Chapter 8
Conclusion

Abstract The conclusion briefly summarises the main arguments of the book. It focuses on the requirements for mitigation options to be used to address the ethical and human rights concerns of artificial intelligence. It also provides a high-level overview of the main recommendations brought forth in the book. It thereby shows how conceptual and empirical insights into the nature of AI, the ethical issues thus raised and the mitigation strategies currently being discussed can be used to develop practically relevant conclusions. These conclusions and recommendations help to ensure that AI ecosystems are developed and shaped in ways that are conducive to human flourishing.

Keywords Requirements for AI ethics · Recommendations for AI · AI governance · Ethics and AI ecosystems · AI regulation

Technology is part of human life. Its development and use have the potential to raise ethical concerns and issues – and this will not change. Ethics, understood as our struggle to determine what is right and wrong and our reflection on how and why we make such a distinction, is not subject to resolution. While we may agree on what is right and wrong in many cases, this agreement is always partial, temporary and subject to revision. We may, however, be able to agree on some general and abstract principles. In this book I have suggested that human flourishing is such a principle. If we agree on that, then we can think through what the application of the principle to a technology such as AI can mean. This exercise can help us understand the specific issues that arise, why they arise and how we can evaluate them. It can also help us think through what we can do about them, and may even help us resolve some of them to universal satisfaction.

Several aspects that I have focused on in this book can, I hope, make a novel and interesting contribution to the AI and ethics debate. I started by looking at the concept of AI. "Artificial intelligence" is not an innocent and morally neutral term. It is emotive because it points to a characteristic of humans (and to some degree of other animals) while implying that this characteristic can be artificially replicated. This implication has consequences for how we as humans see ourselves and our role

B. C. Stahl, *Artificial Intelligence for a Better Future*,
SpringerBriefs in Research and Innovation Governance,
https://doi.org/10.1007/978-3-030-69978-9_8

in the world. Artificial intelligence is also often contrasted with human intelligence, implicitly suggesting or explicitly asserting that machines can or even should replace humans. Again, this touches deeply rooted views of what humans are.

In order to render the discussion more accessible, I have proposed a new categorisation of the AI debate. My suggestion is that we distinguish between three perspectives on AI: machine learning or narrow AI, general AI and converging socio-technical systems. These three perspectives on the technology are enlightening because they align with the categorisation of ethical issues on AI: first, ethical issues related to machine learning; second, general issues related to living in a digital world; and third, metaphysical issues posed by AI. These distinctions thus provide a better understanding and overview of AI ethics in a very busy and often overwhelming public and academic debate.

While these categorisations clarify the debate, they say very little about what could or should be done about the issues. One of the problems in this type of normative discussion is that it is unclear how recommendations or prescriptions can be justified. On what grounds could we say that technical applications should be developed, promoted, avoided or prohibited? Drawing on the idea of human flourishing allows a normative point of reference to be established that is consistent and compatible with the main ethical theories and can provide a framework for thinking about normative questions without presupposing substantive moral positions.

The idea of human flourishing has the added advantage of not requiring a strict distinction between ethics and law, both of which are normative constructs that could promote or inhibit flourishing. This is particularly important in light of the numerous existing legal and regulatory rules that already guide the development and use of technology, including AI.

Drawing on rich empirical work, I analysed ethical concerns and suggested interventions, mitigations and governance approaches to promote the benefits of AI and avoid or address its downsides.

One problem in the AI ethics discussion is its high level of complexity. Any attempt to match individual issues with stakeholders and mitigation options runs into several problems. First, the number of possible combinations of stakeholders, mitigations and ethical issues to be addressed is such that it is impractical to try to understand the field using such a straightforward approach. Second, and more important, the different components of the interaction are not independent, and an intervention in one part is likely to have consequences in another part. As this type of dynamic relationship lends itself to being described using a systems perspective, I have adopted the now widely used ecosystem metaphor and applied it to the AI discourse.

The question of what needs to be done to ensure that AI ecosystems are conducive to human flourishing was then tackled through the ecosystem metaphor. This led me to investigate, from an ethical perspective, the implications of using the ecosystem metaphor, a question that is not yet widely pursued in the AI field. In addition, I analysed the challenges that the ecosystem approach to AI and ethics raises and the requirements that any intervention would need to fulfil, and I concluded with suggestions to take the debate further and provide input into discussions.

