TATuP Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis
Journal for Technology Assessment in Theory and Practice

oekom science

INTRODUCTION

AI-based decision support systems and society: An opening statement

Diana Schneider[1], Karsten Weber*[,2]

Abstract • Although artificial intelligence (AI) and automated decision-making systems have been around for some time, they have only recently gained in importance as they are now actually being used and are no longer just the subject of research. AI to support decision-making is thus affecting ever larger parts of society, creating technical, but above all ethical, legal, and societal challenges, as decisions can now be made by machines that were previously the responsibility of humans. This introduction provides an overview of attempts to regulate AI and addresses key challenges that arise when integrating AI systems into human decision-making. The Special topic brings together research articles that present societal challenges, ethical issues, stakeholders, and possible futures of AI use for decision support in healthcare, the legal system, and border control.

***KI-basierte Entscheidungsunterstützungssysteme und die Gesellschaft**: Der Versuch einer Einordnung*

Zusammenfassung • *Obwohl künstliche Intelligenz (KI) und automatisierte Entscheidungssysteme schon länger existieren, haben sie erst in jüngster Zeit stark an Bedeutung gewonnen, da sie nun tatsächlich eingesetzt werden und nicht mehr nur Gegenstand der Forschung sind. KI zur Unterstützung von Entscheidungen betrifft somit immer größere Teile der Gesellschaft, wodurch technische, vor allem aber ethische, rechtliche und soziale Herausforderungen entstehen, da nun Entscheidungen von Maschinen getroffen werden können, die bisher in der Verantwortung von Menschen lagen. Diese Einführung gibt einen Überblick über die Versuche, KI zu regulieren, und geht auf zentrale Herausforderungen ein, die sich aus der Integration von KI-Systemen in die menschliche Entscheidungsfindung ergeben. Das Special topic versammelt Forschungsartikel, die gesellschaftliche Herausforderungen, ethische Fragen, Akteur*innen sowie mögliche Zukünfte des KI-Einsatzes zur Entscheidungsunterstützung in der Gesundheitsversorgung, dem Rechtssystem und bei der Grenzkontrolle präsentieren.*

Keywords • *artificial intelligence (AI), decision support, socio-technical systems, regulation, social impacts.*

This article is part of the Special topic "AI for decision support: What are possible futures, social impacts, regulatory options, ethical conundrums and agency constellations?," *edited by D. Schneider and K. Weber.*
https://doi.org/10.14512/tatup.33.1.08

Introduction

In recent years the use of artificial intelligence (AI) systems to support decision-making has become established in various areas of application and has therefore also gained societal significance, as more and more individuals are affected by AI systems in very different situations and contexts. In contrast to systems that decide certain aspects autonomously, decision support systems (DSS) are characterized by the fact that they are merely a decision-making aid for human users. By means of AI, for example, decisions can be prepared by analyzing large amounts of data or recognizing patterns in it. While this can increase the efficiency and accuracy of decisions, it could have a variety of serious and far-reaching effects on individuals, groups, institutions, associations, companies, and society as well as the natural environment.

Since the scope of the impacts and the number of parties affected is so vast, or at least appears to be so vast, extreme scenarios are all too often conjured up in which AI systems either subjugate humanity or solve all of humanity's pressing problems, from climate change to combating pandemics, especially in public discussions about AI. The recent debate on large language models in general and ChatGPT in particular also follows this pattern, with proponents – to put it somewhat tongue-in-cheek – declaring the use of AI systems a panacea and opponents label-

* *Corresponding author: karsten.weber@oth-regensburg.de*
[1] *Fraunhofer Institute for Systems and Innovation Research ISI, Karlsruhe, DE*
[2] *Institute for Social Research and Technology Assessment, OTH Regensburg, Regensburg, DE*

© 2024 by the authors; licensee oekom. This Open Access article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).
https://doi.org/10.14512/tatup.33.1.9
Received: 15. 01. 2024; revised version accepted: 19. 01. 2024;
published online: 15. 03. 2024 (editorial peer review); online version corrected: 08 Nov. 2024

ling them the work of the devil. However, Manichean thinking will hardly help to achieve plausible and realistic impact assessments of AI that can help to minimize or even prevent negative consequences and strengthen positive effects of this potentially disruptive technology.

