

Media Evolution

If only the lake could talk

Futures of AI for Sustainability

Media Evolution

If Only the Lake Could Talk

Futures of AI for Sustainability

Colophon

Published by

Media Evolution
as a part of DigiT Hub AI

Editors

Martin Thörnkvist
Reeta Hafner
Rowan Drury

Copy editor

Rowan Drury

Illustrations

Jon Koko

Graphic concept

HolsterGreen Studio

Layout

Linnea Paulsson
Neppelberg

Paper

Munken Polar

Printing

Exakta, Malmö

ISBN

978-91-987389-1-9

Published

August 2022

Contributors

Anders Lindberg
Basudha Bhattarai-
Johansson
Dennis Munetsi
Elin Westerberg
Elisabet Sjölund
Harald Klein
Helle Foldøy
Henning Gross
Henrik Challis
Himanshu Rohilla
Jason Tucker
Jess Haynie-Lavelle
Julia Zajac
Juliija Rukanskaitė
Kajsa Westman
Katja Subrizi Wessling
Kristin Heinonen
Layla Husain
Maria Malcus
Mathias Torp
Michael Strange
Murat Samanci
Nell Watson
Nikolaj Møller
Ninon Moraw
Paolo Nardi Fernandez
Peter Neubauer
Petra Jenning
Rasmus Hedin
Rowan Drury
Sara Murray
Sonja Rattay
Valbon Gurmani
Victor Friberg

Some rights reserved. This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Published by

About Media Evolution

Media Evolution was founded in 2008 as a joint initiative between the private, academic and public sectors to promote the conditions for growth and innovation in the media industries. Over time, the idea of “media” has matured to incorporate all sorts of organisations dealing with digital development and systems change. Today Media Evolution is owned by more than 200 members.

This book is a result of Media Evolution’s process for collaborative foresight, in which experts from member organisations come together to explore and make sense of the future and find ways to impact change for the better.

www.mediaevolution.se

About DigIT Hub AI

In the project DigIT Hub AI ten organisations and three universities in Skåne and Blekinge have joined forces to help small and medium-sized companies improve their operations with the help of AI. It is a part of the EU’s ambition to use digital technology to stimulate a green and sustainable recovery after the pandemic.

DigIT Hub AI is co-funded by the European Regional Development Fund.

www.digithub.se/ai



EUROPEAN
UNION
European Regional
Development Fund

Contents

Preface *I*

Future 1: Nowhere to Hide 8

Demanding Transparency and Accountability *II*

On AI Moral Advisors for Sustainability *19*

Rethinking Governance for Resilient AI Futures *35*

Future 2: Reconnection and Recognition *48*

Putting Nature Back at the Centre *51*

AI, Spirituality and Indigenous Wisdom *59*

Disruptive Possibilities: AI and Planetary Health *73*

Future 3: Who Decides? *80*

Facilitating Decision-making and Governance *83*

Building and Regulating Ethics into AI *89*

Simulation Day *99*

If Only the Lake Could Talk

Future 4: Human and Machine *106*

Enhancing Human Capabilities and Creativity *109*

AI and the Challenge of Speculative Ethics *115*

Is AI Creative or a Tool for Creativity? *123*

Epilogue *143*

Humanity Maturing with Technology *145*

If the Lake Could Talk *151*

Sonja Rattay

AI and the Challenge of Speculative Ethics

Current modern socio-technical imaginaries¹ of AI pull in opposite directions. The last few decades have revealed a multitude of challenges brought on by the digital transformation of society, while many concerns are expressed about the uncertainty of where the current development of AI might end up and where the directive of established development patterns is taking us. There seems to be a shared consensus that AI offers the potential to solve a variety of blurrily defined challenges for humankind, along with the suspicion that “AI taken to the extreme” holds dangers and threats. On the utopian side, many techno-optimistic projects declare AI as the better half of humankind, evening out the fallibility of human bias and our inability to “know it all”². This potential takes the form of many very lofty projections, such as helping humanity understand

Sonja Rattay

itself on a deeper level, uncovering new perspectives and opening paths for global unification and general harmony and balance. Machine learning has enabled some truly innovative approaches to better address the complex demands of our current society. With the world being what it is—hyper-connected, racing towards climate catastrophe and with raging inequality—fast and powerful tools are required to respond to the rising challenges. AI has the potential to address major societal hurdles and the harm already done, which are so large that we might not be able to do better without it. For example, AI and data can help us identify discriminatory patterns that would otherwise be hard to communicate³. AI is also being used to track, analyse and speed up the removal of plastic waste in oceans⁴ and create new sustainable building materials and can thus be the solution we need to address the harms we have already done to the planet. AI can also potentially help us de-centre humanity and move beyond the Anthropocene by decoding languages and developments of natural ecosystems and other species, allowing us to communicate with animals and ecologies, such as smart forests. These projects and approaches claim that if we can rethink AI creatively, we can address it as a social practice rather than a purely technical or even design-making task that's radically re-politicised to address power imbalances and provide foresight for social needs⁵.

