MERCATOR
SCIENCE-POLICY
FELLOWSHIP-
PROGRAMME

WHITHER ARTIFICIAL INTELLIGENCE? DEBATING THE POLICY CHALLENGES OF THE UPCOMING TRANSFORMATION

BIRGITTA WOLFF (ED.)

SCIENCE POLICY
PAPER 2018
03

IMPRINT

Mercator Science-Policy Fellowship-Programme
Editor of the series: Tome Sandevski
Production editor: Andrea Wolf-Dieckmann

Research Strategy Department
Goethe University Frankfurt am Main
Theodor-W.-Adorno-Platz 1
D—60323 Frankfurt am Main

Email: science-policy@uni-frankfurt.de
Phone: +49 69 798 12132

ISSN: 2626-9597
urn:nbn:de:hebis:30:3-478510
Science Policy Paper 3 (2018)

Citation: Wolff, Birgitta (ed.)(2018). Whither Artificial Intelligence? Debating the Policy Challenges of the Upcoming Transformation. Science Policy Paper No. 3. Frankfurt am Main: Mercator Science-Policy Fellowship-Programme.

STIFTUNG
MERCATOR

Rhein-Main
Universitäten
Eine strategische Allianz

The Science Policy Paper Series of the Mercator Science Policy Fellowship Programme features concise contributions by academics of Goethe-University Frankfurt, Johannes Gutenberg-University Mainz and Technische Universität Darmstadt as well as senior policy professionals on current issues. Since economic, social and political challenges of our times are complex the paper series includes articles from various academic and policy-oriented perspectives.

This paper is the result of a public debate on the future of Artificial Intelligence organised by the Mercator Science-Policy Fellowship-Programme and the Centre for Science and Policy at the University of Cambridge at the Representation of the State of Hessen to the European Union in Brussels.

SCIENCE POLICY PAPER 3 (2018)

TABLE OF CONTENTS

PREFACE: WHITHER ARTIFICIAL INTELLIGENCE? DEBATING THE POLICY CHALLENGES OF THE UPCOMING TRANSFORMATION

BIRGITTA WOLFF
GOETHE UNIVERSITY FRANKFURT

In April 2018 the European Commission announced its holistic approach to Artificial Intelligence (AI) based on the following three pillars: first, to boost financial support and encourage uptake by the public and private sectors in order to reach investments in AI-related research and innovation by at least 20 billion Euros by the end of 2020. The second pillar aims at preparing for socio-economic changes in terms of the upcoming transformation of the labour market. Finally, the European Commission will ensure an appropriate ethical and legal framework by developing AI ethics guidelines and providing guidance on the interpretation of the Product Liability directive.

Issues of research and training feature prominently in the EU strategy. As providers of both research and training, universities already play a key role regarding the advancement of AI. Rapid developments in AI challenge both scientists and policy professionals. The opportunities and challenges of AI are multi-faceted. They exceed the boundaries of academic disciplines and the policy portfolios of ministries or Directorates General. Therefore, this paper constitutes a modest attempt to provide a concise overview of the policy challenges related to AI.

In order to foster the dialogue between policy professionals and academics Goethe-University Frankfurt launched the Mercator Science PolicyFellowship Programme in collaboration with Johannes Gutenberg University Mainz and Technische Universität Darmstadt in 2016. Since then, senior professionals from the public sector in Germany, the European Union, media and non-profit organisations have participated in our programme. The Mercator Science Policy Fellowship Programme and the Centre for Science and Policy at Cambridge University organised a joint event on AI in Brussels in July 2018. We decided to compile this publication as a follow-up to our Brussels event.

I would like to thank the Representation of the State of Hessen to the European Union for hosting our public debate. I gratefully acknowledge the funding by the Stiftung Mercator, which enabled us to establish our fellowship programme. I am indebted to Dr. Robert Doubleday and the Centre for Science and Policy at the University of Cambridge for providing invaluable advice for setting-up our fellowship scheme. And special thanks to Tome Sandevski and Andrea Wolf-Dieckmann for putting tremendous effort into preparing this volume.

THE FUTURE OF ARTIFICIAL INTELLIGENCE: POLICY RESEARCH PERSPECTIVES

DEMIAN FRANK AND PETRA AHRWEILER
JOHANNES GUTENBERG UNIVERSITY MAINZ

1. Expectations of how Artificial Intelligence (AI) will influence policy research

Policy research deals with the policy cycle encompassing problem definition, and policy development, implementation, enforcement and evaluation for different policy domains. On the one hand, this research field empirically describes and analyses how the phases of the policy cycle are processed by relevant social actors in interaction with industry, the media and civil society. On the other hand, policy research is concerned with the issue itself, i.e. the reasons for success and failure of policies that have or will run through the process. Here, the research field offers scientific policy advice including exante evaluation and assessments of potential futures, options, developments, and scenarios for policy domains to inform political debates and decisions. It is this latter function AI already has and will further influence policy research.

There is a tension in all phases of the policy cycle outlining the fundamental challenge in this area: political planning and steering for desirable futures is torn between the requirements for planning and the uncertainty of its effects. Usually, the implementation of policy measures for realising planning objectives is a risky enterprise: in complex policy domains, non-linearity exists between suggested interventions and a desired effect. Nevertheless, political planning means to consider and organise a chain of activities and measures to realise a specific objective. Drafting a plan implies scenario analysis and evaluation, i.e. producing knowledge about the future concerning the most likely developments. This in turn allows for preparation to face potential challenges to take place. Of course, such knowledge has obvious benefits, but crucially depends on finding reliable answers to “what-if” questions for evaluating different scenarios before implementing policy measures in empirical reality (Ahrweiler, 2017).

Computational models are increasingly applied by policy research to assist in developing, implementing and evaluating policies (Gilbert et al., 2018). As cause and effect are getting harder to pinpoint with increasing systemic complexity, AI is frequently used to unravel structures of interdependency and thereby improve the advice capacity of policy research. For example, systems mapping approaches such as Fuzzy Cognitive Mapping build qualitative models with different structures and assumptions to represent the situation with and without a policy intervention, thus evaluating different outcomes. Alternatively, techniques such as Logic Mapping are used for ex-post evaluation of policies. These test how a certain policy might have affected an outcome of interest, which supports common theory-based approaches to policy evaluation such as the Theory of Change.

2. AI-induced changes in society and politics

However, it is not only policy research that is supported by AI. These applications have, in fact, already directly seeped into policy practice, and are thereby changing the ways in which policies are processed (Milano et al., 2014; Desouza, 2018). With this change in policy practice, effectiveness, success rates, and applicability of policies are improved, which means that the targets of policy – the policy domains, which principally encompass all parts of our society – are changing too. Through AI-supported political change, this technology enacts AI-supported social change.

In recent years, a range of projects were encouraged that promoted interconnectedness between national and supranational agencies (EU eGovernment Action Plan 2016-2020): the ongoing digitalisation of public administration, advancing eGovernance and the evolution of Smart Cities are all relevant bases for the impact of AI on policy making. One major consequence of increasingly interconnected systems is that government agencies find themselves acting in highly complex social systems. In order to act and react effectively and efficiently within these networks, actors are required to understand the complexity and implied characteristics of a given system. Understanding and managing (within) complex social systems becomes a key skill for any actor involved, as unintended side effects are ever-present.

In policy practice, AI already helps policy makers find the right policy measures to act on a certain situation, for example, by evaluating complex interrelations of major sectors in national economies or by helping candidates in their election campaign via big data analysis. The increase in computing power and access to big data lets other scenarios become conceivable reality. By means of interconnected databases and sensor-supported Smart Cities, decision makers could be able to make ad-hoc decisions informed by simulation models based on real-time data. Although the realisation depends on several factors (e.g. CPU efficiency and legal frameworks), research is already engaged in theoretical approaches to simulation models based on citizen big data for population dynamics analysis.

As AI helps humans keep complex social systems manageable, one of the most important questions in years to come will be to decide if AI should remain a tool for policy making or if it should become a decision-making authority by itself. If AI can help to run a successful campaign for candidates via big data analysis, it may someday also articulate a political agenda of its own. While this thought might provoke scepticism, and rightfully so, AI has already followed through with similar developments, graduating from assistant to decision maker in other contexts of its application.

The current example par excellence where the hypothesis of AI-supported policies enacting AI-supported social change can most strongly be illustrated is the Chinese Social Credit System: As a demographically growing and economically expanding nation, China feels challenged to maintain social cohesion and turns to AI for controllability. With an AI-System monitoring its citizens' behaviour and opinions, the government is establishing a social meritocracy aligned with its own perspective of good citizenship. It is not unlikely that hopes of rationalisation and efficiency gains might prompt policy makers to implement such systems elsewhere.

3. Policy responses required

AI is confronting democracy with the imponderables of the digitalised 21st century. The capabilities of AI will transform policy making on a large scale. It is time to act; to choose between the desire for more state control and the chances of increased participation. Especially in the context of government action, AI has the potential to support or harm democracy and society's cultural values.

On the supportive side, AI has tremendous potential to organise and optimise the distribution of public goods, thereby supporting public welfare. Policies need to invest in the technical management of these social development contexts; for example, more publicly funded research programmes need to address the capabilities of AI in multiple stakeholder contexts. Furthermore, AI empowers people to access more information, more data, and more knowledge sources than ever before. It also enables people to interact and communicate with more people than ever before. Policies need to recognise and react to the resulting potential for opinion formation and increased participation of citizens in the democratic process.

