SMEs, leading to the suspicion that a substantial number of SMEs may simply take no interest in cybersecurity.¹³ In Denmark, SMEs comprise more than 90 % of all businesses and more than 50% of private sector employment.¹⁴

Laura Kocksch and Torben Elgaard from Aalborg University's Techno-Anthropology Lab (TANT-Lab) have investigated the practices of SMEs rather than evaluating them against defined standards set by the security industry and national authorities. The aim of the project was to understand how SMEs deal with cybersecurity, guided by the following research questions:

- How do SMEs organise cybersecurity responsibilities?
- What types of everyday knowledge do SMEs have of cybersecurity?
- How do SMEs handle cybersecurity issues in everyday practice?
- How can the specific situation of SMEs be taken into account when facilitating future dialogue and communication?

The study

To investigate how cybersecurity is practised in SMEs, the research team conducted a multi-sited ethnographic study consisting of visits to 30 SMEs across Denmark spanning a variety of products and services. The study included interviews with people responsible for cybersecurity, observations of on-site security practices, and stakeholder workshops. The multi-sited ethnographic study has given new insights into the cybersecurity practices, circumstances, and rationalities of Danish SMEs. To analyse cybersecurity practices, the research team focused on the mundane art of practising cybersecurity and developed the notion of dilemma thinking to articulate the types of practices SMEs use to address cybersecurity issues, as well as to describe how SMEs manage to deal with many relevant and important cybersecurity issues they face in their everyday.

By applying this term to the analysis of cybersecurity practices, the study offers a unique insight into SMEs' own solid logic for handling everyday cybersecurity issues. Furthermore, the study articulates what the authors call 'good' organisational reasons for 'bad' cybersecurity.

10 Kocksh, L. & Jensen, T. E. (2024). The Mundane Art of Cybersecurity: Living with Insecure IT in Danish Small- and Medium-Sized Enterprises. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2), 1-17. <https://doi.org/10.1145/3686893>

11 Kocksh, L. & Jensen, T. E. (2024). The Mundane Art of Cybersecurity: Living with Insecure IT in Danish Small- and Medium-Sized Enterprises. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2), 1-17. <https://doi.org/10.1145/3686893>

12 Interview with Laura Kocksch at Monday Morning (2024). <https://algoritmer.org/exit-interview-med-laura-kocksch-smver-har-bedre-cybersikkerhedspraksis-end-hvad-man-umiddelbart-tro/>

13 Kocksch, L. & Jensen, T. E. (2023). "Good" organizational reasons for "bad" cybersecurity (2023). Research Report. Aalborg University.

14 Kocksh, L. & Jensen, T. E. (2024). The Mundane Art of Cybersecurity: Living with Insecure IT in Danish Small- and Medium-Sized Enterprises. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2), 1-17. <https://doi.org/10.1145/3686893>



Danish Business Authority (later the Agency for Digital Government)

In August 2021, the research team at TANTLab Aalborg University was contacted by the Business Forum for Digital Security and the Danish Business Authority's office for cyber- and information security (the office moved later to the Agency for Digital Government) with the intention of establishing a research collaboration aiming at conducting an in-depth qualitative investigation of cybersecurity practices of SMEs. The aim of the collaboration was also to supplement and challenge the prevailing perception of cybersecurity in SMEs by providing evidence of everyday cybersecurity practices.

The Agency for Digital Government's overall work is to strengthen companies' digital safety levels through various channels, e.g., through information campaigns. As SMEs constitute over 90% of all companies in Denmark, the agency was interested in how cybersecurity practices are working in SMEs.¹⁵ They have a lot of quantitative data on SMEs but no systematic qualitative data on how everyday cybersecurity is practised and integrat-

ed into the small- and medium-sized companies. The agency's goal in this collaboration was to gain new insights and inspiration for tackling and securing the digital safety of SMEs.

Pathways to impact

The subproject's pathways to impact are grounded in the collaborations with the Danish Business Authority and productive engagement with the small- and medium-sized enterprises across Denmark. A key impact pathway was the development of a dilemma board game used in stakeholder workshops.

Building relationships and transforming cybersecurity dialogue

A core driver of the project's impact was the strong, trust-based collaboration between the research team, the Danish Business Authority, and the 30 small- and medium-sized enterprises (SMEs). This engagement not only enabled the collection of in-depth empirical data but also shaped how cybersecurity is understood, discussed, and communicated both by Danish authorities and cybersecurity experts.

¹⁵ Kocksch, L. & Jensen, T. E. (2024). The Mundane Art of Cybersecurity: Living with Insecure IT in Danish Small- and Medium-Sized Enterprises. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2), 1-17. <https://doi.org/10.1145/3686893>

Collaborative engagement with key stakeholders

From the outset, the research team adopted a partnership-oriented approach, positioning themselves not as external evaluators but as partners in collaboration with the Danish Business Authority.

As written above, the Danish Business Authority reached out to the TANTLab, inviting the research team to partner to get a picture of the landscape of cybersecurity practices in SMEs. The research team's approach of not promising or claiming that they would have all the answers laid the groundwork for a constructive partnership with the Danish Business Authority. The partnership was evident in that an office at the Danish Business Authority took responsibility for the very significant task of finding, contacting, and arranging dates for the researchers' visit to the 30 SMEs included in the study. The Business Authority was also instrumental in organizing a stakeholder workshop at the end of the project. Throughout the project, several meetings were held between the TANTLab researchers and their collaborators at the Business Authority to discuss preliminary findings, potential ways to specify the study's focus, and the study's relevance to the Business Authority's processes.

The 30 SMEs became active contributors, and the collaboration with the research team showed an engagement from the SMEs' side, which was crucial for obtaining relevant information. Their willingness to share detailed accounts of their cybersecurity practices yielded rich, grounded insights that challenged conventional narratives of cybersecurity practices in SMEs.

The willingness to engage was facilitated by the research team's non-judgmental approach, which emphasised understanding over compliance. This became evident in the way that the researchers experienced some of the people in the SMEs asked them to "not judge" them as they initially saw the researchers as experts who were there

to judge and evaluate their cybersecurity practices.¹⁶ This shows that SMEs are used to being judged by authorities for not strictly adhering to cybersecurity guidelines and rules. Knowing that the research team was not there to judge, trust grew, and the SMEs shared detailed accounts of their cybersecurity practices, revealing rich, grounded insights challenging conventional narratives. The collaboration with the SMEs yielded five key insights that help reshape the understanding of how everyday cybersecurity is practised in SMEs:

1. SMEs care about cybersecurity but have intricate practical challenges that would even be hard to crack for experts.
2. SMEs build with what they have locally (not what they could have).
3. SMEs build on local expertise, not cybersecurity professionals.
4. Fixing things is easy; living with almost broken things is hard.
5. A less moralistic conversation is needed.

These insights challenge dominant assumptions about SMEs' inadequate level of cybersecurity and underscore the importance of designing communication strategies that reflect the everyday lives of SMEs, not how their everyday cybersecurity is expected to work.

Transformative use of Dilemma Games

To share findings and initiate practical reflection, the research team designed participatory dilemma board games, presented at the final stakeholder workshop attended by 22 representatives from public authorities, trade unions, employer organisations, and SMEs.

The dilemma games presented participants with realistic cybersecurity dilemmas present in the empirical material. Players used colour-coded "resource cards" representing technical fixes,

16 Kocksh, L. & Jensen, T. E. (2024). The Mundane Art of Cybersecurity: Living with Insecure IT in Danish Small- and Medium-Sized Enterprises. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2), 1-17. <https://doi.org/10.1145/3686893>

management changes, people involvement, and knowledge gathering to explore possible solutions, and were moreover encouraged to add their own “resource cards” in order to solve the problem.

The significant outcome of the dilemma games was how experts and stakeholders engaged actively in discussing the different dilemmas, reflecting on how the dilemmas were tackled in their own organisations while gaining insights from how dilemmas could be tackled differently. The stakeholders’ takeaways from the workshop were the following opportunities for facilitating and improving the dialogue with SMEs about cybersecurity issues:

1. Engaging formats

The game’s format should be developed further and used to drive and facilitate cybersecurity dialogues with SMEs.

2. New collaborators

The stakeholders identified key actors in SMEs who need to be in the centre of attention for cybersecurity advice and dialogue: Accountants and board members.

3. New approach and understanding

A general need to customise communication and take the diversity of SMEs into account.

By ending the presentation of insights with a hands-on game, the project disrupted expert-driven discourses and created a space for more inclusive, situated reflections on how cybersecurity is practised.

Policy and communication: “good” organisational reasons for “bad” cybersecurity

The research collaboration between TANTLab and the Danish Business Authority culminated in the publication of the report *“Good” Organizational Reasons for “Bad” Cybersecurity*, providing the authority with qualitative evidence and five actionable recommendations for improving cybersecurity communication for SMEs:

1. Involving local cybersecurity figures in cybersecurity campaigns.

2. New vocabulary is needed to relate to everyday cybersecurity knowledge.
3. Understand “good” practical reasons for “bad” cybersecurity.
4. SMEs are heterogeneous; hence, communication tactics should be customized.
5. Moral high grounds have a negative effect on SMEs.

These findings and recommendations have directly influenced the authority’s approach towards cybersecurity practices for SMEs. In an interview with the think tank Monday Morning, authority’s representatives praised the project for revealing how cybersecurity is embedded in the everyday routines of SMEs and for expanding their understanding of roles and responsibilities within them. To increase awareness of the everyday cybersecurity challenges SMEs face, the authority has launched a LinkedIn campaign highlighting everyday cybersecurity dilemmas focused on everyday cybersecurity dilemmas based on the report’s findings.

In the interview, the authority also highlights its focus on developing new digital tools for SMEs to self-assess their cybersecurity levels, with tailored recommendations based on the SMEs’ responses. Furthermore, the authority acknowledged the need for alternative communication formats targeted at SMEs, rather than only big companies with cybersecurity units.

Conclusion

The case study shows that cybersecurity is neither absent nor ignored by SMEs. Rather, it is actively practised, though often in informal, locally adapted ways that fall outside the scope of traditional frameworks and therefore remain under-recognised by authorities and in their communication efforts. Viewing SMEs as spaces where cybersecurity is actively negotiated opens new opportunities to rethink how communication strategies and tools are designed and delivered.

The project has contributed to a new vocabulary and set of empirical insights that support more constructive, relevant, and resonant cybersecurity communication strategies targeted at

small- and medium-sized enterprises. Crucially, the study challenges the use of moralising narratives or “shaming” tactics, which often alienate SMEs and hinder meaningful engagement and communication.

A key takeaway from the project is the importance of partnering with stakeholders who are open to alternative approaches and can appreciate the complexity of empirical insights. Such collaborations lay the groundwork for impactful, societal interventions – enabling the ADD project to generate tangible, context-specific outcomes.

To sustain and expand the project’s impact, the research team is exploring ways to further develop and publicly share the dilemma game as a practical engagement tool. By inviting diverse stakeholders into participatory spaces, such as the dilemma games, the project has fostered an alternative dialogue rooted in real-world practices rather than formalised standards targeted at big companies. This shift represents a more inclusive, pragmatic approach to strengthening cybersecurity culture across the SME landscape.