Discourses based on these narratives push a large part of ethical responsibility towards these technical solutions⁶: AI takes over all the difficult aspects that humanity is failing in, such as coordinating production circles that honour planetary health, long-term sustainable economic systems, global communication, interest negotiations between nations (or other social groupings such as tribes etc.) and representing nature as an equal party with rights and interests. AI can calculate the “true costs” of decisions and predict and estimate outcomes on a global and long-term scale. As a result, we can then utilise AI to make better deci-

¹ According to Jasanoff “Sociotechnical imaginaries occupy the theoretically undeveloped space between the idealistic collective imaginations identified by social and political theorists and the hybrid but politically neutered networks or assemblages with which STS scholars often describe reality.” in *Future Imperfect: Science, technology, and the imaginations of modernity in Dreamscapes of modernity* (2015).

² In *Justitia ex Machina: The Case for Automating Morals*, Berg Palm and Schwöbel illustrate a common conflation of the tool with the application, as well as the justification that tools can be fallible because humans are fallible. This approach negates the fact social structures re-embedded and echoed through tools, make it harder to break them apart.

AI and the Challenge of Speculative Ethics

³ D'ignazio and Klein describe in *Data feminism* (2020) how data influences power dynamics and hierarchies and how to work with data to challenge existing structures.

⁴ An overview of different projects using machine learning to discover new raw materials by Neil Savage: *Machines learn to unearth new materials* in *Nature* (2021)

⁵ One example are the efforts of Indigenous AI, a collective that takes a post-human approach to living with AI.

⁶ Aphra et al. draw from the sociology of expectations to outline and examine how “ethical AI” is being constructed in different cases, from commercial as well as governmental angles. They also look into the implications of the resulting discourse in *Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance*. in *Big Data & Society* (2020).

sions, optimise holistically and hold humans accountable for their actions.

While all these projections paint AI as a potential solution to human failings, the scenarios within which these AI agents are set are perceived by many as potentially dystopian. The various use cases of data-driven technologies have grown much faster and broader than our understanding of the interconnected consequences and implications of their enmeshment into our socio-technical environments. Worries about eliminating free will, individual choice and a general abandonment of human development are embedded in many critiques and discourses. Some scenarios predict a stronger divide in humanity between have and have nots, while others focus on a unified global society in which AI levels all needs and interests.

In these dystopian imaginaries, AI will make humanity either obsolete or turn humans into overly optimised cogs in a machine with an unclear purpose. AI and ML have drawn criticism, in particular, for the far-reaching consequences of short-sighted technical implementations. Cases of harmful outcomes on various scales have been discussed repeatedly in the media. For example, Compass, an algorithm used in the US legal system leading to racially biased sentencing, and The Facebook Files, one of the newest investigations on the extent to which ethical problems are known and tolerated by the social network. AI supports the concentration of power in the hands of the already powerful. The required means to build, train and utilise AI systems are limited to those with already massive economic and technical infrastructure in place. This reinforces the separation between the economically, socially and digitally privileged and consumers, who in turn double as data providers and hence building material for this new infrastructure. This infrastructure also harms the planetary ecosystem on a dramatic scale, from lithium mines to the construction of massive data centres⁷. Further, training these models emits huge amounts of carbon for

Sonja Rattay

small increases in model accuracy. These, and plenty of other examples, highlight that technological progress left unattended does not alone provide better solutions for all parts of society and negatively affects already vulnerable groups and individuals. Those suffering the consequences of this separation are the ones already affected most by the breaks and errors in infrastructure, targeted by data drawn from a racist, ableist, classist, misogynist world⁸. AI using training data based upon this neoliberal, violent society, then creates a future based upon the past, reinforcing the bureaucratic form of violence privileging scientific authority and solutionism, where quantitative correlations are praised regardless of substance or causality⁹.

These and similar cases have left the impression upon many that to unlock the potential of AI, we need to address the functional oversight that led to unforeseen and unintended harmful consequences. The general sense seems to be that AI as a technology can save us from dangers that humankind has caused through unsustainable resource management and production practices. To leverage this potential we have to “solve the problem of the ethics” to address the potential negative side effects of the dystopian speculations. While many of the discussed scenarios position AI as benevolent, it is also sketched out to always weigh the needs and interests of the individual against the needs and interests of a global society, including nature, the planet etc.—in short, engaging in the process of ethical decision making. This painting of AI acts as a solution to the fear of making wrong and/or flawed decisions. Here we encounter a structural dilemma in the engaged AI imaginaries—in order to successfully deploy AI to make the right ethical decisions, we need to solve the problem of ethics to avoid the undesirable non-ethical consequences.

As a result, there is increasing investment in designing ethical AI systems. With these conditions, it is questionable, however, whether any kind of debiasing or reforming performed by corporate or governmental actors can change

⁷ Kate Crawford works through the infrastructure and ecosystem necessary to produce what is perceived as AI on the consumer front in *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence* (2021). She traces the ecological impact of the required resources as well as the economic and social consequences of the far-reaching extractive practices deployed for the construction of smart agents and systems and the human labour required for making something material appear immaterial.