On the harmful side, AI can be used for anti-democratic top-down control, or for realising manipulative or hidden agendas of lobbyists: Opinion formation, participation in social movements and networking increasingly take place in digital environments that are owned by private companies. State influence on these social mechanisms has diminished considerably. This environment is highly vulnerable to manipulation and influencing by AI programmes. Policies need to be established that encourage and implement a continuous ethical debate about core societal values and how to nurture them in a digitalised world.

References

Ahrweiler, Petra (2017). Agent-based Simulation for Science, Technology, and Innovation Policy. Scientometrics, Vol. 110 (1), pp. 391-415.

Desouza, Kevin C. (2018). Delivering Artificial Intelligence in Government: Challenges and Opportunities. Washington D.C.: IBM Center Report.

Gilbert, Nigel; Ahrweiler, Petra; Barbrook-Johnson, Pete; Narasimhan, Kavin; Wilkinson, Helen. (2018). Computational Modelling of Public Policy: Reflections on Practice. Journal of Artificial Societies and Social Simulation (JASSS), Vol. 21 (1) 14.

Milano, Michela; O'Sullivan, Barry; Gavanelli, Marco (2014). Sustainable Policy Making: A Strategic Challenge for Artificial Intelligence. AI Magazine, Vol. 35 (3), pp. 22-35.

ARTIFICIAL INTELLIGENCE IN THE WORKPLACE

ROLF VAN DICK
GOETHE UNIVERSITY FRANKFURT

According to a survey by the Institute for Management and Economic Research (manager seminars, September 2018), 41% or almost half of those respondents over 60, considered it unlikely that they would be affected by Artificial Intelligence (AI) in the workplace. On the other hand, younger respondents more realistically estimated that significant AI-related changes would occur in their workplace within the next five years, not only in production and data analysis, but also in customer service and office practices across the board.

For many years, the conviction was that a computer would never completely replace a human being due to its lack of intelligence. It was thought impossible that it could, for example, beat a human being at a game of chess. Then, in 1997, IBM computer Deep Blue won a game against incumbent grandmaster Garry Kasparov. Since that time, man and machine have come together to form teams, made up of one human player and one chess programme each, that are currently known as the best chess players in the world. Nowadays, Deep Blue's successor, Watson, can provide medical diagnoses with astonishing accuracy and will likely replace doctors in diagnostics in the not too distant future (Kelly, 2016).

AI has also long been used in staff recruitment practices. DeepSense, a company in California, can determine an applicant's personality based on an analysis of their Facebook profile, and then sends the information on to the recruiter, whether or not the applicant is aware of this. The US company HireVue holds purely digital interviews with applicants. Applicants' voices, facial expressions and choices of words are analysed and compared with specific success parameters. If similarities with those parameters are found, the AI filters out these candidates as those most likely to be a good fit for the recruiter.

The rate of success of such services may not yet be as high as desired, as the algorithm quality is significantly impacted by the choice of psychological test used to profile a personality. If these tests are of poor quality, even the best algorithm cannot deliver reliable results. Nevertheless, we need to resign ourselves to the fact that companies will increasingly rely on such profiling processes and that these will improve over time. At the very least, they will be used, and in fact already are being used, to sift through large volumes of applicants. This has advantages and disadvantages. Beside considerable time-savings, a company will naturally benefit from being able to identify applicants with a good fit, for example based on their personality. Co-workers also benefit, because a good fit between an employee and their work is crucial for high job satisfaction rates and long-term good health. However, such processes also have their drawbacks. They can contribute to homogenisation of the workforce within a company, as candidates who fail the algorithmic test have no chance of proceeding to the interview stage. Those candidates thus never encounter a human being in a conversation via which their hidden potential could be recognised by chance.

One of the primary responsibilities of policy making is to safe-guard the privacy of any prospective employee. To date, a conventional personality test can only be applied, and the results analysed with the participant's consent. Such tests are also subject to quality control and licensing: the German standard DIN 33430 delineates who can apply what aptitude tests, and for what reasons. The ability of AI to outperform a human being in terms of speediness is no reason to let this standard slip for algorithmic tests.

The changes that AI is effecting in the workplace are not only obvious in the recruitment field. The relationship between an organisation and its employees is also changing. This is exhibited by, for example, platform businesses, a prominent one being Uber. On the one hand, Uber can provide a faster and significantly cheaper taxi service to its customers. On the other hand, Uber's drivers are mis-classified as self-employed, and therefore lack adequate unemployment and other such insurance that comes as standard with a regular employment contract. AI will continue to develop at break-neck speed in this field and offer a wide variety of services. Policy makers must not hamper these developments, yet they must ensure that work standards, relating to, for example, job security and minimum wage requirements, are met.

Replacing the human being in a variety of other fields will be the next step in AI-based automation of the workforce. So far Uber has increased competition for conventional taxi businesses, but look to the future and self-driving cars, perhaps even self-flying taxis as being tested in Dubai and Singapore (Hein, 2017) will relieve the need for a human driver (or pilot). Such developments will rear their head in other spheres of work, too. In the retail banking sector, customer service agents in bank branches have long started to be replaced by online banking. Other retail banking functions, such as mortgage and loan advisers, will increasingly be replaced by algorithms. Soon, we will find such developments affecting jobs that we cannot yet conceive of being carried out by machines. Lawyers' and notaries' simpler tasks are already being automated. Automation is also happening in schools and in university-level education, where the introduction of digital teaching will erase traditional jobs. An experiment has shown that students being taught in a digital environment could not distinguish between the feedback received from a human teacher and that received from an AI-based system. Not only did the 50% of students who were "supervised" by the AI system not notice any difference, they even judged their teacher to be friendlier and more trustworthy (Leopold, 2017).

AI-based automation will have a similar impact on the number and kinds of jobs in the service sector as the introduction of the steam engine and the conveyor belt had on the manufacturing industry and on agriculture. Considering current demographic developments, this new kind of automation might also offer some opportunities. Assuming no noteworthy increase in immigration and a continuation of current demographic trends, the German labour pool will decrease by 40% from around 44 million to 26 million employees (Institut für Arbeitsmarkt- und Berufsforschung, 2011: 2). Thus, increased automation and the ongoing development of AI provide important opportunities. The future will see us generating higher revenues with fewer employees. This same scenario has played out in agriculture and in major areas of the manufacturing industry in the last 100 years. There will be fewer jobs for those less qualified, however we will also see entirely new occupations emerge.

Policy makers must therefore be proactive in investing in and therefore paving the way to flexible and life-long education, rather than having education end with a middle school or university graduation certificate. Breathing spaces will need to be created by, for example, expanding paid educational leave, and financial resources will need to be made available to individuals, as well as to universities and other educational establishments.

In areas other than the workplace, the question of surveillance needs to be prioritised and addressed by policy makers, once a necessary broad-ranging public debate has taken place. Amazon and co., for example, know when we look at different kinds of web content and how long for. Technology and the processing of masses of data, practically turns people into an open book for entities collecting and analysing the collected data. In fact, this is not only the case when surfing the web for online shopping or otherwise, but also applies to people as employees. Finally, policy makers need to develop new corporate and employment taxation models. Why not demand national insurance contributions to be made for machines that are replacing jobs formerly carried out by human beings? Furthermore, where people are mis-classified as self-employed while working for a platform business, whether the world is just or not will depend on how those who profit from the newly-gained flexibility are obliged to "give back" to society for the common good (Lenzen, 2018). This is no longer relevant to taxi drivers only, but also applies to programmers, lawyers, teachers and other professions.

References:

Hein, Christopher (2017). Singapur und Dubai testen fliegende Taxis. Available at: http://www.faz.net/aktuell/gesellschaft/singapur-und-dubai-testen-fliegende-taxis-14977992.html [Accessed 26.11.2018].

Institut für Arbeitsmarkt- und Berufsforschung (German Institute for Labour Market and Occupational Research). IAB-Kurzbericht 16/2011. Nürnberg: IAB. Available at: http://doku.iab.de/kurzber/2011/kb1611.pdf [Accessed 26.11.2018].

Kelly, Kevin (2016). The inevitable. Understanding 12 technological forces that will shape our future. New York: Penguin.

Lenzen, Manuela (2018). Künstliche Intelligenz. Munich: C.H. Beck.

Leopold, Todd (2017). A professor built an AI teaching assistant for his courses — and it could shape the future of education. Available at: https://www.businessinsider.com/a-professor-built-an-ai-teaching-assistant-for-his-courses-and-it-could-shape-the-future-of-education-2017-3?IR=T [Accessed 26.11.2018].

ARTIFICIAL INTELLIGENCE AND LAW

CHRISTOPH BURCHARD
GOETHE UNIVERSITY FRANKFURT

Artificial intelligence (AI)[1], together with big data, is the driving force behind the ever-accelerating digital revolution. AI has what it takes to call into question our fundamental concepts and processes of political, social, economic etc. order (Macron, 2018; Zuboff, 2018), and the law will not be spared. Therefore, all societal actors (inter alia from politics, the economy, legal practice and academia) must take responsibility for the crucial twin tasks of determining the right, balanced relationship between AI and the law, and even to hybridise them.