References

Kocksch, L. & Jensen, T. E. (2023). “Good” organizational reasons for “Bad” cybersecurity (2023). Research Report. Aalborg University. Accessible from the homepage of Digital Security, The Danish Resilience Agency: <https://www.sikkerdigital.dk/virksomhed/publikationer/good-organizational-reasons-for-bad-cybersecurity>

Kocksch, L. & Jensen, T. E. (2024). The Mundane Art of Cybersecurity: Living with Insecure IT in Danish Small- and Medium-Sized Enterprises. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2), 1-17. <https://doi.org/10.1145/3686893>

Interview with Torben Elgaard Jensen from TANTLab and Eva Elisabeth Roland from the Agency for Digital Government, moderated by ADD senior advisor Jakob Kaastrup Sørensen (05.07.2023): <https://algoritmer.org/cybersikkerhed-i-smver-interview-med-digitaliseringsstyrelsen-og-tantlab/>

Interview with Laura Kocksch at Monday Morning (2024): <https://algoritmer.org/exit-interview-med-laura-kocksch-smver-har-bedre-cybersikkerhedspraksis-end-hvad-man-umiddelbart-tror/>

Campaign on LinkedIn in collaboration with the Agency for Digital Government (June 2024): <https://algoritmer.org/add-projektet-bi-draget-til-ny-kampagne-for-cybersikkerhed-hos-danske-smver/>



Read more about the subproject here

Kocksch, L. & Jensen, T. E. (2023): [“Good” Organizational Reasons for “Bad” Cybersecurity: Ethnographic Study of 30 Danish SMEs](#)

Jensen, T. E., Kocksch, L. & Wagenknecht, S. (2026): [The cybersecurity dilemma game: moving cybersecurity beyond solutionism](#)

Kocksch, L. & Jensen, T. E. (2025): [The Cybersecurity Dilemma Game: Collaborative Boardgame for Organizational Cybersecurity](#)



Impact leads:
*Sine Nørholm Just and
Marcus Lantz,
Roskilde University*

2.

Diagnostic AI. The case of the Danish mammography screening program

Subproject: Digital organisation of
datafied health

Context

Studies of the implementation of AI technologies in organisational contexts consistently find that AI solutions serve strategic goals only when humans collaborate with algorithmic agents,¹⁷ establishing sociotechnical relationships of collegiality rather than control or competition.¹⁸ However, achieving such collegial relationships has also proven to be a main challenge; it is simply difficult for humans to get to know their algorithmic colleagues, leading to technical domination or social rejection rather than sociotechnical collaboration.¹⁹ Thus, across the various use cases, one should not only consider potential technological gains in terms of increased efficiency, accuracy, etc., but also take into account what might be termed ‘human pains’ – i.e., reasons why different stakeholder groups might fail to understand that technological innovation is needed or could even object to this claim. In other words, one must begin by assessing the situation, asking: What is the problem that AI might solve? How could AI solve the problem? Why is it a good solution? And what are the drawbacks of and reservations

to this solution?²⁰ Ultimately, when is the implementation of AI technologies a good idea? And how can it also become a good process?

Broadly speaking, the healthcare sector is in many ways well-suited to the implementation of AI systems, not only because of alignment between algorithmic and diagnostic rationalities but also because of broader societal preconditions like changing demographics and limited resources. In the context of the Danish public healthcare sector, this increases political pressure for reform of the sector and for accelerated use of innovative solutions, not least because there is significant public support for and trust in the implementation of AI in the context of healthcare in Denmark.²¹ The promises of innovation include cutting resources. Specifically, the development of treatment, improved well-being for staff and the freeing of human resources. In sum, the development of AI technologies for clinical purposes is raising hopes for better and more efficient diagnoses and treatments.

17 For an overview, see Raftopoulos & Hamari, 2023

18 Meijer, A., Lorenz, L. & Wessels, M. (2021): Algorithmization of bureaucratic organizations: Using a practice lens to study how context shapes predictive policing systems. *Public Administration Review*, 81(5): 837-846.

19 Faraj, S., Pachidi, S. & Sayegh, K. (2018): Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1): 62-70.
Bader, V. & Kaiser, S. (2017): Autonomy and control? How heterogeneous sociomaterial assemblages explain paradoxical rationalities in the digital workplace. *Management Review*, 28(3): 338-358.

20 Gulbrandsen, I. T. & Just, S. N. (2024): Artificial intelligence in organizational communication: Challenges, opportunities, and implications. In: Ndlela, M. N. (ed.), *Organizational Communication in the Digital Era. Examining the Impact of AI, Chatbots, and Covid-19* (pp. 51-77). Cham: Palgrave Macmillan.

21 Ploug, T., Sundby, A., Moeslund, T. B., Holm, S. (2021): Population preferences for performance and explainability of artificial intelligence in health care: Choice-based conjoint survey. *Journal of Medical Internet Research*, 23(12): e2611.

The field of radiology is particularly affected by a shortage of human resources, and technological advances, spearheaded by AI image recognition tools, promise to solve the challenge of delivering sound diagnoses to more patients with fewer human radiologists at hand.²² The Danish mammography screening program, for which all women above the age of 50 are eligible, is a key example. Given changing demographics and limited human resources, the program is struggling to remain successful. Hence, it is not surprising that stories of imminent technological advances have reached the Danish public, raising hopes and expectations among medical professionals, patients, and politicians, while also causing some experts to call attention to the nuances that risk being lost in the heat of anticipation. The point, then, is not to let the sense of urgency overshadow the need for analysis.

The study

The case is part of the ADD subproject on digital health, based at Roskilde University and revolving around the digital organisation of personal and public health. Questions of decision-making that involve datafied health, including ethical concerns about data protection and privacy as well as human autonomy, are at the core of the subproject. Specifically, this case was part of Prins Marcus Valiant Lantz's postdoc project on automated decision-making in healthcare.

Seeking to contribute to an understanding of how human-AI collaboration may be organized to achieve societal goals, the research focused on the strategic decision to develop and implement clinical AI within the Danish mammography screening program. The case is critical, as demographic changes and limited access to human experts suggest that the program will only continue to be successful if it is technologically enhanced. The purpose of the case was to offer best practices for implementing diagnostic AI by exploring principles of organizing explainable AI

diagnostics. This case highlights three principles of best practices for implementing diagnostic AI:

1. Begin from nuanced versions of the debate on AI.
2. Create solutions with all relevant stakeholders: Medical doctors, AI developers, patients, and their relatives.
3. Make use of emerging knowledge from ongoing AI projects.

Collaboration with the Centre for Clinical Artificial Intelligence (CAI-X)

During his postdoc, Marcus established a close collaborative relationship with the Centre for Clinical Artificial Intelligence (CAI-X).

CAI-X is a joint centre between Odense University Hospital (OUH) and two faculties at the University of Southern Denmark (SDU): The Faculty of Engineering and the Faculty of Health Sciences. The purpose of the centre is to bridge technology and healthcare and "to ensure full utilisation of intelligence to benefit patients and staff of the hospital".²³

Given the collaborative nature and research design of the project, methods include not only ethnographic approaches such as observations and interviews but also forms of action research, most notably the organisation of workshops with different stakeholder groups.

Pathways to impact

The foundation for impact pathways was built in the close collaboration with CAI-X and the continuous engagement with a broad range of stakeholders.

22 Olsen, T. L., Martensen, M. M., Yding, H. & Futtrup, L. R. (2023): Minister vil bruge kunstig intelligens og rollebytte til at løse mangel på radiologer og radiografer. DR, <https://www.dr.dk/nyheder/indland/minister-vil-bruge-kunstig-intelligens-og-rollebytte-til-loese-mangel-paa-radiologer>

23 <https://cai-x.com/about-cai-x>

Building relationships and creating context-sensitive insights

Through close collaboration with CAI-X, Marcus explored how diagnostic AI was being developed and implemented differently across Denmark. While the Capital Region adopted a commercially developed AI tool due to pressing clinical demands, such as long waiting lists and radiologist shortages, the Region of Southern Denmark followed a more locally driven, research-informed path. This comparison illuminated the importance of adapting AI strategies to local healthcare contexts – a key insight stemming directly from the collaboration.

Crucially, insights were not developed in isolation but co-produced with clinical researchers, computer scientists at CAI-X, public health officials, patients, and members of the public. Engagements such as workshops, conferences, and panel debates ensured that a wide range of voices informed the research. This co-creation approach enhanced the legitimacy and applicability of the findings.

Enhancing capacity through stakeholder workshops

By bringing the different stakeholders together, workshops served as a core method for generating real-time impact. They enabled participants to share perspectives, learn from one another, and reshape their understanding of diagnostic AI implementation. These workshops were more than knowledge-sharing sessions; they became active arenas for capacity-building and reflection, laying the groundwork for key policy recommendations and insights that participants could immediately apply in their respective domains.

For example, participants came to recognise the need for a nuanced, context-specific approach to AI implementation rather than adopting one-size-fits-all solutions based on past implementations elsewhere. They also developed a deeper appreciation of the socio-technical dynamics

involved in introducing diagnostic AI, including the professional and ethical relationship between humans and technology.

Policy recommendations

The project culminated in a set of co-created policy recommendations for implementing diagnostic AI in the Danish public healthcare system. These are outlined in the policy brief *Healthy scepticism and emerging optimism – A human perspective on artificial intelligence in health-care*, jointly produced by the research team and CAI-X.

The key recommendations include:

1. Begin from nuanced versions of the debate on AI.
2. Create solutions with all relevant stakeholders: Medical doctors, AI developers, patients, and their relatives.
3. Make use of emerging knowledge from ongoing AI projects.

These recommendations represent a paradigm shift from top-down technology implementation to participatory, practice-informed innovation. Patients' perspectives, such as the comment "*It's a really good tool, if there's a human involved all along*" [translated from Danish] support the importance of human oversight and relational trust in the deployment of AI systems.

The theme of human-technology relation is further explored in the book chapter *The ethos of automation*,²⁴ which theorises the replacement of interpersonal trust by sociotechnical reliability when automation enters the healthcare space.

The recommendations are elaborated in the policy brief, co-created between the research team and CAI-X. Together, the recommendations lay the foundation of changing the implementation of diagnostic AI in mammography screening – and in healthcare, more broadly.

24 Lantz, P. M. V. & Just, S. N. (2025): The ethos of automation: Strategy-as-rhetoric in the development of trustworthy clinical AI. In: Hess, A. & Kjeldsen, J. E. (eds.), *Ethos, Technology, and AI in Contemporary Society. The Character in the Machine* (pp. 278-298). New York: Routledge.



Broadening influence through public engagement

To expand the reach and impact of the research, the team engaged in multiple public and professional outreach formats. The research team and CAI-X joined forces and discussed the research findings and their relevance for national health policy at a webinar at the think tank, Monday Morn.²⁵ Key concerns raised in the discussion included data governance, clinician-AI collaboration, and the central role of patients in technology, highlighting the following aspects:

- Insecurity of how and where data will be stored.
- The collaboration between professionals and technology has two dimensions: the direct interaction between professionals and technology, and the professionals' mediating role between patients and technology.
- The importance of including patients and clinicians in the AI development team, as it is not meaningful to leave out the context of implementation in the development phase.