The analysis pointed to three groups of requirements that interventions into AI ecosystems need to fulfil, in order to increase their chances of successfully promoting human flourishing:

- **Interventions need to clearly delineate the boundaries of the ecosystem**: Systems boundaries are not necessarily clear and obvious. In order to support AI ecosystems, the boundaries of the ecosystem in question need to be clearly located. This refers not only to geographical and jurisdictional boundaries, but also to conceptual ones, i.e. the question of which concept of AI is the target of intervention and which ethical and normative concepts are at the centre of attention.
- **Interventions need to develop, support, maintain and disseminate knowledge:** The members of AI ecosystems require knowledge, if they are to work together to identify ethically desirable future states and find ways of working towards those. AI as a set of advanced technologies requires extensive subject expertise in the technologies, their capacities and uses. In addition, AI ecosystems for human flourishing require knowledge about concepts and processes that support and underpin ethical reflections. And, finally, AI ecosystems need mechanisms that allow for these various bodies of knowledge to be updated and made available to members of those ecosystems who need them in a particular situation.
- **Interventions need to be adaptive, flexible and able to learn:** The fast-moving nature of AI-related innovation and technology development, but also of social structures and preferences as well as adjacent innovation ecosystems, means that any intervention into the AI ecosystem needs to incorporate the possibility and, indeed, likelihood of change. Governance structures therefore need to be flexible and adaptable. They need to be open to learning and revisions. They need to be cognisant of existing responsibilities and must build and shape these to develop the ecosystem in the direction of human flourishing.

These requirements are deduced from the nature and characteristics of AI innovation ecosystems. They are likely to have different weights in different circumstances and may need to be supplemented by additional requirements. They constitute the basis of the recommendations developed in this book. Before I return to these recommendations it is worth reflecting on future work.

The work described in this book calls for development in several directions. An immediate starting point is a better empirical understanding of the impact of AI and digital technologies across several fields and application areas. We need detailed understanding of the use of technologies in various domains and the consequences arising. We also need a much broader geographical coverage to ensure that the specifics of different nations, regions and cultures are properly understood.

Such empirical social science research should be integrated into the scientific and technical research and development activities in the AI field. We need a strong knowledge base to help stakeholders understand how particular technologies are used in different areas, which can help technical researchers and developers as well as users, deployers, policymakers and regulators.

The insights developed this way will need to be curated and made available to stakeholders in a suitable way. To a large extent this can be done through existing structures, notably the scientific publication process. However, issues of legislation, regulation and compliance require special gatekeepers who can lay claim not only to a high level of scientific and technical expertise, but also to normative legitimacy. The idea is not to install a tyranny of the regulator, but to establish ways that help stakeholders navigate the complexity of the debate and spaces in which organisations and societies can conduct a fruitful debate about desirable futures and the role that technologies should play in them.

The discussion of the ethics of AI remains high-profile. Numerous policy and regulatory proposals are likely to be implemented soon. The causes of the high level of attention that AI receives remain pertinent. The technologies that constitute AI continue to develop rapidly and are expected to have a significant social and economic impact. They promise immense benefits and simultaneously raise deep concerns. Striking an appropriate balance between benefits and risks calls for difficult decisions drawing on expertise in technical, legal, ethical, social, economic and other fields.

In this book I have made suggestions on how to think about these questions and how to navigate the complexity of the debate, and I have provided some suggestions on what should be done to facilitate this discussion. These recommendations have the purpose of moving AI ecosystems in the direction of human flourishing. They satisfy the three requirements listed above, namely to delineate the ecosystems boundaries, to establish and maintain the required knowledge base and to provide flexible and adaptive governance structures. In slightly more detail (see Chapter 7 for the full account), the recommendations are:

- **Conceptual clarification: Move beyond AI (7.3.1)**
  The concept of AI is complex and multi-faceted (see Chapter 2). The extent of the ecosystems concerned and the ethical and human rights issues that are relevant in them depend to a large degree on the meaning of the term “artificial intelligence”. Any practical intervention should therefore be clear on the meaning of the concept. It will often be appropriate to use a more specific term, such as “machine learning” or “neural network”, where the issues are related to the characteristics of the technology. It may also be appropriate to use a wider term such as “emerging digital technologies”, where broad societal implications are of interest.
- **Excellence and flourishing: Recognise their interdependence (7.3.2)**
  In the current discussion of AI, including some of the policy-oriented discourses, there is a tendency to distinguish between the technical side of AI, in which scientific and technical expertise is a priority, and the ethical and human rights side. This blurs the boundaries of what is or should be of relevance in an AI ecosystem. The recommendation points to the fact that scientific and technical excellence must explicitly include social and ethical aspects. Work on AI systems that ignores social and ethical consequences cannot be considered excellent.