In a nutshell, "AI and Law" thus manifests a relationship of interdependence and mutual penetration. The following three examples illustrate this:

- The use of AI (e.g. in self-driving cars, as “members” of corporate boards, or in the context of bank lending decisions) leads to seemingly classic legal questions (for example, in regard to tort liability: damages may occur because the underlying algorithm has been incorrectly programmed, the output of a self-learning AI process was unpredictable for human beings, or the system was set up in a non-transparent manner). Here, the law is binding upon AI. This highlights the general regulatory dilemma of “law and technology”: legal (e.g. tort law) requirements must not hamper AI innovations, or even make them impossible. However, AI innovations must not compromise legally protected goods and interests (e.g. bodily integrity).
- AI applications often strive for and enable "legal compliance by design". As such, legal compliance is integrated into the source code. Despite the inherent challenges, the law – or rather, the legal community – should accept this “invitation” to make its knowledge of the normative content of and background to legal decision making IT-compatible (Herberger, 2018: p. 2828). Once this is accomplished, the law will no longer emerge from social (interpersonal) practices, but rather as an IT component that determines social interactions from their outset. In extremo, the law as we know it is even likely to become partially superfluous: for instance, traditional traffic criminal law will no longer be

[1] The term AI is enigmatic at best. I am not referring to AI in the sense of artificial general intelligence or strong AI, which covers all human cognitive abilities. I am rather referring to automated and self-taught decision making and classification processes that usually use big data and that are also increasingly being used outside of the usual clearly defined context of so-called weak AI or machine learning.

necessary if road traffic control becomes automated and controlled by AI (Schwintowski, 2018, p. 1608). Note, however, that this implies a change of focus, where criminal law will, for example, focus on the automation process, or on attacks against it. These processes of hybridization, replacement and refocusing must also be incorporated into legal research and education, which in turn calls for support from the relevant political actors.

- AI facilitates the development and use of so-called legal tech, which supports legal work processes, or prepares or even fully automates legal decision making. These practices have traditionally been reserved for human beings. This applies to both the private sector (e.g. if due diligence checks are no longer carried out by lawyers but instead by "machines") as well as to decision making in the public sector (e.g. once bail or probation decisions are automated, or once divorces are carried out using "online" tools). In this way, AI will significantly change job descriptions in the legal profession. This again must be mirrored in legal education (e.g. first studies suggest that thousands and thousands of lawyers will be "automated" in the years to come). Further, when public officials turn to AI and legal tech, this will raise pressing concerns about democratic legitimation and control. (This, for example, will hold true once the police resorts to predictive policing applications that are programmed and developed by private companies, which neither disclose the underlying algorithms, since they are treated as corporate secrets, nor account for the data used to train these algorithms.)

These examples, though few, suggest that the potential of "AI and Law" to mutually transform and merge with one another is significant. However, it is too early to make the call on precisely how AI will transform our existing general political, economic, societal etc. order – and especially our legal order. These days, AI fuels many hopes and fears, which are at times exaggerated. Indeed, the (in many cases inevitable) superficial grasp of complex topics such as AI, the law, and "AI and Law" leads to overstatements, simplifications and distortions. European policy making in the area of "AI and Law" must therefore be rational and cool-headed in order to comprehensively assess its opportunities and hazards from a European and a geostrategic vantage point.

Such an assessment of "AI and Law", which by its very nature calls for interdisciplinary efforts, must not lose sight of two challenges (also cf. Burchard, 2018):

- Firstly, AI is not unbiased – despite the idealisation that associates AI with objectivity and rationality. Rather, AI is – whether consciously or unconsciously – normatively charged. It can thus perpetuate (e.g. political, economic or social) asymmetries, strengthen existing discrimination and make the quantification of the social sphere appear unavoidable. This is worrisome, for example, when AI systems are trained with existing data so that their biases and prejudices etc. are thus "bred into" AI applications.
- Secondly, it is unclear where the political, economic, and social transformations enabled by "AI and Law" will lead us. The use of AI can lead to emancipation and liberation. However, it can also be used to foster and strengthen authoritarianism and populism, to perpetuate (economic, social, etc.) asymmetries, or to concentrate power within private companies. Policy makers who are dealing with "AI and Law" must therefore be committed to fundamental European values. Though there is a wide scope of policy design options, policy makers are called upon to use their decision-making power in a way that lives up to the expectations of democratic legitimacy and the protection of human and fundamental rights etc. If, for example, one wanted to have employment agencies algorithmically rate the employment prospects of unemployed people based on current data which systemically discriminates against elders and women, this would amount to a conscious political decision to "Keep it up!". This illustrates that "AI and Law" does not render politics mute; to the contrary, it calls for good and sensible policy making.

References

Burchard, Christoph (2018). Die normative Ordnung Künstlicher Intelligenz | NO:KI. PI-Project at the Cluster of Excellence "The Formation of Normative Orders", Frankfurt am Main, Germany. On file with the author.

Herberger, Maximilian (2018). „Künstliche Intelligenz" und Recht. Neue Juristische Wochenschrift 2018 (39), pp. 2825-2829.

Macron, Emmanuel (2018). Emmanuel Macron Talks to WIRED About France's AI Strategy. Available at: https://www.wired.com/story/emmanuel-macron-talks-to-wired-about-frances-ai-strategy/ [Accessed 30 Nov. 2018].

Schwintowski, Hans-Peter (2018). Wird Recht durch Robotik und künstliche Intelligenz überflüssig? Neue Juristische Online Zeitschrift 2018 (42), pp. 1601-1609.

Zuboff, Shoshana (2018). Das Zeitalter des Überwachungskapitalismus. Frankfurt/New York: Campus.

ARTIFICIAL INTELLIGENCE:
PERSPECTIVES FOR THE FINANCIAL INDUSTRY

VOLKER BRÜHL
CENTER FOR FINANCIAL STUDIES FRANKFURT

Artificial Intelligence (AI) will be one of the key technologies driving the future competitiveness of numerous industries. However, the term "AI" is defined in a variety of ways. AI could be understood as an umbrella term for technologies and systems that carry out tasks otherwise only executable with human intelligence. This requires specific skills that fall into the broad categories of "Sense", "Comprehend", "Act" and "Learn". Through machine learning, modern AI systems can be trained to adapt to changes in their environment, self-optimise and hence achieve better results than earlier versions of AI systems that were based on clearly defined, pre-programmed rules. Based on AI methods, rational and autonomous agents can be developed that collect and analyse relevant information from their environments, come to optimal conclusions based on certain performance parameters and eventually perform physical actions (e.g. robotics) or virtual actions (e.g. chat bots). Machine learning algorithms ensure that the information base of the system is continuously updated so that performance of the system is optimised in an iterative process.

A broad range of applications

AI methods and technologies are already being used in a wide variety of products and digital services, such as navigation systems, digital cameras, drones, and robots. Industrial robotic systems have already been used for many years, especially in the automotive, process and service industries. Highly automated, yet relatively inflexible manufacturing systems are increasingly being replaced with intelligent, decentralised, self-optimising manufacturing systems. The "Smart Factory" is a substantial element of Industry 4.0. Service robots are taking on assistive functions in the household (vacuum cleaner, lawn mower) or in medical care. AI can be applied in home appliances with internet connectivity ("Smart Home"), intelligent traffic control ("Smart Mobility"), or in energy supply ("Smart Grid"). Further areas of application for AI include search engines, speech recognition systems, digital assistants or cyber security. Humanoid robots that emulate the human structure of head, torso and limbs are particularly fascinating examples of AI. The humanoid robot ASIMO (Advanced Step in Innovative Mobility), developed by Honda, is currently one of the most advanced humanoid robots in the world.

Artificial Intelligence in the financial sector

The financial sector will also be significantly affected by advances in AI as financial products and business processes are increasingly digitised. A look at the banking sector reveals that potentially all parts of the value chain could be substantially optimised, if not revolutionised, by introducing AI. This is valid for Customer Relationship Management (CRM), for example, where "predictive analytics" can lead to a deeper understanding about a bank's customer base, thereby gaining valuable insights for marketing strategies, product development or pricing. In central functions such as risk management and compliance, AI can be deployed to fight money laundering, fraud and regulatory breaches. Furthermore, asset management can benefit from the use of AI by developing smart trading strategies. However, the use of AI in the financial sector is still in its infancy, not only for banks and the asset-management industry, but also for the insurance sector and stock exchanges. Besides, there is still a lack of young financial technology firms ("FinTechs") in the German startup ecosystems that are developing or using AI technologies.

From a macroprudential perspective, AI technologies may contribute to global financial stability by improving early warning systems for detecting potential systemic risk. A further aspect that will become increasingly important in coming years consists of the potential interdependencies between AI and the Blockchain technology, e.g. the combination of both technologies makes innovative, real-time payment and settlement systems possible. Hence, usage-dependent pricing models could be implemented along the supply chain. However, the introduction of AI in the financial sector will create new challenges in areas of data and investor protection.

Policy implications

The German federal government has decided to invest three billion Euros by 2025 to support research and development in AI, as a leading position in AI technologies is decisive for the future competitiveness of the German economy. Although research on AI at German universities and research institutions belongs to the best in the world, e.g. in the field of cyber security, robotics or connected cars, there is a substantial need to accelerate the time-to-market for AI-based products and services. The transformation of fundamental research on AI into commercial enterprises is a field where economic policy can be an important catalyst by providing regulatory framework, IT infrastructure and financial support. The lack of venture capital in Germany is still a limiting factor for innovative technology startups in our country. AI has not only the potential to disrupt economic structures such as the labour market, but may also trigger far reaching consequences for our society. Therefore, an interdisciplinary approach integrating multiple disciplines and perspectives is needed to deal with chances and risks of AI in a broader context.