Furthermore, an op-ed in the Danish Medical Journal, a tailored magazine for medical professionals, was published, translating the research insights into a format accessible to clinicians and reinforcing the practical importance of inclusive, context-aware AI implementation.

By disseminating the research across different platforms and for diverse audiences, the project significantly broadened its influence and contributed to shaping a more ethically grounded, practically feasible, and socially informed pathway for diagnostic AI in healthcare.

Conclusion

The collaboration between Roskilde University and CAI-X has been defined by a dialogical and co-creative partnership, where workshops and joint inquiry formed the core of both the research process and the development of policy recommendations. Thus, the project has demonstrated how impact can be achieved through collaboration, and it emphasises the importance of considering context sensitivity and sustained stakeholder engagement.

Rather than treating stakeholders as passive consultees, the project embraced the principle of *"nothing about us without us"*, embedding active stakeholder participation at most stages of the research process. This model of inclusive, participatory knowledge production has proven critical for addressing the complex, context-specific challenges of implementing diagnostic AI in mammography screening.

Most importantly, the project has contributed to reframing diagnostic AI implementation as a so-

25 ADD webinar: <https://algoritmer.org/nyt-add-webinar-og-policy-brief-om-kunstig-intelligens-i-sundhedsvaesenet/>

cio-technical challenge, where success depends not only on technical efficiency but on trust, human oversight, and meaningful inclusion of those most affected by the implementation.

Looking ahead, the collaboration opens several avenues for further impact. One is to revisit the implementation landscape to assess whether and how the recommendations are being applied in real-world mammography screening programmes. This would offer insights into how evolving public debate and policy developments around AI are affecting practice. Another valuable and interesting direction is to investigate how trust can be fostered not only between patients and professionals, but also between humans and diagnostic technologies.

References

Bader, V. & Kaiser, S. (2017): Autonomy and control? How heterogeneous sociomaterial assemblages explain paradoxical rationalities in the digital workplace. *Management Revue*, 28(3): 338-358.

Faraj, S., Pachidi, S. & Sayegh, K. (2018): Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1): 62-70.

Gulbrandsen, I. T. & Just, S. N. (2024): Artificial intelligence in organizational communication: Challenges, opportunities, and Implications. In: Ndlela, M. N. (ed.), *Organizational Communication in the Digital Era. Examining the Impact of AI, Chatbots, and Covid-19* (pp. 51-77). Cham: Palgrave Macmillan.

Lantz, P. M. V. & Just, S. N. (2025): The ethos of automation: Strategy-as-rhetoric in the development of trustworthy clinical AI. In: Hess, A. & Kjeldsen, J. E. (eds.), *Ethos, Technology, and AI in Contemporary Society. The Character in the Machine* (pp. 278-298). New York: Routledge.

Meijer, A., Lorenz, L. & Wessels, M. (2021): Algorithmization of bureaucratic organizations: Using a practice lens to study how context shapes predictive policing systems. *Public Administration Review*, 81(5): 837-846.

Olsen, T. L., Martensen, M. M., Yding, H. & Futtrup, L. R. (2023): Minister vil bruge kunstig intelligens og rollebytte til at løse mangel på radiologer og radiografer. DR. <https://www.dr.dk/nyheder/indland/minister-vil-bruge-kunstig-intelligens-og-rollebytte-til-loese-mangel-paa-radiologer>.

Ploug, T., Sundby, A., Moeslund, T. B., Holm, S. (2021): Population preferences for performance and explainability of artificial intelligence in health care: Choice-based conjoint survey. *Journal of Medical Internet Research*, 23(12): e2611.

Raftopoulos, M. & Hamari, J. (2023): Human-AI collaboration in organisations: A literature review on enabling value creation. ECIS 2023 Research Papers. 381. https://aisel.aisnet.org/ecis2023_rp/381.



Read more about the subproject here

[MAGIC - AI for Breast Cancer Diagnostics \(CAI-X\)](#)

[ADD webinar on AI in the health care system](#)

Lantz, P. M. V., Sørensen, U. D., Nielsen, P. B. (2024). [Policy brief: Healthy scepticism and emerging optimism – A human perspective on artificial intelligence in healthcare](#)

Agency for Digital Government (n.d.). [Optimisation of mammography screening through artificial intelligence in the Capital Region of Denmark](#)

3.



*Impact lead:
Ida Schrøder,
Aarhus University*

A qualitative evaluation of algorithmification in a Scandinavian NGO

Subproject: Public administration
and prediction

Context

"We only develop digital solutions to support the children – we have no motive at all to develop just to develop."

The above quotation is from the manager at the Children's helpline in the Danish Children's Rights organisation, and it encapsulates the ambiguity often present in AI development projects. While AI-infused systems are frequently introduced with a clear purpose of either solving a problem or improving existing practices, they often entail far-reaching transformations of work processes, qualities, and the environments embedded. This kind of 'algorithmification' is a general trend in Danish welfare society, where the goal is for future welfare services to be supported by personalized data analysis (Government 2022), and where organisations and professions pursue this goal to the best of their ability.²⁶ In the case of AI for decision support in social services, the main stakeholders are local authorities and private firms that support and advise the public sector and how to develop and use AI solutions in their work.²⁷ A third group of main stakeholders are the non-governmental, civil society organisations (NGOs), who aim to supplement public social services and improve the lives of targeted groups of people.

The study

This case study investigates how voluntary counsellors use a specific AI solution, the Counselling Assistant (AICA), and explores what the algorithmification of the Danish Children's Rights organisation entails for their voluntary counsellors and the organisation at large. Due to the focus on evaluation, but without a clear set of evaluation criteria, fostered rather open research questions to guide the data collection:

- What significance does the AICA have in the work of the volunteers?
- How does the significance of the AICA change over time and in different situations?
- Where, when, and who contributes to making the algorithm professional?

To answer the above-stated research question, the research team applied a situational ethnographic approach, aiming to collect data from as many different situations involving the AI solutions as possible. The research team combined data from volunteers, employers, developers, managers and lawyers, focusing on work, expectations, and attitudes, as well as from documents and decision-making meetings where the

26 Ratner, H. F. & Schröder, I. (2022). The emerging ethical plateau of predictive algorithms in the public administration of Danish child protective services. 1-25. Ratner H. F. & Schröder, I. (2023). Ethical Plateaus in Danish Protection Services: The Rise and Demise of Algorithmic Models. *Science & Technology Studies*, 44-61. <https://doi.org/10.23987/sts.126011>.

Laage-Thomsen, J., & Ratner, H. F. (2024). Kunstig intelligens i den offentlige forvaltning: sammenhænge mellem algoritmisk regulering og automatisering af beslutninger i de danske AI "signaturprojekter". *Politica*. Advance online publication. <https://tidsskrift.dk/politica/article/view/153262/195915>

Høybye-Mortensen, M. (2020). Sagsbehandlers roller i den digitale forvaltning. I J. Hundebøl, A. S. Pors, & L. H. Sørensen (red.), *Digitalisering i offentlig forvaltning* (s. 165-186). Samfundslitteratur.

Meilvang, M. L. (2023). Working the Boundaries of Social Work: Artificial Intelligence and the Profession of Social Work. *Professions and Professionalism*, 13(1). <https://doi.org/10.7577/pp.5108>

27 Laage-Thomsen, J., Ratner, H. F., & Schröder, I. (2025). The beginning of AI-driven welfare? An inquiry into how public sector AI experiments shape the Danish welfare state. I *Digitalization, Data and Welfare: Sociotechnical Approaches to Service Delivery* (s. 38-56). Edward Elgar Publishing. <https://doi.org/10.4337/9781035338153.00010>

AI solutions were described and assessed. The theoretical foundation of the research relies on Science and Technology Studies (STS).

A central point in STS is that technologies are never stable.²⁸ They are rather relational, meaning they are integrated differently and have different significances in different situations.²⁹ 'Algorithm' is, in this case, understood as an unbounded model, as its reach and consequences encompass far more than what is hidden in the computer's code language.³⁰ This theoretical approach is particularly suited to uncovering and noticing what we do not take for granted and therefore do not expect to see.³¹

The Danish Children's Rights organisation – partner in the ADD partnership Alliance

The Danish Children's Rights organisation, as the main stakeholder in this subproject, is a well-known NGO in Denmark that offers free counselling to children and young people up to the age of 25. Their mission is to "stop neglect of children". This refers to neglect caused by children's parents and by society, e.g., local governments, schools, day care institutions, etc. Correspondingly, the Danish Children's Rights organisation both helps children and represents their voices in Danish society, often with a critical stance against national policies and public sector organisations.

To meet their mission, they raise funding from private investors, receive charity, and provide direct help services with volunteer staff. To continuously develop, target and improve their services and means of representation, they collaborate with public/private organisations and

researchers. In the specific case of developing AI solutions, they raised funding from a philanthropic foundation and collaborated with a local tech firm. These stakeholders – children, public/private partners, funders and Danish citizens – all contribute to the Danish Children's Rights organisation for moral reasons, because they want to do good. However, there is always the possibility that a partner or funder aims to exploit the brand of the Danish Children's Rights organisation to boost their own social responsibility. To mitigate this kind of exploitation, the Danish Children's Rights organisation insists on a very high ethical standard, continuously evaluating their collaborations, services, ideas, and impact against the consequences for individual children. This regarded our research as well. And, as we shall see in the following, this duality between following an agenda – for instance, the implementation of a new AI support application – and doing good for the individual was very much present in the development and use of their AI solutions.

Pathways to impact

The subproject's pathway to impact was built through close collaboration with the Danish Children's Rights organisation and continuous engagement with its staff and volunteers throughout the research process.

Co-creation and stakeholder engagement

The collaboration between Ida Schrøder and the Danish Children's Rights organisation started with an informal conversation, which evolved into a formalised collaboration between Ida, as the researcher, and the Danish Children's Rights organisation, as a case to study the development and use of an AI solution to support their volun-

28 Suchmann, L. (2023). The uncontroversial 'thingness' of AI. *Big Data & Society* 4(2). <https://doi.org/10.1177/20539517231206794>

29 Høybye-Mortensen, M. (2021). Science Technology Studies (STS) - de ikke-menneskelige aktører i socialt arbejde. I M. Christensen, R. E. Jørgensen, N. H. Lysen, & C. Rosenberg (red.), *Videnskabsteori og socialt arbejde* (s. 217-234). Samfundslitteratur.

30 Ratner H. F. & Schrøder, I. (2023). Ethical Plateaus in Danish Protection Services: The Rise and Demise of Algorithmic Models. *Science & Technology Studies*. 44-61. <https://doi.org/10.23987/sts.126011>.

Seaver, Nick (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data and Society* 4 (2):1-12. doi:10.1177/2053951717738104.

