

Chapter 12

Ethics in AI and Robotics: A Strategic Challenge



The recent success of AI and robotics has massively increased the international awareness of and the interest in these topics as a factor of economic competitiveness. This concerns not only businesses, but also regional and national governments. There are many claims about the potential for AI to create new service and product innovation. Such claims include the benefits of AI for healthcare such as improved diagnosis or therapy; for transport due to improved efficiency; for energy based on more accurate predictions of energy consumption; or ease of computer use with more natural user interfaces as, for example, in the case of speech understanding, gesture and face recognition and automatic translation.

In general, many smart or intelligent technologies have been considered a major driver of innovation (Lee and Trimi 2018; Makridakis 2017), and also an important source of knowledge for innovation (Fischer and Fröhlich 2013). As a result of these promises, there is today a plethora of regional, national, and even supranational strategies and policy papers that aim at maximising the benefits of AI for their citizens. Examples of national strategies range from Canada to Mexico, Japan or India. Regional strategies have been developed in Bavaria and in the Northern Baltic countries. Supranational AI strategies, or at least joint studies and papers are the subject of work in the OECD and the United Nations International Telecommunication Union (ITU).

In early 2019, a broad range of policy papers (Agrawal et al. 2019) and marketing studies from consulting companies have been published. Many of these make the case for the innovation potential and economic benefits of AI (Li et al. 2017; Seifert et al. 2018). Governments around the world have responded to the massive increase in AI applications, but also to an even greater number of predictions of future AI applications and their societal benefits. As early as 2017, the Canadian government published a pan-Canadian AI strategy. It was followed by developed countries traditionally interested in the creating information technology such as Japan, Singapore,

Finland and China. By mid-2018 the European Commission published its Communication on Artificial Intelligence, thus effectively motivating its member states to draft strategies for AI. In December 2018, the EU presented its plan for AI with more concrete actions from research to development, investments in AI, ensuring training and education and a proper computing infrastructure.

AI strategies around the world mostly follow a general model that addresses the **actors** in the AI and robotics environment, such as professionals, society, research organisations, companies and government. These groups require support through **infrastructure**, data, staff, finance, and information to productively create an environment conducive to the deployment of AI solutions. The aim is to create AI **solutions**, practices and improved benchmarking within industry. To support this, government strategies focus on a range of **processes** ranging from cooperation between these actors, on improving research excellence, staff training, and regulation. In addition, many strategies emphasise the need for an elaborate societal discourse and an ethical approach to AI.

It is not only the promises of increased competitiveness or new applications, which are driving the development of national strategies for AI and robotics. On the contrary, it seems that the public discussion in many countries is more focused on the potential damages that AI may induce, for example in the labour market, but also regarding human autonomy, privacy, and even the very future of society. There are two main streams that present one or another version of AI—and often robotics—dystopia. From an economic perspective, authors like Ford (2015) studied the potential impact of AI and robotics on (human) work. He predicts massive job losses in many sectors that have seemed to be immune to automation for a long time. The argument here is that new AI technology is now capable of replacing much more of the work for which human intelligence was required to date. This includes, for example, medical diagnostic knowledge, expert knowledge from tax advisory or the legal knowledge of lawyers. The second line of dystopian publications stems from more journalistic accounts of the potential future of AI. In several cases, these publications draw an image of our future in which AI overlords threaten humanity while others are more careful predictions about losing our privacy and autonomy (Bartlett 2018).

These often rather pessimistic predictions about the impact of AI have been quite successful in terms of their influence on the broader public. It is therefore unsurprising that policy makers around the world include the potential damage created by AI and robotics in their considerations and discussions. For example, the German AI strategy explicitly calls for a broad discussion of AI's societal impacts. In addition, it aims to support a continued discussion between politics, science, industry, and society. Several national policies emphasise the need to continuously monitor and study the impact of AI technology on labour and society. For example, the French AI strategy proposes to study the labour market impacts and to implement ethics-by-design. Thus, the central topics of such societal dialogues include questions of the potential or real impact of AI and robotics on the work force and on society as a whole, and also questions of privacy, security, safety and adequate regulation.

12.1 The Role of Ethics

The question of what should be considered right and wrong in the development and deployment of AI and robotics is central to many published policy papers. For example, the European Commission (EC) now asks for the inclusion of *ethics in the development and use of new technologies* in programmes and courses and the development of ethical guidelines for the use and development of AI *in full respect of fundamental rights*. The EC plan goes as far as aiming to set a global ethical standard towards becoming a world leader in ethical, trusted AI. Whatever we may think of this aspiration, it is certainly true that there is a lack of practical, agreed guidelines and rules regarding systems that are much more autonomous in their calculations, actions and reactions than what we have been used to in the past. Such autonomous systems—or more specifically intelligent autonomous systems—act; they *do* things. Now, *what one should do* is the famous Kantian question underlying all ethical considerations.

Today, most jurisdictions around the world have only just started to investigate regulatory aspects of AI. In this book we have given preliminary answers to many of these issues, however industry has made the case that it requires clear rules for speedy innovation based on AI. Companies may steer away from AI applications in states of uncertainty in which the legal implications of bringing AI and robotic applications to the market are unclear.

Ethics therefore becomes important at many layers of the policy discussion. It is a topic for the engineer designing the system, including the student learning to build AI systems. It is also a topic for society to value the impacts of AI technology on the daily lives of citizens. Consequently, it is a key question for policy makers in discussions about AI and robotic technologies. Note that ethical aspects are not only discussed in questions of regulation. Much more importantly, ethical questions underpin the design of AI and robotic systems from defining the application to the details of their implementation. Ethics in AI is therefore much broader and concerns very basic design choices and considerations about which society we would like to live in.

12.2 International Cooperation

In its AI Action Plan, the EC identifies a need for coordinated action in ethics and in addressing societal challenges, but also the regulatory framework. It calls upon its member states to create synergies and cooperation on ethics. At the time of writing this book, countries around the world are looking for best practices in regulating—or deregulating—AI and robotics. Information and communication technologies in general have a tendency to generate impact across country borders. Already today, AI systems such as Google's translation service, for example, provide good quality translations to people all over the world. These translations are based on documents

in many different languages available throughout the internet. In this way, Google exploits data that users publish only to create and improve its services often without people being aware that they support the development of improved translation services.

Many AI systems rely on massive amounts of data. And in many cases, this data may be considered personal. Questions of personal data protection have definitely become international ever since the EU defined its General Data Protection regulation (GDPR) to apply internationally, if not before. Privacy, data exchange, and AI are intimately related aspects that need to be put into an international context. There is a real international need to exchange concepts and ideas about how to best regulate or support AI and robotics. Countries will often take up regulatory models from other countries as they consider useful for their respective jurisdiction and the area of AI is no exception. Europe's GDPR has influenced policy makers world-wide, for example in the state of California. On the other hand, some countries may also decidedly vote against its underlying rationale and seek different legal approaches to data protection and privacy.

Similarly, the impact of AI on labour laws includes important international aspects. It is therefore not a coincidence that the New Zealand AI strategy, which remarkably is an industry association's paper, calls for more engagement with the international labour policy community. The New Zealand AI strategy also includes the challenging topic of AI for warfare. It is evident that this is a specific topic that needs to be discussed and perhaps regulated internationally.

Finally, international aspects do not stop with questions of regulation, labour or how to best design an AI system. Both the German and the EU strategies for AI include the important question which role AI should play in development policies. It is an yet open issue how to make sure that AI does not become a technology exclusively developed by a small set of industrialised leaders and instead also benefits developing countries.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