One fundamental problem with referring to such extreme scenarios is that the actual opportunities and risks of using AI tend to be obscured. In reality, there exists a huge continuum of effects between saving the world and destroying it, the assessment of which will also depend on point of view; what benefits one stakeholder may have negative consequences for another. In addition, the narrative of 'becoming a victim of technology' might reduce the ability of social actors to intervene – if one always sees oneself as a victim or is considered to be a victim, this can prevent stakeholders from even trying to take measures to shape the technology and the social framework in which it should be deployed. From such a passive position, it seems to be difficult or even impossible to discuss AI dispassionately and, for that matter, to make civic, professional, and/or political decisions based on sound information and rational arguments.

It is currently impossible to predict what effects the proposed EU regulation will have on other countries and within the international discourse.

One example of this rather unfortunate situation is the sometimes quite emotional debate surrounding (AI-based) decision support systems in medicine, social work, the judiciary, and many other professional fields that are strongly characterized by human interactions between clients and professionals. This Special topic of TATuP is intended to help developing a differentiated perspective on AI systems in order to counteract premature judgements of AI. By presenting AI systems in various fields of application as well as the challenges and opportunities they create, the aim is also to encourage interdisciplinary dialogue. In the limited space available, it is impossible to cover all areas in which (AI-based) decision support systems are already being applied or could be employed in the foreseeable future. However, it is to be hoped that the examination of individual use cases will also provide insights into domains not covered in this volume.

Attempts to regulate artificial intelligence

The research articles in this Special topic focus predominantly on outlining possible areas of application for (AI-based) decision support systems, identifying stakeholders and describing any potential impacts. Given the novelty of the subject, this is not only an important but also a difficult task. As a result, regulatory issues could only be dealt with marginally or not at all; therefore, at least a few comments on this should be made.

When a provisional agreement on the Artificial Intelligence Act (AIA) was reached on 9 December 2023 after lengthy negotiations in a trialogue between the European legislator, the European Parliament and the Council of the EU, this was heralded as a historic moment in the regulation of AI. The AIA takes a risk-based approach to regulation: While AI systems with no or only low risk are hardly regulated, special requirements apply to high-risk applications, e.g., specific transparency obligations and extensive requirements for data quality, documentation, and traceability (European Union 2023; European Commission 2021). Transparency is considered to be highly relevant in order to interpret AI-generated results and ensure appropriate use (European Commission 2021, reason 47, p. 30) – and thus ultimately contributes to the explainability of AI analyses and recommendations, so seems to be the assumption of EU lawmakers.

The measures proposed in the AIA are primarily aimed at preventing potential risks to the fundamental rights, health, or safety of EU citizens. The debate on the draft regulation is particularly essential for the discussion of AI-supported decision-making systems, as all the fields of application covered in this TATuP Special topic (jurisdiction, law enforcement, and medicine) in principle must be considered particularly sensitive areas. The AI use cases discussed in the research articles can therefore have a considerable impact on the lives of those affected – not only in the event of an error, but also in regular use.

Yet, the AIA was and is on no account the only attempt to regulate the use of AI (Butcher and Beridze 2019; Schiff et al. 2022; Schmitt 2022; Ulnicane et al. 2021). An arbitrary and by no means comprehensive selection shows the range of types of actors and approaches to regulation: For instance, the OECD has formulated the *Recommendation of the Council on Artificial Intelligence*, the EU the *Ethics Guidelines for Trustworthy AI* and the Future of Life Institute the *Asilomar AI Principles*. These and many other documents appear to propose ethical guidelines and codes of ethics for regulation – at least that is what an initial review suggests. However, the binding force of ethical guidelines and codes of ethics is based on voluntary commitment; there is therefore no enforceability that only laws could offer. Moreover, Schiff et al. (2022) emphasize that most of these documents offer little indication of how requirements, recommendations and/or claims they propose can be translated into actionable instructions for the practice of AI development and use, and instead remain at the rather abstract level of moral imperatives.