31 Star, S. L. & Strauss, A. (1999). A. Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work (CSCW)* 8, 9-30. <https://doi.org/10.1023/A:1008651105359>

tary staff. This kind of co-creation in the field and scope of research is a partnership collaboration.

To make the research tangible, Ida invited Mathilde Høybye-Mortensen, professor at VIA University College, and Marie Leth Meilvang, associate professor at University College Lillebælt, to join her research team. This transdisciplinary team structure, with one member based close to the Danish Children's Rights organisation's branch in Aarhus, enabled a more comprehensible and locally embedded research process, enhancing both data access and interdisciplinary analysis.

Furthermore, to sustain engagement from the representatives of the Danish Children's Rights organisation, the research team held two reflection workshops together with the organisation. The aim was to qualify data and validate ethical aspects of the research. At the Danish Children's Rights organisation, they expressed the usefulness of having the opportunity to discuss why they are doing, what they are doing, and that the discussions in the workshops helped them create a language for talking about their AI initiatives, not least their ethical implications.

Ethics, boundaries, and trust-building

Early in the research, ethical concerns around access to sensitive child data were identified as a potential barrier. Rather than compromising the integrity of the study or risking ethical breaches, the research team and the Danish Children's Rights organisation agreed to revise the scope of the project to only include one AI-solution, the AICA, leaving out other, more tentative AI-solutions.

As part of balancing the ethical concerns and the research uptake, the research team and the Danish Children's Rights organisation had regular meetings throughout the research process where they negotiated in a friendly and respectful manner how the research team could maintain their research code of conduct while also offering the Danish Children's Rights organisation the results of value to their work.

From actionable recommendations to evidence of tangible outcomes

The project culminated in a report outlining seven key points of attention for both the Danish Children's Rights organisation and other NGOs working with AI to support human services:

1. Focus on the border between AI support and AI automation.
2. Focus on the double ethical responsibility.
3. Focus on the unique data foundation a societal responsibility.
4. Focus on the role as an AI frontrunner fosters transparency.
5. Focus on the unique transdisciplinary collaboration.
6. Focus on the human-centred artificial intelligence.
7. Focus on (volunteering) data-work.

While grounded in the context of the Danish Children's Rights organisation, these points of attention have broader applicability across nonprofit and public sectors. They offer a practical and ethical framework for those planning, developing, or implementing AI in socially sensitive contexts.

At the time of finalising the description of this ADD impact case in March 2026, the Danish Children's Rights organisation have gone through several iterations of their AI-solutions. The latest the research team has heard from them is that the AICA is now assisting the voluntary counsellors with their administrative task of filing their conversations, while the AICA is no longer supporting the volunteers in their conversations. This mirrors a shift towards data quality and away from complex task of juggling dual ethical responsibilities for model fairness and situated fairness. While the Danish Children's Rights organisation may have considered such changes independently, they have acknowledged that the research helped them discuss how best to develop the AICA. This alignment between the Danish Children's Rights organisation's adjustments and the report's focal points suggests that the research has impacted their strategic direction on AI use.



Broader societal impact

Beyond the Danish Children's Rights organisation, the research has sparked broader public and professional dialogue. Researcher Ida Schrøder has been repeatedly invited to speak at events and institutions, using the case of the Danish Children's Rights organisation to explore the ethical and practical challenges of AI in professional settings. Invitations have included the Municipality of Copenhagen, the Alumni Network for Master Students at Adult Learning and Organisational Change, the National Agency for IT and Learning's yearly conference on online supervision, and the ADD symposium for researchers, managers, and public sector practitioners.

These engagements demonstrate that the research resonates beyond academia, offering relevant insights into the ethical governance of AI in human-facing professions.

Conclusion

This project exemplifies how research built on co-creation, ethical reflection, and sustained stakeholder engagement can lead to tangible organisational changes, inform broader public debate, and contribute to the development of more ethically responsible AI practices.

Moreover, it shows the importance of working with key stakeholders and engaging them in the research process to guide the research toward useful insights and recommendations. To get the full picture of the project's impact on the Danish Children's Rights organisation's implementation of a new language model, it would be interesting to reach out to them to get feedback and evidence of the impact. It would also be interesting to investigate the perspective of the Danish Children's Rights organisation's partners' reactions to the new AI solutions to grasp the reach of the project's impact.

To better grasp the reach of the impact, it would be interesting to get to learn how Ida's talks for other public sector organisations have impacted their work with algorithmic solutions and ethics.

References

Høybye-Mortensen, M. (2021). Science Technology Studies (STS) - de ikke-menneskelige aktører i socialt arbejde. I M. Christensen, R. E. Jørgensen, N. H. Lysen, & C. Rosenberg (red.), *Videnskabsteori og socialt arbejde* (s. 217-234). Samfundslitteratur.

Laage-Thomsen, J., & Ratner, H. F. (2024). Kunstig intelligens i den offentlige forvaltning: sammenhænge mellem algoritmisk regulering og automatisering af beslutninger i de danske AI "signaturprojekter". *Politica*. Advance online publication. <https://tidsskrift.dk/politica/article/view/153262/195915>

Laage-Thomsen, J., Ratner, H. F., & Schrøder, I. (2025). The beginning of AI-driven welfare? An inquiry into how public sector AI experiments shape the Danish welfare state. *Digitalization, Data and Welfare: Sociotechnical Approaches to Service Delivery* (s. 38-56). Edward Elgar Publishing. <https://doi.org/10.4337/9781035338153.00010>

Meilvang, M. L. (2023). Working the Boundaries of Social Work: Artificial Intelligence and the Profession of Social Work. *Professions and Professionalism*, 13(1). <https://doi.org/10.7577/pp.5108>

Ratner, H. F. & Schrøder, I. (2022). The emerging ethical plateau of predictive algorithms in the public administration of Danish child protective services. 1-25. <https://events.ruc.dk/democracyanddigitalcitizenship/conference>

Ratner H. F. & Schrøder, I. (2023). Ethical Plateaus in Danish Protection Services: The Rise and Demise of Algorithmic Models. *Science & Technology Studies*. 44-61. <https://doi.org/10.23987/sts.126011>.

Seaver, Nick (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data and Society* 4 (2):1-12. doi:10.1177/2053951717738104.

Star, S. L. & Strauss, A. (1999). A. Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work (CSCW)* 8, 9-30. <https://doi.org/10.1023/A:1008651105359>

Suchmann, L. (2023). The uncontroversial 'thingness' of AI. *Big Data & Society* 4(2). <https://doi.org/10.1177/20539517231206794>.



Read more about the subproject here

The Danish Children's Rights organisation (2025): [Data Cleaning, Data Quality, and Data-Driven Counseling: A Study of the Development and Application of Artificial Intelligence to Support Volunteer Counseling Sessions \(Danish version\)](#)

Schrøder, I., Leth Meilvang, M. & Høybye-Mortensen, M. (2026): The (Un)Fair Algorithm: Socio-Technical Ethics Work for Artificial Intelligence in Social Work. *Ethics and Social Welfare*, 1-16. <https://doi.org/10.1080/17496535.2026.2619945>



*Impact lead:
Johann Ole Willers,
Copenhagen
Business School*

4.

Beyond control? Exposing and shaping the global surveillance industry

Subproject: Finance and transparency

Context

This impact case focuses on research on the global private surveillance industry and its role in enabling lawful interception, intrusion, and surveillance capabilities. The work sits at the intersection of International Political Economy, Security Studies, Surveillance Studies, Science & Technology Studies, and global governance research. It contributes to ongoing debates on cyber proliferation, economic security, export controls, human rights, surveillance capitalism, and state-market relations in high-tech sectors.

The research examines the rapid growth of private vendors supplying governments around the world with interception, intrusion, and spyware technologies.

These technologies:

- Strengthen intelligence and military capabilities.
- Blur the boundary between lawful interception and offensive cyber operations.
- Enable domestic repression of activists, journalists, and political opposition.
- Raise concerns about strategic proliferation of offensive cyber capabilities to US/NATO adversaries.

Key stakeholders include national security agencies, law enforcement, intelligence services, EU institutions and export-control bodies, US and

NATO policymakers, civil society organisations, and technology companies affected by spyware abuse.

The study

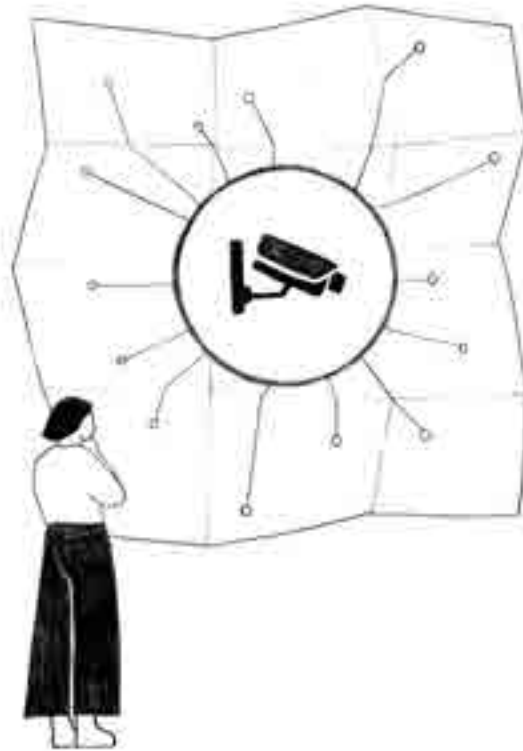
The study investigates the opaque and evolving market of private interception, intrusion, and surveillance (PIIS) tools, which has become central to the expansion of global digital surveillance.³² It asks how this market emerged, why regulation has struggled to keep up, how state and market actors interact, and what this reveals about governance in high tech sectors. The findings reveal a globalized industry, the cross border spread of offensive cyber tools, and persistent regulatory gaps driven by the mismatch between static policy tools and fast changing market conditions.

The analysis draws on a new dataset of 5,973 industry presentations across 64 trade fairs (2003-2020) alongside corporate ownership data, industry reports, leaked documents, and 19 in-depth interviews with insiders and policymakers. Moreover, the Atlantic Council report cross-references 107,542 arms-fair exhibitors and 777 ISSWorld sponsors/speakers, identifying 224 overlapping firms active in both domains.

The study conceptualizes the PIIS market as a dynamic assemblage shaped by evolving state-market interactions and technological change. Drawing on political economy, security studies, and surveillance research, it integrates

32 Deibert, R. J. (2012). Black code. Surveillance, privacy, and the dark side of the internet. *McClelland & Stewart*.

Deibert, R. J., & Pauly, L. W. (2019). Mutual entanglement and complex sovereignty in cyberspace. In E. Ruppert, E. Isin, & D. Bigo (Eds.). *Data politics*. Taylor & Francis
Penney, J., McKune, S., Gill, L. & Deibert, R. J. (2018). Advancing human rights-by-design in the dual-use technology industry. *Columbia Journal of International Affairs*, 71(2), 104.



insights from global security assemblage theory, technological governance literature, and international trade scholarship.

