Eva Buddeberg* and Fruzsina Molnár-Gábor

Introduction

https://doi.org/10.1515/auk-2026-3007

As is evident everywhere, research into artificial intelligence has advanced by leaps and bounds in recent years, leading to a rapid increase in its use across almost all areas of life. As early as 2018, the German government of the time set out a 'Master Plan for Artificial Intelligence' in its coalition agreement, and the current government plans, as announced in a recently published strategy paper, to quadruple Germany's AI capacity by 2030. Artificial intelligence is regarded as a key technology that is essential for Germany's economic competitiveness. The development and use of AI, as well as its societal implications, are therefore a central focus of political decision-making and of debates in media and academia.

The four contributions gathered here emerged from discussions within a research network on 'AI and Responsibility' led by Eva Buddeberg and Fruzsina Molnár-Gábor, which was funded by the German Research Foundation (DFG) from 2022 to 2025.[1] Its focus was on the question of how responsibility can be appropriately described, justified and normatively assessed in the context of artificial intelligence (AI), and under what conditions responsibility can be meaningfully attributed in complex contexts of action characterised by algorithmic systems.

Whilst AI systems are now capable of recognising complex patterns and thus independently preparing decisions or automatically controlling processes, normative assessment and the attribution of responsibility remain the preserve of human actors. This tension between technological capability and human responsibility formed the starting point for this interdisciplinary project, in which both fundamental and application-oriented aspects of responsibility were systematically examined in the light of these technological developments. Through an exchange of perspectives from philosophy, law, sociology and theology, as well as the engineering sciences, the aim – particularly from a legal perspective – was to establish a

1 German Research Foundation (DFG) project „Künstliche Intelligenz: Wer trägt Verantwortung für welchen Schaden einer Handlung" [Artificial intelligence: Who bears responsibility for what damage of an action?] (BU 3980/3-1), Project number: 512921364, December 2022 to November 2025.

***Corresponding author: Eva Buddeberg**, Institut für Philosophie, Philipps-Universität Marburg, Marburg, Germany, E-mail: eva.buddeberg@uni-marburg.de
Fruzsina Molnár-Gábor, Juristische Fakultät der Universität Heidelberg, BioQuant-Zentrum, Heidelberg, Germany, E-mail: fruzsina.molnar-gabor@uni-heidelberg.de

sound framework for allocating areas of responsibility when using artificial intelligence in complex technical systems. To this end, normative criteria were first identified that allow for the appropriate assessment of AI-supported decision-making processes and the definition of duties of care for the application of AI, in order to derive a liability framework that would also apply to future, unforeseeable developments in connection with AI.

We believe that the insights gained are relevant not only to the individual disciplines represented here, but also to the broader public discourse and political decision-making processes. They can provide a basis for future regulatory frameworks in the field of AI and help to create responsible and legally sound conditions for the use of algorithmic systems. Furthermore, they open new fields of research, such as the institutional embedding of prospective responsibility, the further development of liability models for autonomous systems, or the empirical investigation of societal acceptance of algorithmic decisions.

Beyond questions regarding the legal framework of responsibility and the development of specific duties of care, the focus from a philosophical and sociopolitical perspective was also, above all, on alternative conceptions of shared responsibility and issues of justice in the development and deployment of AI systems. Thus, the four contributions gathered here propose – from legal, philosophical and sociological perspectives – a more comprehensive understanding of shared or distributed joint responsibility and examine the normative and political issues arising from the use of AI systems, for example in computer-assisted selection processes.

In her article 'Shared responsibility for the development and use of artificial intelligence', Eva Buddeberg argues from a philosophical perspective that, going beyond liability or duty of care, we bear a discursive shared responsibility to scrutinise the development and use of AI systems against the yardstick of justice, with the aim of uncovering and counteracting injustices. She demonstrates that the now widespread use of AI in almost all areas of life tends to exacerbate existing injustices. These largely rapid developments also require us to further question, modify or, where necessary, redesign the normative framework and the associated understanding of ourselves and the world.

Drawing on insights from sociology and the studies of science and technology, Cordula Kropp and Tobias Renner develop, in their article 'Automation, Co-Agency, and Distributed Responsibility: Caring for Hybrid Therapeutic Networks' a concept of 'co-agency' as an emergent and relational property, with the aim of capturing the responsible interaction between humans and automated systems. Using a concrete example from the medical/therapeutic context, they argue that responsibility should not be attributed to individual actors – whether human or digital – but rather understood as something that arises within relational processes and should

be conceived as distributed responsibility within hybrid human-machine networks. Drawing on this insight, the authors derive guidelines for a relational duty of care regarding the stability and accountability of such infrastructures.

In his article 'Framing computational fairness and non-discrimination', Wilfried Hinsch outlines a framework for computational fairness in the analysis of data-driven statistical discrimination. He claims that the first step is to distinguish between issues of procedural justice and issues of distributive justice. Computational procedural fairness, he argues, can only be measured by the parity of predictive values in all demographic groups. Thus, a computer-based procedure can only be non-discriminatory on the condition that it achieves the same predictive values across all demographic groups.

In 'Perfect lawmaking and perfect legal compliance: two false ideals of normativity in governance by AI', Klaus Günther finally analyses, from a legal-theoretical perspective, the notion that AI could enable a near-perfect normative order through, for example, data-driven lawmaking or the automated enforcement of rules. As Günther demonstrates, this vision is based on two problematic ideals: the ideal of perfectly rational lawmaking and that of complete compliance with rules. As Günther criticises, this overlooks the central importance of conflict, interpretation and democratic negotiation processes in law. This is problematic both normatively and politically, as it contributes to a technocratic conception of social governance.

As is readily apparent, the issue of 'AI and responsibility' is a vast field. The perspectives presented here can and should only offer an insight into what is at stake and how the problems and challenges outlined might best be addressed. Much remains open to debate, and some points may well need to be revised in the near future, not least because technological developments are transforming our entire way of life at a dramatic pace. Yet we do not, in our view, face this situation simply as passive observers. Rather, interdisciplinary and socio-political dialogue could and should make it possible – particularly through political and legal regulation – to at least contain the inherent risks and, above all, the structural injustices. In this way, the positive potential associated with technological developments could then indeed serve to better address the current problems facing our society.