Introduction

Nicholas Kluge Corrêa, Julia Maria Mönig

The rapid development of artificial intelligence (AI) technologies, the potential ethical and legal issues arising from their deployment, and the legal requirements, including those under the European AI Act, have increased the need for reliable, transparent, and independent auditing structures and mechanisms. While certain technical characteristics might be easy to assess, and it might be possible to refer to existing audits, AI technologies potentially present specific challenges to human rights, ethical values, and constitutional democracy.

Hence, what a certification could look like is only one of the questions that should be asked when discussing ethical certification of AI. Furthermore, how all potentially concerned parties (stakeholders) can be included, which values might be affected, and what this means for those actors who (rightfully) need and want to profit from AI applications were central questions that the philosophy team of the KI. NRW-flagship project „Zertifizierte KI" (Certified AI) investigated. Further key questions raised in the contributions and throughout the project period include: Can ethics be standardized? Who audits the auditors? How can auditing institutions be aligned with actual societal concerns? What can an assessment look like for a specific area (e.g., human resources)? How are law and ethics intertwined? Which contribution can philosophy provide, now that there is a European regulation on Artificial Intelligence (e.g., in the discussion whether robots in elderly care should be considered as deceptive technology under the AI Act)? Finally, what role do standardization bodies play in ensuring trustworthy AI?

Answers to these questions are presented in this book, as well as in the project's central outcomes: the *Catalogue of General Ethical Requirements for AI Certification* and the *Catalogue of Field-Specific Requirements for AI Certification*. The *Catalogue of General Ethical Requirements for AI Certification* provides a structured framework for trustworthy AI development, identifying overarching ethical requirements for AI certification and outlining six ethical values: fairness,

privacy, robustness, sustainability, transparency, and truthfulness. For each value, the project team researched existing tools, metrics, and approaches to help developers, programmers, and other practitioners build trustworthy AI systems. The *Catalogue of Field-Specific Requirements for AI Certification* offers a list of requirements, an ethical stakeholder analysis, and a discussion of potential value tensions for six exemplary fields: arts, biometric profiling, industry and critical infrastructure, law enforcement, medicine and care and social media, a list of requirements, an ethical stakeholder analyses and a discussion of potential value tensions.

The volume at hand collects the contributions to the final conference of the sub-project philosophy of the KI.NRW-Flagship-Projekt "Zertifizierte KI". The topic "Standardization of AI Ethics: Stakeholders, Values, and Profit" offered the opportunity to critically discuss the above-mentioned questions and current tendencies in AI auditing and certification, and to raise ethical issues that the current schemes might entail or overlook.

In his paper "Should AI Auditors Be Audited? Challenges and Paths for Meta-Auditing Artificial Intelligence", Marcelo Pasetti argues that while AI auditing in various forms is considered necessary and desirable in the European Union and other countries, it often falls short of an independent assessment. He proposes a meta-audit to guarantee their impartiality and objective analyses.

Chiara Marcoccia examines the challenge of connecting auditing and governance in the case of AI for image generation. In her paper entitled "Citizens infrastructures as a way to govern AI's power to shape our shared representations", she proposes to leverage social sciences to address power asymmetries in genAI auditing, and their consequences across societal levels.

Gaia Contu examines in her contribution "Social Robots in Elderly Care: Ethics, Regulation, and Design" whether this is true and, if yes, whether, under the European AI Act, social robots in elderly care would need to be considered as "high risk" according to Article 6 of the Act. After discussing several other ethically issues with AI and robotics – the (in)famous COMPAS case, the often-quoted Amazon-algorithm for the selection of job applicants and the equally (in)famous Dutch childcare scandal, Contu concludes, that social robots in elderly care are not per se deceptive and suggests therefore a framework to estimate the risk posed by social robots in elderly care.

In the context of robotics for elderly care, it is being debated whether the use of social robots constitutes deception for the persons concerned. Gaia Contu examines in her contribution "To Be Cared For or Deceived? Opera-

tionalizing Ethics in the Case of Elderly Care Robots under the European AI Act" whether this claim holds and whether it can be addressed under Article 5 of the AI Act on prohibited AI practices. Drawing on considerations of the non-neutrality of technology, illustrated by well-known cases such as COMPAS and the often-cited Amazon recruitment algorithm, Contu concludes that ethical considerations must be operationalized from the earliest stages of designing robots for elderly care but that robots used in elderly care are not deceptive per se.

Tina Lassiter and Kenneth R. Fleischmann present the findings of a study in which they investigated what HR Professionals' and Jobs Seekers' perspectives on "Algorithmic Audits in Human Resources" in the US are. The study highlights substantial differences between the two stakeholder groups regarding knowledge of AI use, trust in AI and HR-AI tools, and their knowledge and opinions regarding algorithmic auditing. As specific legislation, they look into the European AI Act and the New York City Law 144.

Finally, Maria Mensch, Adrian Seeliger, and Johannes Wellhöfer offer a practitioner's perspective on how Standard Developing Organizations (SDOs) might contribute to the design, deployment, and development of ethical AI. Their view from the German Standardization Organization DIN suggests that standardization can contribute to aligning technological advancements with societal values, ensuring responsible innovation while safeguarding fundamental rights.

The papers in this volume were presented at the eponymous conference held in September 2025 at the Center for Science and Thought of the University of Bonn. The editors would like to thank those who helped to make this event a success (on stage and behind the scenes). Thanks are also due to Chris Wickenden, who not only enriched our conference with his art works, but also granted us permission to use his image "Suspense of Identified Companionship", created in exchange with the AI character Preti O'Sum, for the cover design. The project including the publication of this volume was made possible through the funding by the Ministerium für Wirtschaft, Industrie, Klimaschutz und Energie des Landes Nordrhein-Westfalen.