⁸ In *Weapons of Math Destruction* Cathy O’Neil works through how multiple closed loop ML systems enforce oppressive social structures on massive scales. Crown Publishing, 2016.

⁹ Dan McQuillann makes a great point about this in *Non-Fascist AI* (2020).

AI and the Challenge of Speculative Ethics

¹⁰ James Moor was one of the first to call out the ethical implication of computer technology in 1985. He called computers logically malleable devices and made the point that computers, more than any other technology before, have a strong influence on how society has to approach its moral structures in *What is computer ethics?* in *Metaphilosophy* (1985).

¹¹ In 2020, the research conference NeurIPS asked researchers to include a reflection on the broader impact of their research in their submissions. Abuhamad, G., & Rheault, C. surveyed the researchers and concluded that many researchers struggled with indicating why the technical work they are doing might have a broader impact on society outside of their own fields. “Like a researcher stating broader impact for the very first time.” (2020)

the systemic structures upholding the harms inflicted through current AI systems, especially since the notion of what constitutes ethical design of AI is fuzzy.

Traditionally ethics has most commonly been described in three ways: deontological ethics (duty-focused), consequentialist ethics such as utilitarianism, and virtue ethics, each presenting a different framework for how to assess the morality of a decision. Consequentialist viewpoints in particular, are broadly established in the discussions around ethical AI, manifested in risk assessments, simulation and assessments, with a deep reliance of much design and computing practices on traditional risk-based approaches originally shepherded by Moor’s work on computer ethics¹⁰. Duty or rule-oriented perspectives can be seen in frameworks that attempt to ensure certain functional safeguards, such as eliminating biases and discrimination from algorithms, on the basis that discrimination is perceived as morally wrong. Especially within the field of computer science, many describe the functional work as disconnected from the high-level concerns that are positioned in an overarching “logical social layer”, which needs to be figured out independently of the technical layer¹¹. In this view, ethics is perceived as a disconnected problem to solve, a step after building the functional aspects of a product. The connecting tissue between these two, which is actually the space in which ethically relevant decisions are made, is not registered. Ethics is instead positioned as a problematic instance that arises when problems with the current functionality are uncovered, which then must be solved in response. Ethics is perceived to be something separate from the actual development and production of “AI”, something that needs to be done on top to keep up with technological development. These tendencies highlight a disconnect between considering functional aspects and the relation to high-level worries. The previously listed imaginaries in themselves, however, already hold ethical considerations as well as normative commitments—what we deem potentially

Sonja Rattay

possible as well as probable is based on the imaginaries we have about both AI as a technology and as technology as a socially relevant force. “What do we value as a society?” is a question that comes up again and again throughout the public discourses regarding judgements towards what a potential global AI system should optimise for. What stands behind this question is much more the concern of “What should we value in an optimal society?” rather than what we as a society value at present. While these questions are not necessarily recognised as active ethical engagement, they have been at the basis of moral philosophy for centuries and have, as yet, not been answered in a successfully generalisable manner by any of the previously mentioned, rationally motivated philosophies.

More recently, relational perspectives to ethics have started to gain traction, most prominently feminist care-focused ethics. Such ethics of care¹², rather than duty or outcomes, recognise that static frameworks and guidelines struggle with contextual interpretations of ethical decision making. When practical everyday life comes in between good intentions and applied implementation, trade-offs and compromises can lead to scenarios in which the difference between utopia and dystopia is related more to perspective and individual interpretation rather than factual reality. This reality will most likely lie somewhere in between, in an uncomfortable grey zone of compromises, trade-offs and negotiations of values and desired futures. In these negotiations, moral values can be interpreted and actualised in many different ways. These grey zones need to be acknowledged as an embedded part of technical development processes rather than being seen as an inconvenience and as an important aspect of the ethical considerations that are entangled in the socio-technical fabric of our society. Most of all, engaging in these grey zones needs to be validated as an act of productive ethics-making, together with shared dreaming. While we need to critically examine the positions and values that we manifest through this

¹² One example is the work of Maria Puig de La Bellacasa, whose research investigates the crossings of science and technology studies, feminist theory and environmental studies and engages in a more-than-human approach to care ethics in *Matters of care in technoscience: Assembling neglected things* in *Social studies of science*.

AI and the Challenge of Speculative Ethics

technology, we also need hopeful visions and stories that motivate structural change and engage us in the intentional reconstruction of the futures we want to live in, with and through AI, in a caring manner.

SONJA RATTAY is a post-disciplinary designer and researcher at the intersection of design, ethics and AI. Her work focuses on investigation how ethics are constructed in everyday design practices for data driven technologies. As part of her PhD research, she investigates practices for alternative ethical frameworks in technology design. She has a background in strategic design and entrepreneurship and is part of the European research network DCODE, which aims to rethink design for new pathways in the future. DCODE has received funding from the European Union's Horizon 2020 research agreement and innovation programme and the Marie Skłodowska-Curie grand agreement No 955990.