It identifies three core dynamics shaping the market:

1. **Territorialization** – shifts in the geographical and organisational scope of the market.
2. **Technological affordances** – evolving infrastructures and tools that expand surveillance capabilities.
3. **Ordering** – changing relationships between vendors, states, intelligence agencies, and military actors.

Together, these dynamics continuously reconfigure market hierarchies, making traditional regulation difficult to sustain.

Collaboration

The case builds on a long-standing research collaboration between Johann Ole Willers from Copenhagen Business School and Lars Gjesvik from the Norwegian Institute of International Affairs, with contributions from Winnona DeSom-

bre from Harvard Belfer Center, and an institutional partnership with the Atlantic Council's Cyber Statecraft Initiative led by Trey Herr. This partnership developed through ongoing dialogue around earlier Atlantic Council work.

The collaboration thus combined academic analysis with a policy focused platform, enabling faster translation of findings into debates on export controls, cyber proliferation, and democratic governance.

Pathways to impact

Impact pathways were enabled by the combination of original academic research that provides novel empirical insights on a sensitive area of digital politics, and a collaboration with a policy-focused platform that facilitated the targeted dissemination of research insights into key stakeholder groups, including politicians, NGOs, and diplomatic communities.

Societal impact

The main academic output is the article *Beyond control? The political economy of private inter-*

ception, intrusion, and surveillance markets by Lars Gjesvik and Johann Ole Willers,³³ published in the highly ranked journal *Review of International Political Economy*.³⁴

The team also co-authored the Atlantic Council Issue Brief *Surveillance Technology at the Fair: Proliferation of Cyber Capabilities in International Arms Markets*³⁵. This report generated rapid attention with more than 1,000 downloads within 24 hours. It was featured in MIT Technology Review "A grim outlook": *How cyber surveillance is booming on a global scale*,³⁶ including an interview with Johann Ole Willers. The research was also featured in a range of specialized forums and newsletters, including the Risky Biz podcast, which reaches around 20,000 listeners per episode. One co-author was interviewed on Swedish national public television broadcaster SVT and Swedish radio.³⁷

The Issue Brief was subsequently cited by the European Parliamentary Research Service in a study on *Countering Spyware Abuse*,³⁸ and the EU's cyber diplomacy initiative Cyber Direct.³⁹

These outcomes highlight the relevance of the findings and their reach into technology and cybersecurity communities as well as broader public debates.

The visibility of the research also supported collaboration with journalists and civil society actors working to hold the surveillance industry accountable. This includes contributions to the

Open Society Foundation workshop on *Financial Enablers of the Targeted Surveillance Industry* on May 25-26, 2023, in Berlin. The workshop brought together journalists, human rights advocates, lawyers, and academics from around the world to share methods and explore how civil society can better address the industry.

Briefings to stakeholders

Targeted dissemination of the Atlantic Council Issue Brief led to engagement with key stakeholders and a series of briefings. These included meetings with Members of the European Parliament, the Norwegian Ministry of Foreign Affairs (export control division), the Norwegian Ministry of Justice, Microsoft, and several NGOs including the Cyber Peace Institute.

In addition, one researcher testified before the U.S.-China Economic and Security Review Commission in February 2022 (USCC, 2022) and the US National Security Council.

Conclusion

The research provides a systematic account of a largely opaque global market, identifies problematic actors, and informs national and international policy debates. It has contributed to high-level security briefings and offers both analytical tools and practical guidance for addressing the governance challenges posed by commercial spyware and surveillance vendors.

-
- 33 Gjesvik, L., & Willers, J. O. (2024). Beyond control? The political economy of private interception, intrusion, and surveillance markets. *Review of International Political Economy*, 31(6), 1840–1864. <https://doi.org/10.1080/09692290.2024.2375499>
- 34 Gjesvik, L., & Willers, J. O. (2024). Beyond control? The political economy of private interception, intrusion, and surveillance markets. *Review of International Political Economy*, 31(6), 1840–1864. <https://doi.org/10.1080/09692290.2024.2375499>
- 35 DeSombre, W., Gjesvik, L., & Willers, J. O. (2021). *Surveillance technology at the fair: Proliferation of cyber capabilities in international arms markets*. Atlantic Council, Cyber Statecraft Initiative. <https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/surveillance-technology-at-the-fair/>
- 36 O'Neill, Patrick Howell (2021, November 8). "A grim outlook": How cyber surveillance is booming on a global scale. *MIT Technology Review*. <https://www.technologyreview.com/2021/11/08/1039395/grim-outlook-cyber-boom-atlantic-council-report/>
- 37 SVT November 21 2021 Svenskt övervakningsbolag kritiseras i amerikansk rapport <https://www.svt.se/nyheter/inrikes/svenskt-overvakningsbolag-kritiseras-i-amerikansk-rapport>
- 38 Mildebrath, H. (2022) Europe's PegasusGate – Countering spyware abuse. *EPRS – European Parliamentary Research Service*. [https://www.eurparl.europa.eu/RegData/etudes/STUD/2022/729397/EPRS_STU\(2022\)729397_EN.pdf](https://www.eurparl.europa.eu/RegData/etudes/STUD/2022/729397/EPRS_STU(2022)729397_EN.pdf)
- 39 Broeders, D. & Kavanagh, C. (2023). Shades of Grey: Cyber Intelligence and (Inter)national Security. Policy Brief. EU Cyber Direct. <https://eucyberdirect.eu/research/shades-of-grey-cyber-intelligence-and-inter-national-security>

More broadly, the case shows how sustained academic research can inform current policy discussions. By combining new data, clear concepts, and academic partnerships, the project translated research insights into actionable recommendations.

References

Broeders, D. and Kavanagh, C. (2023). Shades of Grey: Cyber Intelligence and (Inter)national Security. Policy Brief. EU Cyber Direct. <https://eucyberdirect.eu/research/shades-of-grey-cyber-intelligence-and-inter-national-security>

Deibert, R. J. (2013). *Black code. Surveillance, privacy, and the dark side of the internet*. McClelland & Stewart.

Deibert, R. J., & Pauly, L. W. (2019). Mutual entanglement and complex sovereignty in cyberspace. In E. Ruppert, E. Isin, & D. Bigo (Eds.), *Data politics*. Taylor & Francis.

DeSombre, W., Gjesvik, L., & Willers, J. O. (2021). *Surveillance technology at the fair: Proliferation of cyber capabilities in international arms markets*. Atlantic Council, Cyber Statecraft Initiative. <https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/surveillance-technology-at-the-fair/>

Gjesvik, L., & Willers, J. O. (2024). Beyond control? The political economy of private interception, intrusion, and surveillance markets. *Review of International Political Economy*, 31(6), 1840–1864. <https://doi.org/10.1080/09692290.2024.2375499>

Mildebrath, H. (2022) Europe's PegasusGate – Countering spyware abuse. *EPRS – European Parliamentary Research Service*. [https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729397/EPRS_STU\(2022\)729397_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729397/EPRS_STU(2022)729397_EN.pdf)

O'Neill, Patrick Howell (2021, November 8). “A grim outlook”: How cyber surveillance is booming on a global scale. *MIT Technology Review*. <https://www.technologyreview.com/2021/11/08/1039395/grim-outlook-cyber-boom-atlantic-council-report/>

Penney, J., McKune, S., Gill, L., & Deibert, R. J. (2018). Advancing human rights-by-design in the dual-use technology industry. *Columbia Journal of International Affairs*, 71(2), 104.

Risky Business Media. (2021, November 11). *Risky Bulletin Newsletter*. <https://risky.biz/srsly-risky-biz-thursday-november-3a2/>

USCC (2022). The U.S.-China Economic and Security Review Commission www.uscc.gov/sites/default/files/2022-02/Winnona_DeSombre_Bio.pdf



Read more about the subproject here

ADD Blogpost (2025): [Organizations are saying no to AI – Ole Willers dives into one specific barrier](#)

ADD exit-article (2026): [Ole Willers: Cybersecurity is no longer only a technical problem – it is a systemic one](#)



Impact leads:
*Theresia Veronika Rampisela,
Maria Maistro and
Christina Lioma,
University of Copenhagen*



5.

Fairness in recommender systems

Subproject: Machine learning and bias

Context

Recommender systems are applications that match items such as services or products, to individuals. The goal is for users to interact (click, view, purchase, etc.) with the items recommended by the systems. They are not only part of common everyday applications, such as e-commerce, entertainment websites, and social media, but also high-risk applications, including job recommendations and health. For example, job seekers on a job application platform may miss out on employment opportunities because they are frequently placed at the bottom of the rankings shown to recruiters. It is thus fundamental to ensure that these systems provide fair treatment, avoiding discrimination against specific groups or individuals. Fairness of AI systems, such as recommender systems, has become increasingly important and even compulsory for high-risk applications, according to recent legislation, e.g., the EU AI Act.

To accurately quantify the fairness of these systems, we need robust evaluation measures that are applicable in a wide range of cases and easy to interpret. This subproject at the Department of Computer Science at the University of Copenhagen (DIKU) focuses on analysing and developing evaluation measures to assess the level of fairness of recommender systems. This is necessary for practitioners, both from research and industry, to detect cases of unfair treatment and develop mitigation strategies. Specifically, this subproject involved researchers from Department of Computer Science (DIKU) and the Royal Melbourne Institute of Technology in Australia, and used data from well-known providers of recommendation systems, such as Amazon and Yelp.

To investigate fairness in recommender systems, the project addressed the following questions:

1. What are the strengths and limitations of the current evaluation measures of fairness in recommender systems?
2. How do we overcome the limitations of the current evaluation measures of fairness?
3. How does increasing fairness affect effectiveness (e.g. accuracy) in recommender systems?
4. Which evaluation measures should we use to evaluate fairness in recommender systems?

The study

To evaluate fairness in recommender systems, the research team has used real-world datasets provided by tech companies such as Amazon, including ratings given by users to products or content, or click logs. These kinds of data were used as input to AI models, which generated a ranked list of personalised recommendations based on the user interaction history. Then, the research team: a) evaluated the theoretical and empirical properties of existing fairness measures, for example, by analysing their mathematical formulations, computing the agreement among measures, and performing stress tests to assess their robustness; and b) created new improved evaluation measures of fairness. This work received the prestigious best journal paper award from Women in RecSys and was presented at the 18th ACM Conference on Recommender Systems (RecSys 2024).



Some of the key findings and contributions to the field of recommender systems are as follows:

- Several existing evaluation measures of fairness in recommender systems are unreliable because they are mathematically flawed and often conflict with each other in what they deem fair. This is problematic because, for instance, the measures cannot even be computed for some data, or they can return misleading results, e.g., a system being assessed as fair when in practice, it is not.
- The research team corrected the above flaws in existing evaluation measures. This resulted in making the evaluation measures more reliable, more usable, and more interpretable to humans. The research team's reformulated evaluation measures of fairness can now be

used across a wider range of evaluation scenarios than the original evaluation measures.