NO ACCEPTANCE WITHOUT CONTROL

KLAUS MÜLLER
THE FEDERATION OF GERMAN CONSUMER ORGANISATIONS

The thought of using Artificial Intelligence (AI) and algorithmic decision-making (ADM) processes in our daily lives makes many of us feel insecure. Most consumers see more risks than opportunities, an attitude brought about by the black-box nature of algorithms and AI. When an organisation or public authority makes a decision supported by an algorithm, one can feel that one is at the algorithm's mercy, finding it incomprehensible. Widespread consumer distrust of AI and ADM processes will make it difficult to improve their societal acceptance and therefore make it challenging to apply them in the business sector and in policy-making. Without trust on the consumer side, there can be no progress.

Consumer distrust is justified

There are various examples that justify consumers' distrust of AI and ADM. A hotel's ranking on a hotel booking platform is determined, in part, by the commission paid to the platform: the higher the commission, the higher the ranking. However, often such platforms appear to be neutral and independent to the consumer. It is therefore understandable that a consumer might feel misled.

Online shopping must also be scrutinised carefully. The sheer amount of information that is available on individual consumers enables companies to continuously refine their price differentiation down to the individual level. Pricing strategies can - like price fixing at a macro level - lead to a very different problem: when intelligent algorithms finetune pricing or conditions, not only does market competitiveness take a hit, but the consumer is also hurt by being charged higher prices.

ADM processes are not risk-free for the consumer in the financial services and insurance sectors either. Selected groups of consumers can be discriminated against, once a profile of their financial capability and psychological and socio-economic characteristics has been created. The potential for damage is great, with the possibility of being charged higher interest rates on loans and higher insurance premiums.

The use of algorithms can also raise questions of liability, for example in the case of Smart Homes. The highly complex nature of ADM systems and AI makes it practically impossible for the consumer to prove the existence of a defect in the system. Who is liable, if the manufacturer of such connected devices has low safety standards, which allow a hacker to break into the Smart Home network?

No green light without trust

A multitude of questions arises from the use of algorithms and AI. The challenges faced by political processes, the economy and society at large in this respect must be addressed sooner rather than later. The Federation of German Consumer Organisations, "Verbraucherzentrale Bundesverband" or "vzbv" in short, is looking to the future, asking itself what measures need to be taken for algorithms and AI to be advantageous to society. Measures that build trust in the technology will be the only ones to succeed in this.

The two factors that can inspire trust and acceptance in consumers are transparency and control: on the one hand, automatic decisions must be transparent and easy to understand. On the other hand, there must be a system in place which ensures that decisions are made within legal and accepted ethical frameworks.

As such, vzbv is advocating the establishment of an independent audit system that is capable of checking and monitoring socially relevant AI and ADM processes. The system can thus verify that the AI system or ADM process being audited adheres to legal requirements, such as rights to equal and fair treatment, and consumer protection legislation, such as data protection. It could potentially analyse AI's impact on individuals and its broader impact on society. Emphasis should be placed on processes that may have a significant negative influence on consumers and/or that affect large numbers of consumers and society at large.

Urgent action is required. Policy makers must now re-align the legal framework with what is happening in practice. In addition, political decision makers must drive the discourse about ethical principles and engage civil society in it. There are 5 essential elements to this discourse:

1. Enabling decisions to be appealed:
   Consumers should have the right to request the review of a specific decision, including the right to outline why they disagree with a decision, to request an explanation of the decision, and to be able to appeal the decision. This review should be carried out by a human being. The right to appeal should include outcomes from ADM processes that do not rely on

personal data. This right to challenge decisions would, for example, apply to cases in which erroneous, distorted data was used in the decision-making process or to correct an unreasonable decision.

2. Introducing consumer rights to information, a labelling and information requirement and disclosure requirement:
   In order to fulfil the consumer's need for information on societally relevant ADM processes, consumers must be able to obtain information whenever an ADM process has been or is being employed. They must be able to request information on the databases and criteria used in the decision-making process, and the basic logic of how the ADM process itself functions. If the data used in algorithms is incorrect or incomplete, mistakes will be pre-programmed. ADM processes must also be subject to a labelling and disclosure requirement.

3. Aligning liability:
   A lack of transparency in ADM processes coupled with the increasing complexity in determining cause and effect in the case of damage to the consumer, has the potential to force consumers to pick up the tab for any damage caused. Thus, any gaps in contract and liability legislation must be closed.

4. Intensifying research efforts:
   Any research that analyses the potential consequences of ADM processes is still in its infancy. In order to achieve a level of transparency in ADM processes which makes them comprehendible, and to better assess potential individual and societal impacts from the use of ADM processes, research efforts, for example in the field of "explainable AI", must be fostered and intensified.

5. Debating ethical principles and societal consequences:
   The ways in which to deal with societal and ethical consequences of the deployment of ADM processes, such as the loss of human autonomy, must be publicly and broadly debated, and finally negotiated. Outcomes from such a debate could for instance become the tenets of an Ethics by Design approach, in which the creators of ADM processes must include certain legal and ethical principles from the get-go when programming and designing the ADM process.

What success looks like today

The German regulatory authority supervising financial markets, the "Bundesanstalt für Finanzdienstleistungsaufsicht" or "BaFin" in short, already subjects algorithm-based systems used in high-frequency trading on the stock market to strict rules. Algorithmic trading is required to be labelled as such and any changes to the algorithm must be documented. The supervisory authority of the stock exchange has the right to inspect an algorithm at any time and prohibit its use in order to prevent breaches to and eliminate abuse of trading rules.

This example proves that ADM processes and AI can be transparent and that their surveillance is feasible. Considering this success, critics will find arguments against regulation, for example that regulation of ADM processes stalls progress and competition, or that it is simply impossible to put into practice, difficult to wield. Introducing algorithm supervision would assuage consumers' doubts and fears. Critics who block such supervision are standing in the way of ADM processes and AI being more widely accepted. It is evident that without consumer trust, there will be no societal acceptance, and without transparency and a regulatory mechanism, public distrust of ADM processes and AI is here to stay.

ARTIFICIAL INTELLIGENCE IN IT SECURITY

MARTIN STEINEBACH AND MICHAEL WAIDNER
TECHNISCHE UNIVERSITÄT DARMSTADT

Expected developments of AI in IT Security

In IT security today, the usage of AI is already established in multiple domains. SPAM detection is a well-known example where support vector machines try to distinguish wanted from unwanted emails. Author attribution combines natural language forensics and machine learning. Deep learning helps in identifying illicit images and has improved malware detection as well as network intrusion detection.

Besides these applications, we can observe that the role of AI shifts from a tool used by IT security to a technology protected by it. This additional perspective is a common development: IT security is most often added to existing software that was developed without taking security into account. Many business solutions have embraced AI recently, often under pressure as users wanted to harvest the benefits of it as quickly as possible. This commonly leads to applications with high security risks due to bugs and error-prone routines. This includes the systems where AI training is executed and those where decisions are made based on the trained nets. Besides the hardening of code and thereby closing gaps for attacks, they require common methods of access control and rights management; AI systems do not differ from other Big Data or cloud systems here.

Not only the systems though, but also the AI algorithms and trained nets become the target of attacks. Various approaches try to mislead or influence AI -based decisions, requiring countermeasures of IT security protecting the core assets of AI.

As a third aspect, AI will also become a tool for attackers. IT security needs to be prepared for attacks that are capable of adapting more quickly to complex security measures, just like intrusion detection systems today aim to identify complex attacks with the help of AI.

Adversarial machine learning will become more common in IT security. Whenever a security challenge can be described by a relatively simple concept on the one hand, but can also be addressed by machine learning on the other hand, the other side, be it defender or attacker, will use adversarial machine learning to efficiently identify weaknesses in the strategy of the other party and deploy specialised attacks or defences against it.

The detection of SPAM, spear phishing or fake news as well as image attribution in security scenarios are examples of where this can be expected or is already common. For example, if an attacker wants to learn about the SPAM filtering capabilities of an online provider or a company, he can probe the solution with the help of a stolen or otherwise acquired email account within a certain perimeter of the target. AI can now send SPAM to the address and verify its receipt. The AI automatically modifies a blocked message until it passes through the SPAM filter. At this point, the attacker AI has beaten the SPAM filter and can now deliver the SPAM to its targets.

In the case of identity documents, an AI can take a passport photo of a person and modify it until it is recognised by a biometric system or any other trained AI. The attacker AI will eventually find the level of modification that is sufficient enough to fool the verification system with minimal visual changes to the photo.

It is most likely that the trend of these two examples will occur in most scenarios where AI has the role of a security checkpoint: the defender AI will become an oracle for attacker AIs trying to circumvent the defender. A race between defender and attacker AI will take place, similar to the obfuscation and recognition of malware or the hacking of systems and fixing of security issues.

Impact on Society

A prominent example of AI with a potential impact on society is predictive policing. While the actual performance and benefits are still being discussed and evaluated, it can be seen as a herald for future developments. The availability of data and the capability of complex analytics will lead to a wave of prediction approaches in a variety of domains.