It can only be assumed that, in view of competing regulatory approaches, the AIA is rather not the last word on the regulation of AI, even more so as criticism of the AIA has not been long in coming. Furthermore, it is currently impossible to predict whether and what effects the proposed EU regulation will have on other countries and within the international discourse. How the application of AI will be regulated in, say, ten years' time in the areas covered in this Special topic as well as in other areas is therefore difficult to predict today.

Human decisions and artificial intelligence

With regard to the question of how AI systems can be specifically integrated into human decision-making, mainly theoretical considerations and only a few empirical studies exist. Many of the following considerations originate from the medical context, as the impact of AI systems on human decision-making has long been the subject of intensive research, particularly in the healthcare professions. For example, Braun et al. (2020) outlined various modes of interaction for the healthcare sector (e.g., the integrative AI-DSS, which can independently request and collect patient data, or the fully automated AI-DSS, which does not require the involvement of professionals); however, most considerations primarily assume direct, essentially bilateral interaction between the professional and the AI system, i. e., a conventional AI-DSS. Simultaneously, there is widespread agreement that the integration of AI systems into the professional decision-making process – regardless of the respective mode of interaction – will have an impact on established work relations, e.g., on the relationship between professionals and patients or employees and employers (Schneider et al. 2022 b). On the one hand, the use of DSS is expected to reduce the workload and the potential time saved is associated with a more empathetic approach to patients (Topol 2019) – expectations that appear questionable given the increasing costs of purchasing and maintaining technology and educating personnel as well as labor shortages. On the other hand, there are concerns that computer paternalism could undermine the essential relationship of trust between healthcare professionals and patients (Čartolovni et al. 2022; Heyen and Salloch 2021) – or, to put it in more general terms, between professionals and clients. Studies already indicate that time and again automation bias occurs (Sujan et al. 2019), i. e., recommendations from AI-based decision-making systems are adopted without question. In view of the frequent lack of data literacy, this poses an enormous challenge, as the AI-generated recommendations can only be used responsibly if they can be correctly understood and interpreted by the users. The perception of AI systems as a second opinion could also raise further ethical questions regarding responsibility (Kempt and Nagel 2022). How AI-based systems could be meaningfully incorporated into shared decision-making processes (e.g., between patients or clients and professionals) also appears to be largely unresolved.

As the use of AI-based systems results in a stronger focus on data and the information, patterns, or meta-information it contains, other forms of professional knowledge could become jeopardized. Particularly in areas of application in which human experience, intuition, tactile or implicit knowledge, but also interpersonal relationships are highly valued (e.g., in the social and healthcare sector as well as in case of judicial or administrative decisions), an inappropriate focus on dataism is a concern and has been strongly criticized in some cases (Pedersen 2019; Devlieghere et al. 2022; Webb 2003). Various papers have pointed out that the data sets used for AI-based systems are fragmented (Tucker 2023), may contain deliberate omissions (Schneider 2022), or that administrative data sets are unsuitable for assessing professional issues (Gillingham 2015, 2020). Besides the fact that most training datasets have a strong bias and are poorly representative in terms of ethnic origin and gender, for example, there is also the problem that particularly vulnerable and/or stigmatized groups of people are often underrepresented.

However, analyzing large data sets opens up the possibility of contributing to evidence-based practice. But this requires that the algorithms underlying the AI-based recommendations are not exclusively pattern-based, but also incorporate concepts and theories from current research and knowledge – otherwise it would be almost impossible to make valid statements (Schneider et al. 2022 a).

If AI systems suggest decisions and users routinely adopt them, it is not the technology that has changed but the way it is used.

