Summary

Mission AI

The New System Technology

WRR

Mission AI
The New System Technology

About the Netherlands Scientific Council for Government Policy
The Netherlands Scientific Council for Government Policy (WRR) is an independent strategic advisory body for government policy in the Netherlands. It advises the Dutch government and Parliament on long-term strategic issues that are of great importance to society. The WRR provides science-based advice aimed at opening up new perspectives and directions, changing problem definitions, setting new policy goals, investigating new resources for problemsolving, and enriching the public debate.

The studies of the WRR do not focus on one particular policy area, but on crosscutting issues that affect future policy-making in multiple domains. A long-term perspective complements day-to-day policy-making, which often concentrates on the issues that dominate today's policy agenda.

The WRR consists of a Council and an academic staff who work together closely in multidisciplinary project teams. Council members are appointed by the Crown, and hold academic chairs at universities, currently in fields as diverse as economics, sociology, law, public administration and governance, health, and water management. The WRR determines its own work programme, as well as the content of its publications. All its work is externally reviewed before publication.

The council's current term of office runs until 31 December 2022.
It is composed of the following members:

Professor C.C.J.H. (Catrien) Bijleveld,
Professor A.W.A. (Arnoud) Boot,
Professor M.A.P. (Mark) Bovens,
Professor G.B.M. (Godfried) Engbersen,
Professor S.J.M.H. (Suzanne) Hulscher,
Professor J.E.J. (Corien) Prins (chairperson),
Professor M. (Marianne) de Visser,
Professor C.G. (Casper) de Vries,

Secretary: Professor F.W.A. (Frans) Brom.

The Netherlands Scientific Council for Government Policy
Buitenhof 34
PO Box 20004
2500 EA The Hague – The Netherlands
Telephone +31 (0)70-356 46 00
E-mail info@wrr.nl
Website www.wrr.nl/en

Summary

Mission AI

The New System Technology

Corien Prins
Haroon Sheikh
Erik Schrijvers
Eline de Jong
Monique Steijns
Mark Bovens

WRR

THE NETHERLANDS SCIENTIFIC COUNCIL FOR GOVERNMENT POLICY

This is a summary of the Dutch report *Opgave AI. De nieuwe systeemtechnologie* published by the Netherlands Scientific Council for Government Policy (WRR) in 2021. The Dutch report can be downloaded free of charge in PDF format from www.wrr.nl.

Content design: VormVijf, The Hague
Cover image: Steffie Padmos

Executive summary

- In recent years, AI has left the confines of the lab and proliferated throughout society. Advances in the fundamental science of AI have led to an increase in the number of patents and, in turn, emergent interest from businesses and governments. The technology has also captured the public's imagination.
- AI is now being used throughout the economy and society at large, affecting the daily lives of citizens in manifold ways. Therefore, the objective for societal actors, particularly governments, is to develop ways to adequately embed AI in society. To achieve this, we need to understand precisely what type of technology AI is.
- Like what have been called general-purpose technologies, AI is characterized by pervasiveness, continual improvement and innovational complementarities. However, the WRR has coined the term 'system technology' for AI in an effort to emphasize the systemic nature of its impact on society. Other examples of system technologies are the steam engine, electricity, the combustion engine and the computer.
- Embedding system technologies within society entails five overarching tasks:
  1. **Demystification**: Tackling overly optimistic and pessimistic images and learning to focus on the right questions.
  2. **Contextualization**: Making the technology work in practice by creating an enabling socio-technical ecosystem.
  3. **Engagement**: Democratizing the technology by involving relevant actors, in particular civil society.
  4. **Regulation**: Developing appropriate regulatory frameworks that safeguard fundamental rights and values in the long-term.
  5. **Positioning**: Investing in competitiveness and assuring security in an international context.
- For each of the tasks that society faces when embedding AI, we make two recommendations, respectively. Our recommendations for governments are as follows:
  1. *Make learning about ai and its application an explicit goal of governmental policy.*
  2. *Stimulate the development of 'AI wisdom' amongst the general public, beginning by setting up algorithm registers to facilitate public scrutiny.*
  3. *Explicitly choose an 'AI identity' and investigate in which domains changes in the technical environment are required to realize this.*
  4. *Enhance the skills and critical abilities of individuals working with AI, and establish educational training and forms of certification to qualify people.*