In predictive policing, aspects like the influence of prejudice and biased data have already been discussed. Similar discussions have taken place on the topic of customer “scoring”, even before the rise of AI. The lesson learned from these discussions is the need for transparency if these technologies are to be widely accepted. Transparency not only addresses equations or algorithms as in scoring, but also the data used for training the AI. Data transparency is closely linked to the concept of privacy by design, which also includes the demand to inform data subjects and for methods to correct false data.

Interpreting the results of AI will be an even more important challenge in the future than it is already today. For example, while it is relatively simple to describe the architecture of deep learning, a common tool in AI, the trained net resulting from combining the architecture with tagged training data is often beyond human interpretation. Given that deep learning is still subjected to relatively high error rates depending on the training quality, an important decision based on an AI result should only be made with a second opinion coming from a human expert. In other words, it is important that AI is not seen as an incomprehensible decision engine, but as an assistant for human experts as long as AI results cannot be interpreted comprehensibly by the subject of the decision.

But at some point in the future the fundamental choice must be made regarding whether we allow algorithms to quickly decide upon important questions on their own based on the collected data. When the AI technology becomes ubiquitous, the sheer number of decisions made by it will render it impossible to execute effective human control. We see preliminary discussions about AI in cars, where its behaviour is questioned from an ethical perspective, for example in the case of an accident. Similar discussions will be necessary in multiple domains, since only a set of accepted rules will allow AI usage that is trusted by society. The role of IT security will then be to verify if the regulations have been implemented successfully, comparable with today's software and hardware penetration tests.

Need for action

Data collected today may (and most likely will) be used in unexpected ways tomorrow. As we see in the advancement of AI-assisted data analysis, it will be possible in the future to combine different data to breach the privacy of individuals up to a level beyond anything currently known and already criticised. Social media accounts, fitness tracker data, consumer behaviour and smart home data will be linked to advanced user profiles if the data is not sufficiently protected. Therefore, both data privacy and user awareness need to be improved.

As already mentioned above, AI will at some point require regulations about its behaviour to prohibit bias or discrimination. Interdisciplinary discussions about the nature and context of this regulated behaviour are necessary to design rule sets that are interpretable by a machine and that can be verified in the case of doubt.

Given the quick development of AI and the growth of its use cases, an interdisciplinary discussion should address these basic issues as soon as possible.

Otherwise, technical regulations will be implemented that are based on engineering concepts but do not consider ethical or legal aspects. This will considerably reduce public acceptance and will thereby hinder or slow down utilising the advantages that AI will offer. For example, it is obvious that the heavy amount of social media traffic cannot be effectively monitored by human observers when it comes to filtering out fake news or hate speech. AI can assist these human observers by identifying modified versions of already known and filtered content, or by pointing to the content that most likely needs moderation. This will raise the accusation of censorship . A political and legal discussion is therefore required on which role AI can take in the control of communication. An accepted trade-off between freedom and regulation is necessary before the actual technical solution evaluates social media traffic.

To summarise, we want to point out that machine learning is not a recent trend in computer security but an already established set of methods in some areas like spam detection. Still, recent advances in AI brings algorithms, which are able to significantly improve security solutions , to domains not addressed by AI so far. From a system security perspective, we will see the need to harden AI applications in the future, as the fast development of software using AI will inevitably bring numerous design and implementation risks with it. Nevertheless, the most important challenge will be the way in which AI is used in the future and which decisions it is allowed to make on its own. This is a question to be addressed by the whole of society; the tasks of IT security will be to provide means of protection, verification and privacy.

Further Reading

Chio, Clarence; Freeman, David (2018). Machine Learning and Security: Protecting Systems with Data and Algorithms. Sebastopol, CA: O'Reilly Media.

Barreno, Marco; Nelson, Blaine; Joseph, Anthony D.; Tygar, J. D. (2010). The security of machine learning. Machine Learning, 81(2), pp. 121-148.

Hitaj, Briland; Ateniese, Giuseppe ; Perez-Cruz, Fernando (2017). Deep Models Under the GAN: Information Leakage from Collaborative Deep Learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17). New York: ACM, pp. 603-618. Available at: https://dl.acm.org/citation.cfm?doid=3133956.3134012.

Shokri, Reza; Shmatikov, Vitaly (2015). Privacy-Preserving Deep Learning. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15). New York: ACM, pp. 1310-1321. Available at: https://dl.acm.org/citation.cfm?doid=2810103.2813687.

Montavon, Grégoire; Samek, Wojciech; Müller, Klaus-Robert (2018). Methods for interpreting and understanding deep neural networks. Digital Signal Processing, Vol. 73, pp 1-15.

Stone, Nathan; Ngoc, Tran Nguyen; Phai, Vu Dinh; Shi, Qi (2018). A Deep Learning Approach for Network Intrusion Detection System. IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 2, no. 1, pp. 41-50.

ARTIFICIAL INTELLIGENCE IN THE MILITARY: MORE THAN KILLER ROBOTS

NIKLAS SCHÖRNIG
PEACE RESEARCH INSTITUTE FRANKFURT

“Artificial Intelligence (AI) is the future. […] Whoever leads in AI will rule the world” (Russia Today, 2018). This was the central message that President Vladimir Putin conveyed to more than one million Russian school students in September 2017. He also promised to ensure that Russian knowledge of AI would benefit the world. However, the competition in this field is already playing itself out globally. Besides Russia, the USA and China are already in the race, with China, for example, having recently published an ambitious AI strategy, namely the „New Generation Artificial Intelligence Development Plan” (Webster et al., 2018). This document predicts China’s world leadership in the AI field as soon as 2030. The EU and several other countries – among them Germany in the autumn of 2018 - have followed suit with their own AI strategies.

Most of these strategies are primarily targeting the use of AI in civil society. However, rapid developments in AI and Machine Learning (ML) are having a significant and perhaps even disruptive impact on the military sector, especially in technological areas largely characterised by “dual-use”. These areas include, for example:

- automation, ranging from automating specific tasks to the “autonomous” behaviour of individual systems;
- comprehensive interconnectedness and the analysis of the resulting large amount of data; and
- swarming.

Subtasks within weapons systems and platforms can, as is the case in the civilian arena, be executed (semi)autonomously. Drones are not only able to fly along certain defined routes independently but they also take off or land without human intervention – even from and on an aircraft carrier.

The importance of complex algorithms is also steadily rising in data capture and analysis. The German F124 Sachsen-class Frigate is, for example, able to simultaneously pursue over 1,000 different airborne targets, with each of these targets being up to 400 kilometres away from the ship (RK Marine Kiel, 2016).

To date, the international debate has focused on controversial so-called Lethal Autonomous Weapons (LAWS), also known as “killer robots” among their critics. In 2014, unofficial conversations concerning LAWS started to take place under the umbrella of the UN Convention on Certain Conventional Weapons (CCW) in Geneva. The discussions turned official in 2017 with the formation of a Group of Governmental Experts (GGE) on LAWS.

No consistent definition of an autonomous weapon has so far been developed in Geneva. However, a definition can be gleaned from an American document authored in 2012, which states that an autonomous weapon is one which selects and engages targets without human input (Department of Defense, 2012/2017: p.13).

Critics, including some nation states and several NGOs, disapprove of and reject the development and use of autonomous weapons that target humans, on the grounds of the violation of international law and human dignity. They argue that a computer is unable to translate fundamental tenets of International Humanitarian Law (IHL), such as non-discrimination and proportionality, into action. Ethically speaking, allowing an algorithm to decide on the life and death of a human being would be untenable. Therefore, critics are demanding a legally binding ban on LAWS, and that an imperative of Meaningful Human Control (MHC) be enacted (Rosert, 2017). However, not all countries share this view. Russia, for example, is arguing that it is premature to discuss a ban, because there is still too much uncertainty around the technology. The United States see potential benefits in the deployment of LAWS, because they might, in fact, lead to improved adherence to international law.

The increasing deployment of software and AI is, however, creating additional problems in security policy that are not in the spotlight in Geneva, and that should be carefully scrutinised. Two examples of the issues are:

Firstly, the speed of military decision-making and the pace of battles is increasing due to the use of computer systems and extensive networks (Scharre, 2018). Decisions need to be made in ever less time yet must draw on ever more available data.

This inhibits the ability to take a critical look at a crisis while faced with it and undermines strategic stability between military opponents. What's more, a vicious cycle of acceleration is created: in order to retain the ability to make decisions under time pressure, more processes need to be handed over to computers, and the cycle thus goes on and on. There were good reasons for arms control measures to aim at the deceleration of critical processes (Altmann and Sauer, 2017).

Secondly, the stronger the dependence on computers and algorithms in military decision making and action, the greater the chances of unforeseen behaviour and system vulnerabilities. This raises questions of the reliability of systems in a crisis, i.e. when subject to extreme circumstances, and of the systems' resilience in the face of external manipulation. Past events have shown, for example, that even systems that were once considered secure have become victims of attack.

Finally, it should not be obscured that AI can be advantageous for security policy. In the area of arms control, AI-based processes can improve verification, i.e. the accuracy and speed at which contract breaches can be identified, and therefore deter potential fraudsters.

The spectrum of possible applications ranges from the analysis of trade data to uncover clues for the proliferation of weapons of mass destruction, to the identification of landmines that is boosted by AI with improved ground-penetrating radars.