These short comments can only highlight a few aspects regarding the use of AI systems for decision support. For instance, the differentiation between systems for automatic decision-making (human-out-of-the-loop) and for decision support (human-in-the-loop, human-on-the-loop) should certainly be dealt with in much more detail, as different questions are raised depending on how AI systems are actually employed. It would also be worth to examine whether and how transitions from decision support systems to automatic decision-making systems might take place; this is less a technical issue than an organizational and practical one, because if AI systems suggest decisions but users routinely adopt them, it is not the technology that has changed but the way it is used. Such transitions in modes of use can in turn lead to far-reaching changes in the respective understanding of the profession and this in turn can again change modes of use

https://doi.org/10.14512/tatup.33.1.9

(Schneider et al. 2022 b). In other words: When talking about AI, this must always be done in terms of a socio-technical system.

Contributions in this Special topic

The six contributions to this TATuP Special topic cover the use of AI systems in three different domains: healthcare, legal system, and law enforcement or, more precisely, border control. We decided to cluster thematically related research articles and to sort them in the clusters according to the alphabetical order of the names of the first authors – this seemed to us to be the best variant of an ultimately arbitrary arrangement. The first three research articles deal with the use of AI in the legal system, followed by a research article on AI systems used to identify illegal migrants at the border, and finally two research articles on the application of AI systems in medical contexts.

- Anna-Katharina Dhungel and Moreen Heine are investigating whether, how and who would benefit from the use of AI systems in the legal system. They will attempt to answer this question based on interviews with judges from Germany. On the one hand, this means a limitation, as the results are only meaningful in the context of the German legal system, yet on the other hand, comparable studies have so far mostly been carried out in countries, such as the United States, whose legal system is structured significantly differently to the German system. The article therefore fills a research gap.
- Brandon Long and Amitabha Palmer focus on a different target group, as they look at the question of whether AI systems could have an advantage for jurisdiction from the perspective of the users of the legal system in the United States. Their primary interest is whether AI systems could be used as cost-effective advisors, in particular for socially disadvantaged people, providing, for example, reliable information about the potential success of a lawsuit.
- With reference to the legal system, Giovana Lopes asks whether AI systems could be used to identify prejudices and partialities – bias for short – among judges, to alert them to the existence of bias, thereby inducing behavioral and attitudinal changes on the part of judges and thus ensuring fairer judgments.
- In their research article on the use of AI in the context of border controls, Daniel Minkin and Lou Therese Brandner explore the theoretical assumptions underlying the 'iBorderCtrl' system. They conclude that the theoretical foundations of the system are questionable and that the use of the system to recognize false statements made by people entering a country should therefore be called into question. This indicates that in the context of technology assessment not only technology, but also fundamental and/or theoretical assumptions that are incorporated into technology should be taken into account.
- Manuela Marquardt, Philipp Graf, Eva Jansen, Stefan Hillmann, and Jan-Niklas Voigt-Antons use a scenario-based interview study to investigate which requirements AI systems in the medical context must fulfil in terms of explainability so that their outputs are comprehensible for all stakeholders. Without situationally adapted and comprehensible explanations, building trust in AI systems would hardly be possible and their use would therefore be called into question.
- Finally, Jan C. Zoellick, Hans Drexler, and Konstantin Drexler address the use of AI systems for the detection of melanoma. Using three scenarios, they consider shifts in competences from physicians to the systems used, examine the need for regulation of the use of AI systems in diagnostic contexts and finally look at a conundrum that is often brought up but rarely clearly resolved: What should be done if humans and machines come to different conclusions? Whose conclusion should take priority?

Conclusion

As already indicated, the contributions in this Special topic cannot address all aspects of AI-based decision-making support. Regarding technology assessment they do, however, provide examples of the topics and issues raised by the rapid and ubiquitous introduction of AI technologies. In the case of country-specific studies, the potential for drawing generalizations may be limited, but particularly from a technology assessment perspective, such specific studies cannot be dispensed with, as the effects of technology are determined to a considerable extent by the prevailing conditions. Discussions of specific topics usually differ not only from country to country, but also within a profession in a country where there are different and long-established strands of discourse with corresponding arguments and assumptions, e.g. in the field of the digitalization of social work in Germany (Waag and Rink 2023). Country comparisons and interdisciplinary studies can therefore help to make such discourses more comprehensible and transparent. The comparison of arguments and assumptions can also help to uncover blind spots in one's own argumentation. A differentiated view that is considering country-specific characteristics and social conditions is also indispensable with regard to impact assessment and evaluation of technology, as otherwise there is a risk of not being able to move beyond thinking in terms of extreme scenarios. While the research articles in this TATuP Special topic refer to similar challenges and issues, they also illustrate the importance of detail and differentiation, despite the variety of subjects covered.