- **Measurements of flourishing: Understanding expected impacts (7.3.3)**
  In order to react appropriately to the development, deployment and use of AI, we must be able to understand the impact they can be expected to have. It is therefore important to build a knowledge base that allows us to measure (not necessarily using quantitative metrics) the impact across the range of AI technologies and application areas. While it is unlikely to be possible to comprehensively measure all possible ethical, social and human rights impacts, there are families of measurements of aspects of human flourishing that can be applied to AI, and these need to be developed and promoted.
- **AI benefits, risks and capabilities: Communication, knowledge and capacity building (7.3.4)**
  The knowledge base of AI ecosystems needs to cover the technical side of AI technologies, to ensure that the risks and potential benefits of these technologies can be clearly understood. This knowledge, combined with the measures of human flourishing in the preceding recommendation, is required for a measured view of the impact of AI systems and a measured evaluation of their benefits and downsides. This knowledge base that AI ecosystems must be able to draw on, in order to make justifiable decisions on AI, is dynamic and can be expected to evolve quickly. It therefore needs to develop mechanisms for the regular updating and development of expertise and means of disseminating it to those who need it.
- **Stakeholder engagement: Understanding societal preferences (7.3.5)**
  The broad and all-encompassing nature of AI and its possible impacts means that decisions shaping the development, deployment and use of AI and hence its societal impact must be subject to public debate. Established mechanisms of representative democracy have an important role to play in guiding AI governance. However, the dynamic and complex nature of the field means that additional mechanisms for understanding the views and perceptions of stakeholders should be employed. Involving stakeholders in meaningful two-way communication with researchers, scientists and industry has the advantage of increasing the knowledge base that technical experts can draw on, as well as improving the legitimacy of decisions and policies resulting from such stakeholder engagements.
- **Responsibility for regulation and enforcement: Defining the central node(s) of AI ecosystems (7.3.6)**
  AI ecosystems do not develop in a vacuum but emerge from existing technical, social, legal and political ecosystems. These ecosystems have developed a plethora of mechanisms to attribute responsibility with a view to ensuring that the risks and benefits of emerging technologies are ascribed appropriately. The emergence of AI ecosystems within these existing environments means that existing roles and responsibilities need to be suitably modified and developed. This calls for a way of coordinating the transition to AI ecosystems and integrating them into established contexts. The shifting networks of responsibilities that govern emerging technologies will therefore need to evolve ways of developing formal and informal governance structures and monitoring their implementation. This calls for the establishment of central nodes (e.g. regulators, agencies, centres of excellence)

that link, guide and oversee AI ecosystems, and relevant knowledge and structures to ensure the technologies contribute to human flourishing.

I hope that this book and the recommendations that arise from it help strengthen the debate on AI and ethics. The book aims to support the appropriate shaping of AI ecosystems. In addition, its message should reach beyond the current focus on AI and help to develop our thinking on the technologies that will succeed AI at the centre of public attention.

Humans are and will remain tool-using animals. The importance of technical tools will increase, if anything, in times of ubiquitous, pervasive, wearable and implantable technologies. While novel technologies can affect our capabilities and our view of ourselves as individuals and as a species, I believe that some aspects of humanity will remain constant. Chief among them is the certainty that we will remain social beings, conscious of the possibility and reality of suffering, but also endowed with plans and hopes for a good life. We strive for happiness and seek to flourish in the knowledge that we will always be negotiating the question: how exactly can flourishing best be achieved? Technology can promote as well as reduce our flourishing. Our task is therefore to ask how novel technologies can affect flourishing and what we can do individually and collectively to steer such technologies in directions that support flourishing. I hope that this book will help us make positive use of AI and move towards a good, technology-enabled world.