To conclude, using AI in military applications can pose major problems, and at the same time be of value in specific areas in the field of arms control. Thus, policy makers, researchers and military officials should discuss the pros and cons of AI more deeply and openly. Concurrently, efforts to negotiate the ban of LAWS should be re-doubled and the debate brought to a resolution. It will be particularly important to flesh out the available options for reaching an agreement on an international ban. This will likely only be possible if the industry representatives who have spoken out against LAWS in the past (Future of Life Institute) are brought to the table, and the discussion can therefore benefit from their technical expertise. It would be worth a try.

References:

Altmann, Jürgen; Sauer, Frank (2017). Autonomous Weapon Systems and Strategic Stability. Survival, Vol. 59 (5), pp. 117-42.

Department of Defense (2012/2017). Directive 3000.09. Available at: http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf [Accessed 1.11.2018].

Future of Life Institute. Lethal Autonomous Weapons Pledge. Available at:

https://futureoflife.org/lethal-autonomous-weapons-pledge/ [Accessed 1.11.2018].

RK Marine Kiel (2016). Fregatte Klasse F124. Available at: https://www.rk-marine-kiel.de/files/bundeswehr/fahrzeuge/fregatte_f124.pdf [Accessed 1.11.2018].

Rosert, Elvira (2017). How to Regulate Autonomous Weapons: Steps to Codify Meaningful Humanitarian Control as a Principle of International Humanitarian Law. Frankfurt/M: PRIF Spotlight 6/2017.

Russia Today (2018). 'Whoever leads in AI will rule the world': Putin to Russian children on Knowledge Day. Available at: https://www.rt.com/news/401731-ai-rule-world-putin/ [Accessed 1.11.2018].

Webster, Graham; Creemers, Roger; Triolo, Paul; Kania, Elsa (2018). Full Translation: China's 'New Generation Artificial Intelligence Development Plan' (2017). [online] New America Foundation. Available at: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017 [Accessed 1.11.2018].

Further reading:

Boulanin, Vincent; Verbruggen, Maaike (2017). Mapping the Development of Autonomy in Weapon Systems. Stockholm: SIPRI.

Heinrich Böll Foundation (ed.)(2018). Autonomy in Weapon Systems. A Report by Daniele Amoroso, Frank Sauer, Noel Sharkey, Lucy Suchman and Guglielmo Tamburrini. Berlin: Heinrich Böll Foundation.

Scharre, Paul (2018). Army of None. Autonomous Weapons and the Future of War. New York, London: W.W. Norton & Company.

Schörnig, Niklas (2014). Automatisierte Kriegsführung - Wie viel Entscheidungsraum bleibt dem Menschen? Aus Politik und Zeitgeschichte (APUZ) Vol. 64 (35-37).

BIG DATA AND ARTIFICIAL INTELLIGENCE: ETHICAL AND SOCIETAL IMPLICATIONS

ROBERTO V. ZICARI
GOETHE UNIVERSITY FRANKFURT

THIS CONTRIBUTION IS BASED ON A SERIES OF INTERVIEWS CONDUCTED WITH HIGH PROFILE EXPERTS FROM INDUSTRY, ACADEME, AND THE EUROPEAN COMMISSION. THE QUOTATIONS FEATURE KEY STATEMENTS FROM THE INTERVIEWEES.

Artificial Intelligence (AI) seems the defining technology of our time. The Big Data revolution and the rise of computing power has made recent AI advances possible.

It is now possible to analyze massive amounts of data at scale and in real time.

Data as a new economic asset

Data has value only if it can be analyzed and if certain insights can be derived from it. Therefore, data has now become a new economic asset. Companies with big data pools have the potential to gain great economic power. The use of big data technologies initially began in marketing. Currently one stage in the life cycle of an emerging science, marketing is a low-risk – and, yes, it is also lucrative. Now, technology is moving beyond increasing the odds of making a sale, to being used in areas that involve higher-stakes decisions, such as medical diagnoses, loan approvals, hiring and crime prevention. An example of this are so-called 'Digital Biomarkers', i.e. the application of mobile and sensor technology to monitor symptoms, disease progression and treatment response. By applying 'Digital Biomarkers', data is used to create a longitudinal real-world profile that, in case of complex syndromes, such as Multiple Sclerosis and Parkinson's Disease, may help researchers to identify signals and changes in symptoms or general living factors, which may have several potential benefits (Zicari, 2018).

Understanding Decisions made by AI

But what if the decision made using an AI-driven algorithm harmed somebody, and you could not explain how the decision was made?

At present, we do not really understand how advanced AI-techniques, such as those used in Deep Learning, really works. It can be extremely difficult to understand which features of the data the machine used, as well as how they were weighted, to contribute to the outcome. This is due to the technical complexity of such advanced neural networks, which need huge amounts of data to learn properly. It is a trial and error. This poses ethical implications.

New ethical and legal questions

Let`s consider an autonomous car that relies entirely on an algorithm that had taught itself to drive by watching a human do it. What if one day the car crashed into a tree, or even worse killed a pedestrian?

“If the learning took place before the car was delivered to the customer, the car’s manufacturer would be liable, just as with any other machinery.
The more interesting problem is if the car learned from its driver. Did the driver set a bad example, or did the car not learn properly?
And in many cases we also don’t know what humans do: for example, we know how to drive a car, but we don’t know how to program a car to drive itself. But with machine learning the car can learn to drive by watching video of humans drive.” - Pedro Domingos, University of Washington (Zicari, 2018d).

Some AI applications may raise ethical and legal questions related to potentially biased decision-making. If the data are skewed and/or the design of the system contains a bias, then the decisions recommended by such systems may be discriminatory against certain categories or groups.

“When AIs learn by themselves, how do we keep them from gowing astray? Fixed rules of ethics are too rigid and fail easily. But if we just let machines learn ethics by observing and emulating us, they will learn to do lots of unethical things. So maybe AI will force us to confront what we really mean by ethics before we can decide how we want AIs to be ethical.” - Pedro Domingos, University of Washington (Zicari, 2018d).

Trust and Explainable AI

“Citizens and businesses alike need to be able to trust the technology they interact with. In order to increase transparency and minimise the risk of bias, AI systems should be developed and deployed in a manner that allows humans to understand the basis of their actions. Explainable AI is an essential factor in the process of strengthening people’s trust in such systems.” - Roberto Viola, European Commission (Zicari, 2018a).

This is a directive for policy makers. But is it really possible to implement it? And if so, then how? Ethical issues need be core considerations during the design phase of an AI project. (Zicari, no year).

Testing and Validating AI

“Robust [standardized?] procedures for testing and validating AIs would be a pragmatic solution, even if we don’t understand fully the heuristics. Perhaps, by extensive testing with actual or synthetic data sets and extreme scenarios, an AI could be validated for its intended purpose, including likely paths of future learning?" - Bryn Roberts, Roche Pharmaceutical Research & Early Development (Zicari, 2018b).

Human Motivations, Intentions, and Data

The overall human motivation is the key to create a 'safe' AI.

"Good data reflects reality and thus can help us gain insights into how the world works. That does not make such discovery ethical, even though the discover is correct. Good intentions point towards an ethical use of data, which helps protect us against unethical data uses, but does not prevent false big data analysis. We need both, albeit for different reasons." - Viktor Mayer-Schönberger, Oxford University (Zicari, 2018e).

"I'm not worried about robots deciding to kill people, I'm worried about politicians deciding robots should kill people." - Oren Etzioni, CEO at the Allen Institute for Artificial Intelligence (Zicari, 2016b).

Regulatory Frameworks and AI

"While self-regulation can be a first stage in applying an ethical approach, public authorities must ensure that the regulatory framework that applies to AI technologies is fit for purpose and in line with our values and fundamental rights." -- Roberto Viola, European Commission (Zicari, 2018a).

There is an intrinsic tension between innovation and regulation. Regulations are normally meant to protect citizens, but some of these are no longer fitting to the modern capabilities of technology and instead drive cost and slow innovation down (Zicari, 2018c).

Stakeholders

Data, AI and intelligent systems are becoming sophisticated tools in the hands of a variety of stakeholders, including political leaders.

Are computer system designers (i.e. software developers, software engineers, data scientists, data engineers, etc.), the ones who will decide what the impact of these technologies are and whether these technologies are to replace or augment humans in society?

In my personal opinion, it is mandatory that the designers of AI systems (and their managers too) be part of the overall discussion on the ethical and societal implications of AI, so as not to leave this discussion (and possible regulations) entirely in the hands of policy makers, politicians, lawyers and philosophers.

AI vision for the future

"Citizens and professionals [...] should become aware of what AI is and what we can do with it. How can I use AI to do my job better, to find the things I need, to build a better society? Just like driving a car does not require knowing how the engine works, but it does require knowing how to use the steering wheel and pedals, everyone needs to know how to control an AI system, and to have AIs that work for them and not for others, just like they have cars and TVs that work for them." – Pedro Domingos, University of Washington (Zicari, 2018d).

References

Zicari, Roberto V. (2018a). On the Future of AI in Europe. Interview with Roberto Viola. ODBMS Industry Watch, 9 Oct. 2018. Available at: http://www.odbms.org/blog/2018/10/on-the-future-of-ai-in-europe-interview-with-roberto-viola/ [Accessed 06.12.2018].