- Increasing fairness sometimes comes at the cost of reduced effectiveness, e.g. less precise recommendations. This is a real-life problem, because practitioners need technology that is both accurate and fair. To address this problem, the research team created a tool that quantifies the trade-off between recommendation fairness and effectiveness, so that fairness and effectiveness can be jointly measured. This allows practitioners to calibrate their recommendation technologies without compromising accuracy or fairness, while also ensuring that users of this technology receive recommendations that balance accuracy and fairness.⁴⁰

Collaboration with the Royal Melbourne Institute of Technology

As part of her work on the project of fairness in recommender systems, ADD Ph.D. student Theresia Veronika Rampisela visited the Royal Melbourne Institute of Technology in Australia and collaborated with Professor Falk Scholer at the ARC Centre of Excellence for Automated Decision-Making and Society. Prof. Scholer provided expertise on evaluation measures, statistical testing, and human-computer interaction. This collaboration resulted in a new approach to evaluate recommendation fairness for individuals but also for groups of users, an area that had not received much attention until then, making this contribution particularly impactful.

Pathways to impact

The research has generated impact mainly through “participation by product” in the form of scientific outputs, such as journal articles and conference papers at top international venues in the recommender systems field. In addition, the research has been communicated to the wider audience through contributions to the printed media and social media, such as opinion pieces, podcasts, and presentations, both in Denmark and internationally.

Collaboration with internal and external peers

The researchers have been very active in in-person participation in the world’s elite scientific events, where the state of the art research by leading industry and academia is presented every

year. The work has been disseminated through presentations at top international conferences, as well as accepted invitations for talks and seminars with internal and external peers, for instance at universities in Denmark, Sweden, and the UK.⁴¹

The research team has released to the public all programming code of the tools and evaluation measures developed in this project. This allows other scientists and technologists, not only to verify the robustness of these tools, but also to reuse them and adapt them to their needs. All the knowledge produced in this project has therefore been communicated back to the broad public, not only in scientific publications, but also as ready to use tools, free for anyone, accompanied by guidelines for practitioners.

Guidelines for selecting fairness evaluation measures

There exist many fairness evaluation measures for recommender systems, making it difficult for people to choose the correct evaluation measure for a specific purpose. Therefore, one outcome of this project was to develop guidelines to help researchers and practitioners to determine which evaluation measures of fairness to use and how to use them.⁴² For instance, these guidelines specify:

1. Which evaluation measures should be used when assessing the fairness of an individual recommender system, and which should be used when comparing the fairness of several recommender systems to each other. Ignoring this guideline can lead to gross misestimation of fairness by practitioners.⁴³

41 Rampisela, T. V. (2024, July 14–18). Can we trust recommender system fairness evaluation? The role of fairness and relevance. ACM SIGIR 2024, Washington D.C., USA. Maistro, M. (2024, May 27). Evaluation beyond relevance: accounting for credible, correct and fair information. Talk at the Search Engines course, A.Y. 2023/2024, University of Padua, Padua, Italy.

42 Rampisela, T. V., Maistro, M., Ruotsalo, T., Scholer, F., & Lioma, C. (2025). Relevance-aware Individual Item Fairness Measures for Recommender Systems: Limitations and Usage Guidelines. ACM Transactions on Recommender Systems. <https://doi.org/10.1145/3765624>
Rampisela, T. V., Ruotsalo, T., Maistro, M., & Lioma, C. (2024). Can We Trust Recommender System Fairness Evaluation? The Role of Fairness and Relevance. I SIGIR 2024 - Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (s. 271-281). Association for Computing Machinery, Inc. <https://doi.org/10.1145/3626772.3657832>

43 Rampisela, T. V., Maistro, M., Ruotsalo, T., Scholer, F., & Lioma, C. (2025). Relevance-aware Individual Item Fairness Measures for Recommender Systems: Limitations and Usage Guidelines. ACM Transactions on Recommender Systems. <https://doi.org/10.1145/3765624>
Rampisela, T. V., Ruotsalo, T., Maistro, M., & Lioma, C. (2024). Can We Trust Recommender System Fairness Evaluation? The Role of Fairness and Relevance. I SIGIR 2024 - Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (s. 271-281). Association for Computing Machinery, Inc. <https://doi.org/10.1145/3626772.3657832>

2. How humans should interpret the scores of fairness evaluation measures. Different evaluation measures can be more or less generous, i.e. some of them can give mostly high scores, while other measures can give mostly low scores. This should be taken into account when a human tries to understand what a score of 90% fairness means for a system, for example. A score like this does not mean that the recommendations of that system are almost perfect, if the evaluation measure that produced this score tends to generally give scores above 85% for most recommendations.⁴⁴
3. When interested in assessing both the fairness and the accuracy of a recommendation system, the tool developed by the research team, Distance to Pareto Frontier (DPFR), should be used, because it considers the fairness-effectiveness trade-off, while most existing evaluation measures tend to favour either relevance or fairness.⁴⁵
4. Both group and individual fairness should be evaluated, because they do not always agree. A recommender system can be fair to a group, but at the same time highly unfair to individuals within that group. For example, a recommender system might be fair across different genders. But when we consider a specific group, for example, women, the model may still discriminate against women based on ethnicity.⁴⁶

By providing guidelines to stakeholders to apply in selecting fairness evaluation measures, the research team has laid the foundation for future impact for a diverse set of key stakeholders as the guidelines for fairness in recommender systems can be applied by both consumers and providers of recommender systems as well

as researchers within the field of study, system developers, companies or institutions that utilise these systems, or regulators, ethical experts and legislators who must understand this technology. These stakeholders can all benefit from the research by following the best practices in fairness evaluation and using the tools produced in this project. For instance, policymakers can use the results in this project to set measurable fairness criteria for high-risk applications, based on easily interpretable fairness measures. In this way, the project's results can benefit a diverse set of key stakeholders, all of whom engage with or work with recommender systems in one way or another.

Communication of findings and recommendations in different fora

An important part of the research team's work for communicating findings and recommendations has been through presentations at conferences and guest lectures around the world, from Australia to Europe. Asia and the United States, underlining the importance and relevance of the project in the field of recommender systems.

In addition to the impact achieved by presenting the work to top research conferences and publishing journal articles, the research team has also communicated broadly in the printed and social media. For instance, Theresia Rampisela collaborated with a podcast series and also published an online article on the ADD-project's website discussing fairness in recommender systems, where she defines what fairness in recommender systems means: "What do I mean by fairness? (...) Unfortunately, there is no single definition for this, but it is about ensuring users are treated without discrimination".⁴⁷ Results of this project have also been communicated in

44 Rampisela, T. V., Maistro, M., Ruotsalo, T., Scholer, F., & Lioma, C. (2025). Relevance-aware Individual Item Fairness Measures for Recommender Systems: Limitations and Usage Guidelines. *ACM Transactions on Recommender Systems*. <https://doi.org/10.1145/3765624>

Rampisela, T. V., Ruotsalo, T., Maistro, M., & Lioma, C. (2024). Can We Trust Recommender System Fairness Evaluation? The Role of Fairness and Relevance. *I SIGIR 2024 - Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (s. 271-281). Association for Computing Machinery, Inc. <https://doi.org/10.1145/3626772.3657832>

45 Rampisela, T. V., Ruotsalo, T., Maistro, M., & Lioma, C. (2025). Joint Evaluation of Fairness and Relevance in Recommender Systems with Pareto Frontier. *WWW '25: Proceedings of the ACM on Web Conference 2025* (s. 1548-1566). Association for Computing Machinery, Inc. <https://doi.org/10.1145/3696410.3714589>

46 Rampisela, T. V., Maistro, M., Ruotsalo, T., Scholer, F., & Lioma, C. (2025). Stairway to Fairness: Connecting Group and Individual Fairness. *I RecSys '25: Proceedings of the Nineteenth ACM Conference on Recommender Systems* (s. 677-683). Association for Computing Machinery, Inc. <https://doi.org/10.1145/3705328.3748031>

a keynote talk by Christina Lioma at Infosecurity Expo in Copenhagen, as well as opinion pieces and interviews at Ingeniøren Version 2, videnskab.dk, tjeekdet.dk, and Kristeligt Dagblad, among others.

Conclusion

This project makes contributions to the area of measuring the fairness of automated recommendations. Being able to measure the fairness of automated recommendations is vital to ensuring that technology serves all segments of society equitably. As these systems increasingly influence decisions in areas like hiring, lending, healthcare, and criminal justice, establishing clear fairness metrics is essential not only for ethical accountability, but also to inform and support emerging legislation that protects against systemic bias and discrimination. This work therefore contributes to shaping fairer digital ecosystems and guiding policy that safeguards civil rights in an algorithm-driven world.

Before this project, it was simply not possible to reliably measure how fair an automated recommendation is. A few evaluation measures of recommendation fairness existed, but no one knew how well they worked or how to interpret their scores. This project uncovered critical, previously unrecognized limitations that call into question the validity of widely used evaluation approaches and addressed them by introducing corrections and building new robust fairness evaluation tools that have set a new standard for the field. These tools are state of the art, and they are freely available to all at <https://github.com/orgs/diku-irlab/repositories?q=fairness>.

References

Rampisela, T. V., Maistro, M., Ruotsalo, T., & Lioma, C. (In press). Measuring Individual User Fairness with User Similarity and Effectiveness Disparity. 1. 48th European Conference on Information Retrieval (ECIR 2026), Delft, Netherlands.

Rampisela, T. V., Maistro, M., Ruotsalo, T., Scholer, F., & Lioma, C. (2025). Stairway to Fairness: Connecting Group and Individual Fairness. I RecSys '25: Proceedings of the Nineteenth

ACM Conference on Recommender Systems (s. 677-683). Association for Computing Machinery, Inc. <https://doi.org/10.1145/3705328.3748031>

Rampisela, T. V., Maistro, M., Ruotsalo, T., Scholer, F., & Lioma, C. (2025). Relevance-aware Individual Item Fairness Measures for Recommender Systems: Limitations and Usage Guidelines. ACM Transactions on Recommender Systems. <https://doi.org/10.1145/3765624>

Rampisela, T. V., Ruotsalo, T., Maistro, M., & Lioma, C. (2025). Joint Evaluation of Fairness and Relevance in Recommender Systems with Pareto Frontier. WWW '25: Proceedings of the ACM on Web Conference 2025 (s. 1548-1566). Association for Computing Machinery, Inc. <https://doi.org/10.1145/3696410.3714589>

Rampisela, T. V., Ruotsalo, T., Maistro, M., & Lioma, C. (2024). Can We Trust Recommender System Fairness Evaluation? The Role of Fairness and Relevance. I SIGIR 2024 - Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (s. 271-281). Association for Computing Machinery, Inc. <https://doi.org/10.1145/3626772.3657832>

Rampisela, T. V., Maistro, M., Ruotsalo, T., & Lioma, C. (2024). Evaluation Measures of Individual Item Fairness for Recommender Systems: A Critical Study. *ACM Transactions on Recommender Systems*, 3(2), Artikel 18. <https://doi.org/10.1145/3631943>

Rampisela, T. V., & Marjanovic, S. V. (2024). Can you trust algorithms to be fair? New research says 'it depends'. <https://algorithms.dk/2024/11/18/can-you-trust-algorithms-to-be-fair-new-research-says-it-depends/>

Presentations

Rampisela, T. V. (2026, Mar 2). Connecting Group Fairness and Individual Fairness in Recommender Systems. Online talk at University of Glasgow, Scotland, United Kingdom <https://samoa.dcs.gla.ac.uk/events/viewtalk.jsp?id=20555>

Maistro, M. (2024, May 27). Evaluation beyond relevance: accounting for credible, correct and fair information. Talk at the Search Engines course, A.Y. 2023/2024, University of Padua, Padua, Italy.