Funding • This work received no external funding.

Competing interests • The authors declare no competing interests except that author Karsten Weber is a member of TATuP's scientific advisory board. This did not affect the peer-review process.

Acknowledgement

The Special topic editors would like to thank the authors and reviewers for the professional and most cooperative collaboration.

References

Braun, Matthias; Hummel, Patrik; Beck, Susanne; Dabrock, Peter (2020): Primer on an ethics of AI-based decision support systems in the clinic. In: Journal of Medical Ethics 47 (12), p. e3. https://doi.org/10.1136/medethics-2019-105860

Butcher, James; Beridze, Irakli (2019): What is the state of artificial intelligence governance globally? In: The RUSI Journal 164 (5–6), pp. 88–96. https://doi.org/10.1080/03071847.2019.1694260

Čartolovni, Anto; Tomičić, Ana; Lazić Mosler, Elvira (2022): Ethical, legal, and social considerations of AI-based medical decision-support tools. A scoping review. In: International Journal of Medical Informatics 161, p. 104 738. https://doi.org/10.1016/j.ijmedinf.2022.104738

Devlieghere, Jochen; Gillingham, Philip; Roose, Rudi (2022): Dataism versus relationshipism. A social work perspective. In: Nordic Social Work Research 12 (3), pp. 328–338. https://doi.org/10.1080/2156857X.2022.2052942

European Union (2023): Briefing – EU legislation in process. Artificial intelligence act. Available online at https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf, last accessed on 18. 01. 2024.

European Commission (2021): Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts. Brussels: European Commission. Available online at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52021PC0206, last accessed on 18. 01. 2024.

Gillingham, Philip (2015): Electronic information systems in human service organisations. The what, who, why and how of information. In: British Journal of Social Work 45 (5), pp. 1598–1613. https://doi.org/10.1093/bjsw/bcu030

Gillingham, Philip (2020): The development of algorithmically based decision-making systems in children's protective services. Is administrative data good enough? In: The British Journal of Social Work 50 (2), pp. 565–580. https://doi.org/10.1093/bjsw/bcz157

Heyen, Nils; Salloch, Sabine (2021): The ethics of machine learning-based clinical decision support. An analysis through the lens of professionalisation theory. In: BMC Medical Ethics 22 (112), 9 pp. https://doi.org/10.1186/s12910-021-00679-3

Kempt, Hendrik; Nagel, Saskia (2022): Responsibility, second opinions and peer-disagreement. Ethical and epistemological challenges of using AI in clinical diagnostic contexts. In: Journal of Medical Ethics 48 (4), pp. 222–229. https://doi.org/10.1136/medethics-2021-107440

Pedersen, John (2019): The digital welfare state. Dataism versus relationshipism. In: John Pedersen and Adrian Wilkinson (eds.): Big Data. Promise, application and pitfalls. Cheltenham: Edward Elgar Publishing, pp. 301–324. https://doi.org/10.4337/9781788112352.00019

Schiff, Daniel; Laas, Kelly; Biddle, Justin; Borenstein, Jason (2022): Global AI ethics documents. What they reveal about motivations, practices, and policies. In: Kelly Laas, Michael Davis and Elizsabeth Hildt (eds.): Codes of ethics and ethical guidelines. Cham: Springer International, pp. 121–143. https://doi.org/10.1007/978-3-030-86201-5_7

Schmitt, Lewin (2022): Mapping global AI governance. A nascent regime in a fragmented landscape. In: AI and Ethics 2 (2), pp. 303–314. https://doi.org/10.1007/s43681-021-00083-y

