Book review: Coeckelbergh, Mark (2022): The political philosophy of AI
Schmidt, Michael W.

Veröffentlichungsversion / Published Version
Rezension / review

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

oekom
science

Book review: Coeckelbergh, Mark (2022): The political philosophy of AI

Michael W. Schmidt*,1 ⓘD

Mark Coeckelbergh starts his book with a very powerful picture based on a real incident: On the 9th of January 2020, Robert Williams was wrongfully arrested by Detroit police officers in front of his two young daughters, wife and neighbors. For 18 hours the police would not disclose the grounds for his arrest (American Civil Liberties Union 2020; Hill 2020). The decision to arrest him was primarily based on a facial detection algorithm which matched Mr. Williams' driving license photo with the picture of a man who was suspected of watch theft two years earlier. Not only did the computer 'get it wrong' as one of the detectives said, when Mr. Williams made them aware that the picture of the suspect obviously wasn't resembling him, the probably unreliable algorithm very likely contributed to racial discrimination (Hill 2020).

It is well documented that many available facial detection algorithms at this time had significant problems (e.g. a comparably high false positive rate) with respect to black persons, like Mr. Williams (NIST 2019). Multiple causes may exist, such as unbalanced training datasets and insufficient optimization. Coeckelbergh compares the disturbing case of Mr. Williams with a political interpretation of Franz Kafka's *The Trial*, where the protagonist, Josef K., is accused of an unspecified crime by an opaque, oppressive and absurd bureaucracy: "In the 21st-century United States, Josef K. is black and is falsely accused by an algorithm, without explanation" (p. 2).

This dire picture highlights that the emerging technology of artificial intelligence (AI), in its various forms, is ever more pervading our societies and impacting our collective or individual lives. And, that it is naïve to consider AI technologies, or any other technology, as a *per se* politically neutral tool (pp. 3 f., 59 ff.; Winner 1980). Arrests based on AI processed evidence might, especially for marginalized groups, lead to serious harm. Decisions by an automated vehicle are also sometimes a mat-

ter of life and death, especially for vulnerable road users. Deep fakes in social media might significantly impact the outcome of democratic elections. From a social perspective, therefore, the application of AI involves considerable associated risks. However, there are also relevant opportunities. How to assess the impacts of AI technologies on our values and how to act on this assessment? According to Coeckelbergh, we need clarity concerning inherently political concepts, such as *freedom*, *democracy* and *justice*, to adequately answer these questions. What do we *mean* when we say an AI based decision racially discriminated? Political philosophy deals with the theoretical and normative reflection of such concepts and can help explicating our concerns and expectations, situating them within our web of beliefs and assessing their strength and plausibility. Coeckelbergh's book aims to enable scholars from multiple disciplines and fields to constructively pick up threads from political philosophy for their research and to contribute to the quality of the general public debate on AI.

Freedom, justice, democracy, power and a challenged anthropocentrism

Coeckelbergh organizes the overview of the possible combination of political philosophy and reflection on AI by discussing important concepts and theoretical approaches from political philosophy. To give an example from chapter 2, which is focused on notions connected to the concept of freedom: With regard to John Stuart Mill's liberalism we can locate the burden of proof for the legitimacy of AI aided predictive policing or pervasive surveillance on the side of those who approve, or use, these coercive or intrusive measures (pp. 11–16). Further issues discussed in this chapter include AI-aided manipulation and the critique of libertarian paternalism or self-realization and emancipation regarding the commodification of personal data and AI-based automation.

The other main chapters are concerned with discussions related to justice (chapter 3), democracy (chapter 4), power (chapter 5) and challenges to anthropocentrism by post- and transhumanist theory (chapter 6). In general, Coeckelbergh succeeds in not only providing a comprehensive overview of important

* *Corresponding author: michael.schmidt@kit.edu*

1 *Institute for Technology Assessment and Systems Analysis,*
Karlsruhe Institute of Technology, Karlsruhe, DE
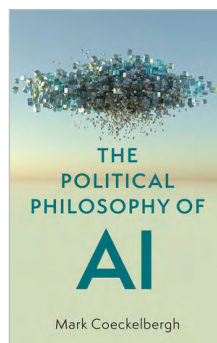
debates from political philosophy for the reflection on AI, but also presents a considerable variety of theoretical approaches in a well-informed and accessible way. These include the systems of thinkers such as Hegel, Marx, Arendt or Foucault, varieties of liberalism, libertarianism, critical theory, radical or agonistic theories of democracy, as well as identity-based approaches and post-colonialism.