Zicari, Roberto V. (2018b). On using AI and Data Analytics in Pharmaceutical Research. Interview with Bryn Roberts. ODBMS Industry Watch, 10 Sep. 2018. Available at: http://www.odbms.org/blog/2018/09/on-using-ai-and-data-analytics-in-pharmaceutical-research-interview-with-bryn-roberts/ [Accessed 06.12.2018].

Zicari, Roberto V. (2018c). On AI and Data Technology Innovation in the Rail Industry. Interview with Gerhard Kress. ODBMS Industry Watch, 31 July 2018. Available at: http://www.odbms.org/blog/2018/07/on-ai-and-data-technology-innovation-in-the-rail-industry-interview-with-gerhard-kress/ [Accessed 06.12.2018].

Zicari, Roberto V. (2018d). On Artificial Intelligence, Machine Learning, and Deep Learning. Interview with Pedro Domingos. ODBMS Industry Watch, 18 June 2018. Available at: http://www.odbms.org/blog/2018/06/on-artificial-intelligence-machine-learning-and-deep-learning-interview-with-pedro-domingos/ [Accessed 06.12.2018].

Zicari, Roberto V. (2018e). Big Data and Society. Interview with Viktor Mayer-Schönberger. ODBMS Industry Watch, 8 Jan. 2018. Available at:

http://www.odbms.org/blog/2016/01/on-big-data-and-society-interview-with-viktor-mayer-schonberger/ [Accessed 06.12.2018].

Zicari, Roberto V. (2016a). Big Data and The Great A.I. Awakening. Interview with Steve Lohr. ODBMS Industry Watch, 19 Dec. 2016. Available at: http://www.odbms.org/blog/2016/12/big-data-and-the-great-a-i-awakening-interview-with-steve-lohr/ [Accessed 06.12.2018].

Zicari, Roberto V. (2016b). On Artificial Intelligence and Society. Interview with Oren Etzioni. ODBMS Industry Watch, 15 Jan. 2016. Available at: http://www.odbms.org/blog/2016/01/on-artificial-intelligence-and-society-interview-with-oren-etzioni/ [Accessed 06.12.2018].

Zicari, Roberto V. (no year). Personal communication with Steven Finlay.

FIFTEEN RECOMMENDATIONS: FIRST STEPS TOWARDS A GLOBAL ARTIFICIAL INTELLIGENCE CHARTER

THOMAS METZINGER
JOHANNES GUTENBERG UNIVERSITY MAINZ

Introduction

In what follows, I will present a condensed and non-exclusive list of the five most important problem domains in the development and implementation of Artificial Intelligence (AI), each with practical recommendations.

The first problem domain to be examined is the one which, in my view, is constituted by those issues with the smallest chances of being resolved. It should therefore be approached in a multi-layered process, beginning in the European Union (EU) itself.[1]

The “race-to-the-bottom” problem

We need to develop and implement world-wide safety standards for AI research. A Global Charter for AI is necessary, because such safety standards can only be effective if they involve a binding commitment to certain rules by all countries participating and investing in the relevant type of research and development. Given the current competitive economic and military context, the safety of AI research will very likely be reduced in favour of more rapid progress and reduced cost, namely by moving it to countries with low safety standards and low political transparency.

- Recommendation 1
  The EU should immediately develop a European AI Charter.
- Recommendation 2
  In parallel, the EU should initiate a political process and lead the development of a Global AI Charter.
- Recommendation 3
  The EU should invest resources into systematically strengthening international cooperation and coordination. Strategic mistrust should be minimised, and commonalities can be defined via maximally negative scenarios.

The second problem domain to be examined is arguably constituted by the most urgent set of issues, and these also have a rather small chance of being resolved to a sufficient degree.

[1] For a slightly longer treatment, see the following open access publication: Metzinger (2018).

Prevention of an AI arms race

- Recommendation 4
  The EU should ban all research on offensive autonomous weapons within its borders and seek international agreements.
- Recommendation 5
  For purely defensive military applications, the EU should fund research into the maximum degree of autonomy for intelligent systems that appears to be acceptable from an ethical and legal perspective.
- Recommendation 6
  On an international level, the EU should start a major initiative to prevent the emergence of an AI arms race, using all diplomatic and political instruments available.

The third problem domain to be examined is one for which the predictive horizon is probably still quite distant, but where epistemic uncertainty is high and potential damage could be extremely large.

A moratorium on synthetic phenomenology

It is important that all politicians understand the difference between artificial intelligence and artificial consciousness. The unintended or even intentional creation of artificial consciousness is highly problematic from an ethical perspective, because it may lead to artificial suffering and a consciously experienced sense of self in autonomous, intelligent systems. Therefore, it may also lead to artificial subjects or a historically new category of legal persons. Such systems would have to be treated as bearers of rights, because they confer an intrinsic value on themselves by desiring their own, self-conscious existence as an end in itself.

- Recommendation 7
  The EU should ban all research that risks or directly aims at the creation of synthetic phenomenology within its boundaries and seek international agreements. [2]
- Recommendation 8
  Given the current level of uncertainty and disagreement within the na-

[2] This includes approaches that aim at a confluence of neuroscience and AI with the specific aim of fostering the development of machine consciousness. For recent examples see Dehaene, Lau, Kouider (2017), Graziano (2017), Kanai (2017).

scent field of machine consciousness, there is a pressing need to pr promote, fund and coordinate relevant interdisciplinary research projects: evidence-based conceptual, neurobiological and computational models of conscious experience, self-awareness and suffering.

- Recommendation 9
  On the level of foundational research there is a need to promote, fund and coordinate systematic research into the applied ethics of non-biological systems that are capable of conscious experience, self-awareness and subjectively experienced suffering.

The next general problem domain to be examined is the most complex one and likely contains the largest number of unexpected problems and "unknown unknowns".

Threats to social cohesion

- Recommendation 10
  Within the EU, AI-related productivity gains must be distributed in a socially just manner. Obviously, past practice and global trends clearly point in the opposite direction: We have (almost) never done this in the past, and existing financial incentives directly counteract this recommendation.
- Recommendation 11
  The EU should carefully research the potential for an unconditional basic income or a negative income tax on its territory.
- Recommendation 12
  Research programmes are needed to assess the feasibility of accurately timed retraining initiatives for threatened population strata. These initiatives should aim to develop creative and social skills.

The next problem domain is difficult to tackle, because most of the cutting-edge research in AI has already moved out of publicly funded universities and research institutions.

Research ethics

- Recommendation 13
  Any AI Global Charter or its European precursor should always be complemented by a concrete Code of Ethical Conduct guiding researchers

in their practical day-to-day work.

- Recommendation 14
  A new generation of applied ethicists specialised in problems of AI technology, autonomous systems and related fields must be trained.
- Recommendation 15
  The EU should invest in researching and developing new governance structures that dramatically increase the speed at which established political institutions can respond to unexpected problems and actually enforce new regulations.

References

Dehaene, Stanislas; Lau, Hakwan; Kouider, Sid (2017). What is consciousness, and could machines have it? Science (New York, N.Y.), Vol 358 (6362), pp. 486–492.

Graziano, Michael S. A. (2017). The Attention Schema Theory. A Foundation for Engineering Artificial Consciousness. Frontiers in Robotics and AI 4, p. 61.

Kanai, Ryota (2017). We Need Conscious Robots. How introspection and imagination make robots better. Nautilus (47). Available at: http://nautil.us/issue/47/consciousness/we-need-conscious-robots [Accessed 22.11.2018].

Metzinger, Thomas (2018). Towards a Global Artificial Intelligence Charter. In European Parliament (ed.), Should we fear artificial intelligence?, Brussels: European Union PE 614.547, http://www.europarl.europa.eu/RegData/etudes/IDAN/2018/614547/EPRS_IDA(2018)614547_EN.pdf [Accessed 22.11.2018].

ABOUT THE AUTHORS

Petra Ahrweiler is professor of Sociology of Technology and Innovation, Social Simulation at Johannes Gutenberg University Mainz since 2013, but she got a leave of absence for being director and CEO at the EA European Academy of Technology and Innovation Assessment in Bad Neuenahr-Ahrweiler, Germany, until 2017. Before 2013, she had been professor of Technology and Innovation Management at Michael Smurfit School of Business, University College Dublin, Ireland, and Director of its Innovation Research Unit IRU. Furthermore, she was Research Fellow of the Engineering Systems Division at Massachusetts Institute of Technology (MIT), Cambridge/USA.

After a degree in social sciences at the University of Hamburg, Germany, she received her PhD for a study on Artificial Intelligence at Free University Berlin, Germany, and got her postdoctoral qualification for a study on simulation in science and technology studies at the University of Bielefeld, Germany. Her main interests in research and teaching are the mutual relationship of new technologies and society, inter-organisational innovation networks, and agent-based models as means of methodological innovation in social sciences. Petra Ahrweiler won various research prizes, and has been awarded with fellowships of various scientific societies such as the German Academy of Technical Sciences acatech or AcademiaNet, the network of excellent female scientists in Germany.

Volker Brühl is Managing Director of the Center for Financial Studies at Goethe University Frankfurt. Furthermore, he is Professor for Banking and Finance at the University of Applied Sciences for Economics and Management. Brühl serves on the board of various academic and non-academic institutions. He conducts applied research projects in the fields of financial markets, corporate finance, digital transformation, Big Data Analytics, Blockchain and Artificial Intelligence. He has received several awards for his research, e.g. from the McKinsey Global Institute, MIT and Giessen University. Before he returned to academia, Volker Brühl held senior management positions at McKinsey & Company, WestLB, Dresdner Kleinwort, Roland Berger and Deutsche Bank.