Rampisela, T. V. (2024, July 14–18). Can we trust recommender system fairness evaluation? The role of fairness and relevance. ACM SIGIR 2024, Washington D.C., USA.



Read more about the subproject

Septiandri, A. A. (Host). (2024, October 27). [S04E03 Keadilan dalam Sistem Rekomendasi \(Fairness in Recommender Systems\) \[Audio podcast\]](#).

Rampisela, T. V., Maistro, M., Lioma, C. (2025, Oct 1). Ansvarlig AI kræver pålidelige metrikker: Tre mangler går igen. Ingeniøren <https://pro.ing.dk/datatech/holdning/ansvarlig-ai-kræver-paalidelige-metrikker-tre-mangler-gaar-igen> [Magazine article]

ADD Blogpost (2026). [The fault in AI fairness evaluation metrics – worrying flaws and steps forward](#)



6.

Impact lead:

*Carl Stefan Roth-Kirkegaard,
University of
Southern Denmark*

How to lead AI adoption as a manager

Subproject: Innovation and democracy

Context

This subproject of the ADD project on innovation focus on understanding the status and implementation pathways of algorithms and AI in the context of leading Danish private corporations, and in turn, the future consequences this implementation holds for democratic power relations both in society and at work.

As AI and algorithmic systems increasingly shape private sector workflows and decision-making processes, Danish executives report significant transformations in organisational structures, employee roles, and customer relationships. Currently, organisations are finding it difficult to scale and broaden the use of customised AI products beyond specialised applications, and at the same time, the implementation of language modes has raised efficiency and freed up resources to engage in innovation, new tasks, or new servicing of customers. While this is seen by executives mostly as a positive development, both financially, for customers and for employees, they are also concerned about the long-term impact these technologies can have on power relations and socioeconomic relations in the broader society, where the fear is socio-economic instability due to unemployment and in their organisations that will lose human dispositions and values.

To mitigate the potential disruptive effects of algorithms and AI on power relations in the workplace, an important dimension the research team has explored is the way adoption of algorithms and AI is managed and led. The research team's primary focus has been on the ways algorithms and AI might come to automate or augment work practices, and the technological and structural ways to ensure that adoption is benevolent. Executives in Danish companies are beginning to discuss the need for managers to have a more pronounced role in leading AI adoption. However, it has not yet been discussed broadly how managers can lead such changing circumstances. Furthermore, many managers lack the tools, confidence, and strategic guidance to oversee the responsible integration of AI. The research team has, through dialogues with companies since the advent of GenAI, experienced that the topic of algorithms and AI is less prevalent for competencies for managers, despite it being prevalent in the discourse surrounding labour market competencies presently. Managers have been reluctant to engage with the managerial role of regulating and encouraging specific AI uses and therefore risking uses that are not converging with the Danish democratic values, which the unmitigated adoption of AI might problematize.

Collaboration with the Danish union Djøf

The research team is not the only one interested in unfolding the topic of AI adoption and managers, but it is also a prevalent topic for Djøf, a Danish union representing professionals in law, economics, public administration and business administration.

The union is divided between a regular union and a part of the union concerned with managers who are not owners and therefore have labour interests. While most labour unions are concerned with labour negotiations and general upskilling of their members, Djøf is especially concerned with upskilling because its members are engaged in knowledge-intensive work practices and are often organisationally close to the ownership structure of organisations, which makes aggressive labour negotiations less attractive.

This collaboration emerged from a research-partner matching session at the ADD conference in November 2024. Both Djøf and the research team identified a shared concern and interest: managers are underprepared to lead AI implementation efforts effectively. This mutual interest laid the groundwork for a partnership model of collaborative co-creation with Djøf providing stakeholder insights and a target group for the research team.

Djøf was interested in providing resources to their managerial members on the ways these members could improve their engagement with AI adoption and their readiness towards increased future adoption in their organisations.

In this case study, the research team, and in collaboration with Djøf, explores the guidelines between managerial work and algorithms and the broader relationship between management and leadership in organisations. To explore the relationship between managerial work and algorithms, the research team has applied sampling-based, network-based, and snowballing-based interview studies, mainly with corporate leaders and other relevant parties.

Pathways to impact

The research has mainly generated impact through stakeholder collaboration with the outcomes being distributed to practicing managers who lead AI adoption.

Collaboration with Djøf

Throughout the collaboration, Djøf and the research team had continuous dialogues where the research team's existing findings were validated. One of the validated findings is that managers lack structured approaches to lead AI adoptions. Furthermore, the dialogues also reaffirmed the researchers' assertion that middle managers are increasingly faced with managing AI adoption but do not feel they have the tools to manage such a process well. This validation was used to refine and ground the project's output in real-world needs. As stated above, Djøf expressed their interest in a partnership with the researchers because they were interested in providing resources to their managerial members on the ways these members could improve their engagement with AI adoption and their readiness towards increased future adoption in their organisations.

By collaborating and partnering up with Djøf, the researchers secured embedded engagement. Furthermore, the collaboration featured knowledge exchange between Djøf and the research team.

Practical recommendations for managers

The specific outcome of the continuous dialogue with Djøf resulted in a set of practical recommendations for managers responsible for managing AI adoption, with a focus on how they should engage with AI adoption and lead AI adoption among their employees. Furthermore, the set of recommendations is directed towards a broad set of managers, as they should be relevant for both managers who have just started managing AI adoption and a reminder for managers with more experience with vital organisational aspects of an efficient and benevolent adoption. Below are the recommendations divided into two sections: General Recommendations for Leaders on Data and Recommendations for Leaders on Generative AI.

General Recommendations for Leaders on Data:

Be curious and proactive about the AI solutions and data available in your organisation.

Be mindful of data sharing with providers when using data solutions.

Discuss with your employees about your and the organisation's use of data, especially data about them.

Create a data policy regarding ethics and responsibility.

Recommendations for Leaders on Generative AI:

Be aware that AI might be used more by your employees than you expect.

Investigate your employees' AI competencies, as they may be unevenly distributed and not necessarily aligned with general technological understanding.

Experiment with the use of AI in a selection of your work processes where AI might be relevant.

Evaluate the quality of AI use in these experiments.

Develop a policy with your employees on what AI should be used for, including best practices, and particularly what AI should not be used for.

Continuously gather employee input on opportunities and experiences in using AI.

Enhance your employees' understanding of AI.

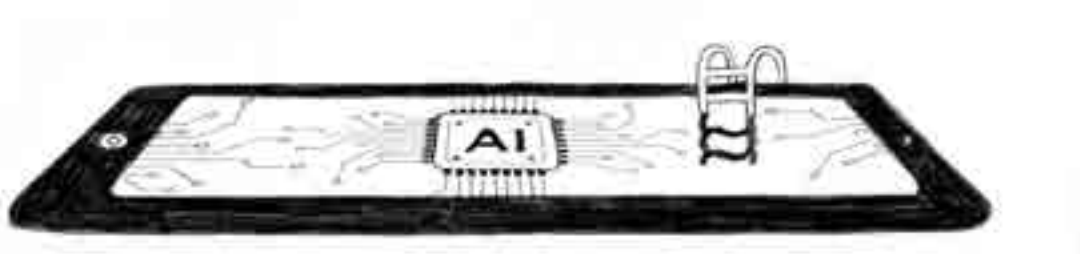
Reflect on where AI should and should not be used in your own leadership tasks.

The research team has not received concrete evidence for the implementation of the recommendations as Djøf has not yet found a media outlet for them. That said, some of the suggestions in the above stated recommendations have been included in a campaign "AI for leaders" on Djøf's website.⁴⁷ This shows that the project has had some impact in the collaborator's development of webinars and courses for managers working with AI.

Future impact

Initially, the research team is attempting to publish the above recommendations in a business-directed mainstream media in order to get the recommendations out to the public. Although managers as the target group might be difficult to reach by using membership-communicated communication, the goal is still to produce a text with the set of recommendations in a more specified and directed format towards managers.

47 <https://www.djoef.dk/kampagner/ai-for-ledere?cardid=ebc56c6c-fea8-4bfd-81b0-9af91fae2555>



The actual impact on managers working with AI adoption might be difficult to evaluate, but the research team will endeavour to collect feedback in collaboration with Djøf from the members to see if the recommendations are improving the resources available to managers in their engagement with algorithms and AI adoption. By collecting feedback together with Djøf laid the foundation for the research team to fine-tune and edit the recommendations and thereby further impact the way managers work with AI adoption.

Conclusion

The actual impact of the recommendations can be hard to evaluate, as the audience of the recommendations is external in relation to the research team. That said, the recommendations lay the foundation of instrumental actions to improve how managers in general and maybe especially middle managers work with AI adoption.

The research project has the potential to enhance the capacity for using the insights with

the recommendations as a stepping stone for workshops for managers leading AI adoption, as they are the external key knowledge users.

This case demonstrates the impact that partnerships between researchers and stakeholders can have in shaping actionable responses to societal challenges, such as leading AI adoption in organisations. Key outcomes of the case:

- A growing awareness among managers of their critical role in AI governance.
- Strengthened organisational capacity within Djøf to support its members' AI leadership.
- An evolving framework that allows managers to experiment with AI responsibility, while aligning practices with democratic values.
- The importance of conceptual clarity: By promoting a shared vocabulary and set of expectations around managerial responsibilities in AI adoption, it can reduce ambiguity and resistance among middle managers.

References

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics*, 40(S1), S293-S340.

Brynjolfsson, E. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies* (Vol. 336). WW Norton & Company.

Fleming, P. (2019). Robots and organization studies: Why robots might not want to steal your job. *Organization Studies*, 40(1), 23-38.

Madsen, D. Ø., & Slåtten, K. (2018). HR-analyse som ledelseskonspekt og ledelsesmote. *Samfundsliderskab i Skandinavien*, 33(1), 42-68.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of management review*, 46(1), 192-210.

Roth-Kirkegaard & Rehn (2025, forthcoming): On Pulsatile Dynamics, Arrhythmic Revolutions, and AI as an Event. In Rombach & Svärd (eds.), *Debating the Speed of Organizational Change*. Palgrave.

Roth-Kirkegaard & Rehn (2024, conference): Ideological imaginaries of identity, innovation and insecurity. On the complex politics of AI adoption. Organisation studies workshop: organization, organizing and politics. Mykonos.

Roth-Kirkegaard, Rehn & Fetteh (2023, conference): Three narratives of Artificial Intelligence adoption: Sense-making about Artificial Intelligence adoption consequences among executives in the Nordics. EGOS. Cagliari.