Of course, some aspects are open to criticism. Chapter 4 focuses intensively on social media. However, it is not easy to identify the degree to which AI or rather other features of current social media, like the for-profit orientation, are responsible for the formation of filter bubbles and echo chambers. More attention

ingly consist of complex socio-technical systems. Correspondingly, there is a recent trend to call for political philosophy or normative inquiries in collective action – as Maarten Franssen framed it at the Forum for Philosophy, Engineering, and Technology 2023 in Delft (Mitcham forthcoming) – in philosophy of technology and engineering as well as in technology ethics (Himmelreich 2019).

Coeckelbergh's book is a much-needed introduction of how to relate political philosophy to fields and disciplines concerned with the reflection on technology. As Coeckelbergh stresses, political philosophy cannot be simply applied to technology and engineering (p. 150) – conceiving socio-technical systems as es-

In light of currently emerging disruptive technologies the political dimension of engineering and technology seems to become more apparent again.

69

could have been devoted to the increasing risk deep fakes pose for democracies' public sphere.

Not all relevant topics and concepts are covered extensively. One example is utopian thinking within political philosophy. Naturally, it must be said that it is simply unfeasible to cover everything in an introductory book. A related fact is that the book tends to focus on the risks of AI. The opportunities with regard to our social values, also deserves due consideration. Can AI-aided translation, for example, enable a real agora for language wise rather Babylonian structures such as the European Union?

Reflection on technology and engineering needs more political philosophy – and vice versa

Coeckelbergh concludes (chapter 7) with the provocative thesis that "political philosophy in the 21st century can no longer be done, and should no longer be done, without responding to the question concerning technology" (p. 150) and speculates, whether there should ultimately be a merging between reflection on technology and reflection on politics. The latter might be slightly exaggerated; however, I fully agree with the diagnosis that philosophy of technology and engineering and related fields need more political philosophy – and vice versa.

Carl Mitcham explains that "[…] despite the early presence of political philosophy in 1970s philosophy of technology, it has tended for more than 50 years to be marginalized: First, by an emphasis on ethics alone separated from politics; second, by a turn away from ethics itself" (Mitcham forthcoming). In light of currently emerging disruptive technologies – besides and in combination with AI, for example, gene editing, quantum computing or nuclear fusion – the political dimension of engineering and technology seems to become more apparent again. To speak in Rawlsian terms: the basic structures of our societies increas-

sential parts of the basic structure of society, will lead us to rethink and interpret established theories, such as Rawls's political liberalism (Binns 2018; Gabriel 2022; Hoffmann 2020). Work, defending a theoretical position at this intersection, is a necessary next step.

References

American Civil Liberties Union (2020): Wrongfully arrested because of flawed face recognition technology. Available online at https://www.youtube.com/watch?v=Tfgi9A9PfLU (video), last accessed on 23. 01. 2024.

Binns, Reuben (2018): Algorithmic accountability and public reason. In: Philosophy & Technology 31 (4), pp. 543–556. https://doi.org/10.1007/s13347-017-0263-5

Gabriel, Iason (2022): Toward a theory of justice for artificial intelligence. In: Daedalus 151 (2), pp. 218–231. https://doi.org/10.1162/daed_a_01911

Hill, Kashmir (2020): Wrongfully accused by an algorithm. In: The New York Times, 24. 06. 2020. Available online at https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html, last accessed on 23. 01. 2024.

Himmelreich, Johannes (2019): Ethics of technology needs more political philosophy. In: Communications of the ACM 63 (1), pp. 33–35.

Hoffmann, Anna Lauren (2020): Rawls, information technology, and the socio-technical bases of self-respect. In: Shannon Vallor (ed.): The Oxford handbook of philosophy of technology. Oxford: Oxford University Press, pp. 230–249. https://doi.org/10.1093/oxfordhb/9780190851187.013.15

Mitcham, Carl (in press): Brief for the political philosophy of engineering and technology. In: Science and Engineering Ethics.

NIST – National Institute of Standards and Technology (2019): NIST study evaluates effects of race, age, sex on face recognition software. Gaithersburg, MD: NIST. Available online at https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software, last accessed on 23. 01. 2024.

Winner, Langdon (1980): Do artifacts have politics? In: Daedalus 109 (1), pp. 121–136.