Schneider, Diana (2022): Ensuring privacy and confidentiality in social work through intentional omissions of information in client information systems. A qualitative study of available and non-available data. In: Digital Society 1 (26), 21 pp. https://doi.org/10.1007/s44206-022-00029-9

Schneider, Diana; Maier, Angelika; Cimiano, Philipp; Seelmeyer, Udo (2022 a): Exploring opportunities and risks in decision support technologies for social workers. An empirical study in the field of disabled people's services. In: Sozialer Fortschritt 71 (6–7), pp. 489–511. https://doi.org/10.3790/sfo.71.6-7.489

Schneider, Diana; Sonar, Arne; Weber, Karsten (2022 b): Zwischen Automatisierung und ethischem Anspruch. Disruptive Effekte des KI-Einsatzes in und auf Professionen der Gesundheitsversorgung. In: Mario Pfannstiel (ed.): Künstliche Intelligenz im Gesundheitswesen. Entwicklungen, Beispiele und Perspektiven. Wiesbaden: Springer, pp. 325–348. https://doi.org/10.1007/978-3-658-33597-7_14

Sujan, Mark et al. (2019): Human factors challenges for the safe use of artificial intelligence in patient care. In: BMJ Health & Care Informatics 26 (1), p. e100 081. https://doi.org/10.1136/bmjhci-2019-100081

Topol, Eric (2019): Deep medicine. How artificial intelligence can make healthcare human again. New York, NY: Basic Books.

Tucker, Catherine (2023): Algorithmic exclusion. The fragility of algorithms to sparse and missing data. Working Paper. In: Brookings Institution, 02. 02. 2023. Available online at https://www.brookings.edu/research/algorithmic-exclusion-the-fragility-of-algorithms-to-sparse-and-missing-data/, last accessed on 17. 01. 2024.

Ulnicane, Inga; Knight, William; Leach, Tonii; Stahl, Bernd; Wanjiku, Winter-Gladys (2021): Framing governance for a contested emerging technology. Insights from AI policy. In: Policy and Society 40 (2), pp. 158–177. https://doi.org/10.1080/14494035.2020.1855800

Waag, Philipp; Rink, Konstantin (2023): Digitalisierung als Irritation. Von ideologischen zu reflexionstheoretischen Selbstbeschreibungen der Sozialen Arbeit im Zuge ihrer Auseinandersetzung mit digitalen Technologien. In: Neue Praxis 23 (4), pp. 292–306.

Webb, Stephen (2003): Technologies of care. In: Elizabeth Harlow and Stephen Webb (eds.): Information and communication technologies in the welfare services. London: Jessica Kingsley Publishers, pp. 223–238.

DIANA SCHNEIDER

is research associate at the Competence Center Emerging Technologies at the Fraunhofer Institute for Systems and Innovation Research ISI since 2021. She was a PhD candidate of the NRW Digital Society research program from 2018–2022. Her research focuses on innovations in social and healthcare systems, in particular their ethical and social implications.

PROF. DR. KARSTEN WEBER

is research professor for Technology Assessment and AI-based Mobility at OTH Regensburg since 2022 and has been working there as senior researcher since 2013. His research is focused on impacts of technology on individuals, groups, society, and environment, namely information and communication technology, particularly in healthcare, mobility, and energy.

Correction to Schneider and Weber (2024)

Received: 15. 01. 2024; revised version accepted: 19. 01. 2024;
published online: 15. 03. 2024 (editorial peer review); **online version corrected: 08 Nov. 2024**.

In the article "AI-based decision support systems and society: An opening statement" by Diana Schneider and Karsten Weber (TATuP – Journal for Technology Assessment in Theory and Practice, https://doi.org/10.14512/tatup.33.1.9), there was an error in the 'Competing Interests' section. The correct **'Competing Interests' statement** is as follows:

The authors declare no competing interests except that author Karsten Weber is a member of TATuP's scientific advisory board. This did not affect the peer-review process.