Smith, A. M., & Green, M. (2018). Artificial intelligence and the role of leadership. *Journal of Leadership Studies*, 12(3), 85-87.

Quaquebeke, N. V., & Gerpott, F. H. (2023). The now, new, and next of digital leadership: How Artificial Intelligence (AI) will take over and change leadership as we know it. *Journal of Leadership & Organizational Studies*, 30(3), 265-275.



Read more about the subproject here

AI for leaders on DJØF's website: [Examine what AI means for your leadership \(Danish version\)](#)

Rehn, A. Roth-Kirkegaard, C. S. (2026): [On Pulsatile Dynamics, Arrhythmic Revolutions, and AI as an Event](#)



*Impact lead:
David Budtz Pedersen,
Aalborg University*

7.

Responsible use of AI in scientific advice

Subproject: ADD Knowledge
Broker Unit

Context

Across the world, science advisers act as knowledge brokers providing decisionmakers, civil servants, and politicians with information on how science and technology intersect with societal issues. Recent advances in artificial intelligence (AI) have stoked various discussions around large language models (LLMs), such as ChatGPT and others, that can generate text in response to typed prompts. Less discussed is how such technologies might be used constructively, to create tools that summarize scientific evidence for policymaking.

AI-based tools could increase the capacity of science advisers and help policymakers become informed about rapidly evolving issues, emergencies, and debates. But how should such tools be designed and used? In this ADD subproject, we engaged with an international group of scholars and policy professionals to highlight the need for rigorous testing, bias screening, and safeguarding against errors and hallucinations. While LLMs could increase the availability of evidence in government, the use of AI to inform policymakers pose serious risks and uncertainties that research need to address.

Our main finding of the international project is that institutions should experiment with using LLMs to provide decision-support but that such exercises should be closely monitored, supported by codes of conducts, and embedded in ethical reviews. This call for action and the policy recommendations based on the subproject resonated with a broader policy community who

attended workshops, downloaded our research commentaries, and expressed deep engagement with the findings.⁴⁸

The study

AI-based tools are emerging, presenting new opportunities to improve scientific advice to policymakers, making access to evidence more agile, rigorous and targeted. But leveraging such tools for good will require science advisers and policy institutions to create guidelines and to carefully consider the design and responsible use of this nascent technology. In this study, we explored two tasks for which generative AI tools hold promises for policy guidance: (1) synthesising evidence and (2) drafting briefing papers.

Currently, evidence searches are time-consuming and involve various human resources. Hard-pressed science advisers must take what they can get. But what if the evidence searches could utilize algorithmic outcomes? Two main approaches are used to synthesise evidence for policymakers: systematic reviews and subject-wide evidence syntheses. Both require time and resources to run. AI-based platforms are increasingly seeking to make such syntheses less time-consuming, freeing subject-matter experts to focus on more complex analytical aspects.

Systematic reviews – such as Cochrane reviews in health and medicine – identify a question of interest and then systematically locate and analyse all relevant studies to find the best answer. Increasingly, machine learning can automate the

⁴⁸ Tyler, C., Akerlof, K. L., Allegra, A., Arnold, Z., Canino, H., Doornenbal, M. A., Goldstein, J. A., Budtz Pedersen, D. & Sutherland, W. J. (2023) AI tools as science policy advisers? The potential and the pitfalls: *Nature*. 622, 7981

search, screening, and data-extraction processes that form the early stages of systematic reviews. We found that AI tools can be useful in making sense of emerging domains of research, in which review papers and disciplinary journals might be lacking. For instance, techniques for natural language processing can systematically classify research on AI itself, and graph algorithms are being used to detect emerging 'clusters' of research in the broader literature.

Main finding

Large language models are rapidly increasing their ability to synthesise scientific evidence for policymakers. While such technologies might be used constructively for identifying research gaps and simulating policy outcomes, experiments should be guarded and closely monitored by humans and professionals, supported by codes of conducts and ethical review guidelines.

The study found the following key aspects of AI in science for policy:

- **Strengths:** AI can rapidly process evidence, identify gaps, and create predictive models (e.g., in public health or environmental policy). Specialized AI can also flag errors in manuscripts.
- **Risks and pitfalls:** Large Language Models (LLMs) can produce "hallucinations" (false info), inherit biases from training data, and lack the moral judgment necessary for, ethical policymaking.
- **Human-in-the-loop:** Effective use requires human experts to verify AI-generated content, ensure transparency, and manage ethical concerns.
- **Policy adoption:** Research institutions should be developing frameworks to ensure responsible AI use, prioritizing accountability and transparency in, scientific research.

Pathways to impact

The research project's pathway to impact was achieved through strategic and interactive communication with international science-to-policy

communities, networks, and decision-makers. Meetings, panels, and conferences were used to convey the messages and open avenues for mutual discussion between project participants and policymakers, creating relational and conceptual impact pathways. The research paper from the project was published in the leading international journal *Nature*, which helped to disseminate the project's results to a wider audience.

The article generated the following PlumX metrics:

Metrics Details

Citations:	27
Citation Indexes:	24
Scopus:	24
PubMed Central:	1
Policy Citations:	3
Policy Citation:	1
Captures:	54
Readers:	54
Mendeley:	54
Mentions:	1
News Mentions:	1
News:	1
Social Media:	18
Shares, Likes and Comments:	18
Facebook:	18

This article has three policy citations and one further report mention:

1. Tuomi, Ilkka (2024). Fostering knowledge-sharing within and among S4P actors. 7 October 2024. Publications Office of the European Union by Directorate-General for Research and Innovation (European Commission).
2. European Commission (2024). Successful and timely uptake of artificial intelligence in science in the EU. 15 April 2024. Publications Office of the European Union by Directorate-General for Research and Innovation.

3. Royal Netherlands Academy of Arts and Sciences (2024). Successful and timely up-take of Artificial Intelligence in science in the EU. 12 April 2024. KNAW.
4. The Behavioural Insights Team (2024). A Blueprint for Better International Collaboration on Evidence. 9 Sep 2024.

Furthermore, the article and the research were presented at the following events and policy meetings:

- World Science Forum, Budapest, 20-23 November 2024.
- American Association for the Advancement of Science Annual Meeting, Denver, 2024.
- European Commission & US State Department Meeting, “How Might Artificial Intelligence Support Evidence Informed Policymaking?”. 10 November 2023.
- International Network for Government Science Advice (INGSA) Global Meeting “The Transformation Imperative: Expanded Evidence for Inclusive Policies in Diverse Contexts”. Kigali, Rwanda 1-3 May 2024.
- EU Scientific Advice Mechanism and ALLEA – The European Federation of Academies of Sciences and Humanities, “Upholding Integrity in Scientific Advice: Key Principles and Challenges”, Budapest, 4 November 2025.
- OECD and Quebec Chief Scientist Meeting in Brussels, “AI as a critical infrastructure for evidence-informed policymaking?”. General Delegate of Québec. 20 November. 2024.



Policy impact

The study resulting from the subproject was published in *Nature* in September 2023. It was circulated and mentioned in several news posts and online social media, e.g., highlighted by the *Nature* journal itself (see PlumX metrics). The article was mentioned by the **US House Admin Committee Hearing** (September 27, 2023) in Washington DC. Timothy M. Persons (Chief Scientist and Managing Director of the Science, Technology Assessment, and Analytics team of the United States Government Accountability Office) included a reference to the project in his remarks.

On November 10, 2023, the ADD Knowledge Broker David Budtz Pedersen presented the recommendations to the US State Department and US National Science Foundation in Washington DC. On the same occasion, high-level policymakers from the European Commission were present, and interacted with the findings and recommendations.

The article was debated at the US-EU Transatlantic Science Policy Forum “How Might Artificial Intelligence Support Evidence Informed Policymaking?”. The article formed the basis for a scientific session adopted for the official programme at the American Association for the Advancement of Science (AAAS) Annual Meeting in Denver on February 14, 2024, at which occasion David Budtz Pedersen presented findings to an audience of 150 persons.

Conclusion

To realise the potential of AI tools in drawing together evidence while minimising possible drawbacks, the following three recommendations from the project were communicated in open conversation with policymakers:

1. Consistency. Many academic journals use standardised formats for reporting study results, but there is great variation across disciplines. Other sources of information, including working papers, project reports and publications from international agencies, non-governmental organisations and industry, are even more mismatched. Such diversity in presentation makes it difficult

to develop fully automated methods to identify specific findings and study criteria. For example, it is usually important to know over what period an effect was measured or how large the sample was, but this information can be buried in the text. Presenting the research methodology and results in a more consistent manner could help.

2. Credibility. Science advisers judge whether evidence is trustworthy in five ways: the plausibility of the findings (assessed on the basis of the advisers’ subject knowledge and evaluation of the research); the authors’ reputations; the standing of the authors’ institutions; the views of others in the field; and the perspectives of colleagues and peers. This multifaceted judgement is hard to replicate in an AI tool. Publication metrics, such as impact factors and citation counts, are found to be poor measures of research quality. Which dimensions of credibility are most important might also differ, depending on the policy question and context. Experts will need to agree on standards for research quality before these can be automated in AI-based tools – a significant task, although progress is being made.

3. Database selection and access. Currently, conducting systematic reviews requires searching across databases – mostly proprietary ones – to identify relevant scientific literature. The choice of database matters and can have a substantial impact on the outcome. But requirements by governments to publish funded research as open access could make it easier to retrieve study results. For research topics that governments deem as funding priorities, eliminating paywalls will enable the creation of evidence databases and ensure alignment with copyright laws.

While policymakers might use LLMs in preparation of policies, experiments should be guarded and closely monitored by humans and professional staff, supported by codes of conducts and ethical review guidelines and continued joint organisational learning.

This case study shows that AI tools may strengthen science-for-policy work by helping advisers synthesise evidence, identify research gaps, and support the preparation of briefings in fast-moving policy contexts. At the same

time, the study makes clear that such tools are not ready to replace expert judgement: risks of hallucinations, embedded bias, weak credibility assessment, and uneven access to research all make human oversight essential.

The project's main contribution has been to frame a responsible pathway for experimentation, emphasizing consistency in research reporting, clearer standards of credibility, and improved access to evidence as key preconditions for trustworthy use. Through publication and engagement with international science-for-policy networks and policy institutions, the study has helped place these questions on the agenda and contributed to ongoing discussion about how AI can be used responsibly in evidence-informed policymaking.

References

Tyler, C., Akerlof, K. L., Allegra, A., Arnold, Z., Canino, H., Doornenbal, M. A., Goldstein, J. A., Budtz Pedersen, D. & Sutherland, W. J. (2023) AI tools as science policy advisers? The potential and the pitfalls: *Nature*. 622, 7981

Pedersen, D. B., (2025). Principles for Science Advise: A Comparative Framework of International Guidelines., (submitted): *Issues in Science and Technology*.

Pedersen, D. B., (2024). Mapping and Strengthening Ecosystems of Science for Policy, In: *The Transformation Imperative: Expanded Evidence for Inclusive Policies in Diverse Contexts*. Quebec Office of Science: International Network for Governmental Science Advice.