The Center for Financial Studies is an independent research institute which closely cooperates with Goethe University. In 2013 the Center for Financial Studies and Goethe University jointly launched the Research Center "Sustainable Architecture for Finance in Europe" (SAFE).

Christoph Burchard is full professor for German, European and International Criminal Justice at the Goethe University, Frankfurt am Main, Germany. He holds a Doctor jur. from the University of Passau, Germany, and a LL.M. from New York University Law School. He researches the foundations, transformations, and crises of criminal justice, inter alia its Europeanization, Internationalization, and digitalization.

Rolf van Dick studied psychology at Philips-University Marburg and earned his PhD at the interfaces of social, organisational and health psychology in 1999. After having been a professor of social psychology and organisational behaviour at Aston University, Birmingham, he became professor of social psychology at Goethe University in 2006. He was visiting professor in Tuscaloosa (USA, 2001), Rhodos (2002), Katmandu (2009), Rovereto (2016), Bejing and Shanghai (2016) und worked as Professor at the Work Research Institute, Oslo (2016-2018). Rolf van Dick is one of the directors of the Center for "Leadership and Behaviour in Organizations" (CLBO) at Goethe University, an interdisciplinary platform for researchers from economics, sociology and psychology and aimed at the exchange with practitioners. He served as dean of the department of psychology and sports sciences. In 2018 he became vice president of Goethe University and is responsible for international affairs, PhD students and post-doctoral researchers as well as for diversity and equality.

Rolf van Dick published more than 200 books and scientific articles and he served as (associate) editor of the European Journal of Work and Organizational Psychology, the British Journal of Management, the Journal of Personnel Psychology, and Leadership Quarterly.

Demian Frank is a Research Fellow at the unit of Sociology of Technology and Innovation, Social Simulation at Johannes Gutenberg University Mainz. He holds a BA in sociology and political science from Johannes Gutenberg University and an MA in social sciences of sports from Goethe University. Demian Frank currently pursues his PhD in the field of AI and its impact on decision making processes in policy relevant contexts. He is part of project CECAN, which develops methods for the evaluation of policies in complex settings across the nexus of energy, environment and food. Moreover, he is involved in the development of a knowledge base-informed expert system for enhancing innovation processes. Prior to his current studies, he was a researcher at the RGZM Leibniz Research Institute in Mainz, conducting evaluations on virtual reality applications and a project assistant in global market research analysing big data.

Thomas K. Metzinger is full professor and director of the theoretical philosophy group at the department of philosophy, Johannes Gutenberg University of Mainz. From 2014-2019 he was a Fellow at the Gutenberg Research College. He is the founder and director of the MIND group and Adjunct Fellow at the Frankfurt Institute of Advanced Studies, Germany. His research centers on analytical philosophy of mind, applied ethics, philosophy of cognitive science, and philosophy of mind. In 2018 Metzinger has been appointed as a member of the European Commission's High-Level Expert Group on Artificial Intelligence (https://ec.europa.eu/digital-single-market/en/high-level-group-artificial-intelligence). It serves as the steering group for the European AI Alliance' and Metzinger is currently working on the European Ethics Guidelines for AI, which will be published in March 2019.

Klaus Müller is the Executive Director of The Federation of German Consumer Organisations (Verbraucherzentrale Bundesverband e.V. – vzbv) since May 2014 after having been the Executive Director of the Consumer Centre of the Federal State North-Rhine Westphalia from 2006 to 2014. In the Federal State of Schleswig-Holstein he had previously been Member of Parliament (2005/6), Minister of Environment, Nature Protection and Agriculture (2003-2005) and Minister of Environment, Nature and Forestry (2000-2003). From 1998 to 2000 he was Member of the German Bundestag. He holds a degree in economics from Kiel University.

Niklas Schörnig is senior research fellow with the Peace Research Institute Frankfurt (PRIF), Germany, and visiting lecture at Goethe-University, Frankfurt. He received his Ph.D. in 2005 with a thesis on American defense industrial policy during the 1990s. In 2012 he received the "Best Article Award 2006-2011" of the German Zeitschrift für Internationale Beziehungen (Journal of International Relations). His research focuses, inter alia, on current trends in warfare, military robotics, military missions of Western democracies and Australian foreign and security policy. His most recent publications include: „Just when you thought things would get better. From Obama's to Trump's drone war" (in: Orient 58: 2 (2017), 37-42) and "Learning Unit 15: Emerging Technologies" (in: EU Non-Proliferation and Disarmament Consortium eLearing Course, https://nonproliferation-elearning.eu/learningunits/emerging-technologies/; 2017, with Frank Sauer).

Martin Steinebach studied computer science at Technische Universität Darmstadt from 1992 to 1999. In 1999 he became a PhD student at the Integrated Publication and Information Systems Institute (IPSI) of the German National Research Centre for Information Technology (GMD).

In 2003 he received his Doctor of Engineering in computer science at the TU Darmstadt, having focused on the topic of digital audio watermarks. In 2007, after the dissolution of IPSI, he moved to the Fraunhofer Institute for Secure Information Technology (SIT), where he became head of the Media Security and Forensics department in 2010. Since November 2016 he has been an honorary professor at the TU Darmstadt, where his lectures include multimedia security, among other subjects. He is the author of over 170 publications and, together with his colleagues, Mr. Steinebach won second place at the Horst Görtz Foundation's IT Security Prize in 2012 for his work on the ForBild project. He currently leads numerous projects on IT forensics and big data security for both the private and public sectors.

Michael Waidner is director of the Fraunhofer Institute for Secure Information Technology SIT in Darmstadt and deputy chairman of the Center for Research in Security and Privacy (CRISP). He holds a professorship in computer science at Technische Universität Darmstadt and is a visiting professor at the Hebrew University of Jerusalem, where he also oversees the development work at the Fraunhofer Project Center. Since 2017, Waidner has also been Chief Digital Officer (CDO) of the City of Darmstadt.

He received his doctorate from the University of Karlsruhe (now KIT) in 1991 and joined the IBM Research - Zurich Lab in Rüschlikon in 1994. During his time there until 2006, he was in charge of research in the field of IT security and data protection and was one of the initiators of the Zurich Information Security & Privacy Center (ZISC) at ETH Zurich. He then transferred to IBM in New York, where until 2010 he was an IBM Distinguished Engineer and Chief Technology Officer for Security, responsible for the technical security strategy and architecture of IBM Corporation. Since 2010 he has been institute director and professor in Darmstadt.
Michael Waidner is the author of over 130 scientific publications and the inventor of over 20 patents. He is an IEEE Fellow and ACM Distinguished Scientist.

Birgitta Wolff has been president of Goethe University Frankfurt since 2015. She has vast experience in both academia and the policy sector. After studying at Witten/Herdecke University, Ludwig Maximilian University of Munich and Harvard University, she became professor of business administration at Otto von Guericke University Magdeburg. She has also conducted research stays at Georgetown and Stanford. Professor Wolff has published various works on human resources and international management. In 2010, she left the university to become Minister of Education and Culture in Saxony-Anhalt. Between 2011 and 2013, Birgitta Wolff served as the Minister of Science and Economy in Saxony-Anhalt.
As member of the High Tech Forum, Professor Wolff has advised the German government on issues of innovation and technology. She is serving on several boards in academia, business and media, including the Board of Directors of the German Public Television (ZDF), the Advisory Board of Deutsche Bank and the Board of Trustees of the Konrad Adenauer Foundation. Since December 2018 she is Member of the Executive Board of the HRK (German Rectors' Conference) and Vice President for Research, Cooperation and Early Career Researchers.

Roberto V. Zicari is professor of Database and Information Systems (DBIS) at the Goethe University Frankfurt and founder of the Frankfurt Big Data Lab. For many years he was the director of Goethe Unibator, a network in Frankfurt supporting bright minds in translating their innovative ideas to market-ready products, a member of the Global Venture Lab network, the Object Management Group and the editor of ODBMS.org – an internet platform informing about big data and new trends in data management and data science – and its blog. Roberto has a sound scientific publishing record in the field of data science and was visiting scientist in the US, Switzerland, Mexico and Denmark. He is also an internationally recognised expert in database and information systems.

SCIENCE POLICY PAPERS
OF THE SERIES

Science Policy Paper 3 (2018)

Wolff, Birgitta (Hg.)
Whither Artificial Intelligence? Debating the policy challenges of the upcoming transformation
urn:nbn:de:hebis:30:3-478510

Science Policy Paper 2 (2018)

Harms, Philipp; Landwehr, Claudia; Scharfbillig, Mario; Schunk, Daniel (Hg.)
Ungleichheit: Interdisziplinäre Perspektiven auf Ursachen und Implikationen
urn:nbn:de:hebis:30:3-478505

Science Policy Paper 1 (2018)

Benz, Arthur (Hg.)
Populismus als Herausforderung für Wissenschaft und Praxis
urn:nbn:de:hebis:30:3-478590

Rhein-Main Universitäten
Eine strategische Allianz

ISSN: 2626-9597

Funded by:
STIFTUNG MERCATOR

SCIENCE POLICY PAPER 2018
03