5. *Strengthen the capacity of organizations in civil society to expand their work to the digital domain, in particular with regard to AI.*
6. *Ensure strong feedback loops between the developers of AI, its users, and the people that experience its consequences.*
7. *Connect the regulatory agenda on AI to debates on the principles and organization of the 'digital environment' and develop a broad strategic regulatory agenda.*
8. *Use regulation to actively steer developments of surveillance and data collection, the concentration of power, and the widening gap between the public and private sector in the digital domain.*
9. *Bolster national competitiveness through a form of 'AI diplomacy' that is focused on international cooperation, specifically within the European Union.*
10. *Know how to defend yourself in the AI era; strengthen national capacities to combat both information warfare and the export of digital authoritarianism.*

– Finally, we formulate a recommendation to address the institutional aspects of embedding AI within society:

11. *Establish a policymaking infrastructure for AI, starting with an AI coordination centre that is embedded into the political process.*

Introduction

This document was written by the Netherlands Scientific Council for Government Policy (WRR) as part of an advisory project for the Dutch cabinet. The main result of the project was a report, titled "Opgave AI. De nieuwe systeemtechnologie", which can be downloaded at wrr.nl. A translation in English will be published in open access by Springer Publisher by the end of 2022. Although the report was initially written for the Dutch government, we feel the findings are relevant to a much broader international audience. It is therefore that by means of this 'Report in Brief' we present the essence of our analyses as well as recommendations.

Artificial intelligence (AI) has undoubtedly captured the public's imagination in recent years. Once merely a scientific discipline that interested experts and science-fiction fans alike, now AI is routinely the subject of front page headlines. One should not be surprised by this fact. The last several years, we have seen AI extend beyond the confines of the laboratory to society at large. Notable scientific breakthroughs led to patents and various new applications, which, in turn, caught the attention of the private sector. While major technology firms shifted towards adopting 'AI-first' approaches in the early 2010s, numerous governments followed suit in the latter part of the decade with their own AI strategies. As its usage increases, so too does public debate on AI, and civil society, activist scientists and citizens all increasingly become engaged with the technology.

There is an extensive array of literature on the various ways in which AI affects society, ranging from studies on privacy, inclusion and autonomy to proposals for principles, norms, and regulation for AI, with the most prominent of these being the European Union's (EU) proposal for an Artificial Intelligence Act. Alongside this, numerous advisory reports have been published, focusing on general AI strategies as well as specific domains.

We seek to add to this body of work by first taking a step backwards. That is to say, rather than directly looking at the impact of ai upon society, we instead focus on the following question: *What type of technology is AI?* Focusing on this question allows us to gain some perspective on the impact of AI, by drawing on analogies and learning lessons from similar technologies. We argue that AI is a **system technology**, comparable to the steam engine, electricity, the combustion engine and the computer. This approach helps us to look beyond the issues of the day and instead make long-term recommendations on how to embed AI within society. Also, it is at this level that the recommendations can be of value to governments across the globe.

Text box – Artificial Intelligence

What is AI?

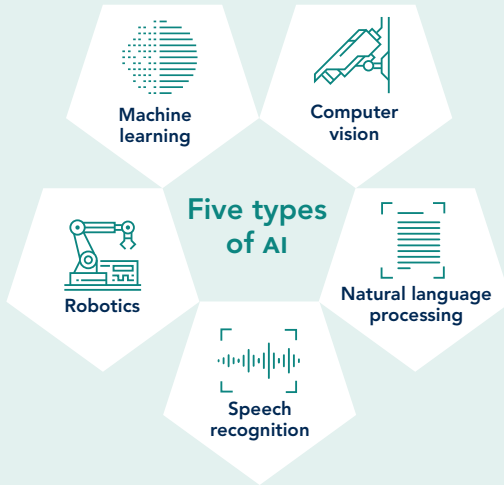
There is no consensus within extant literature on how to define AI. This is not due to a lack of precision on behalf of researchers, but rather to difficulties inherent to the concept itself. Consider the following definitions. The broadest one defines AI as the use of algorithms. While AI does indeed always involve algorithms, these have been in use for a long time and applied within many more basic applications than what we now consider to be AI, such as mathematical formulas or calculators. The strictest definition understands AI as the complete replication of human intelligence by machines. While that is undoubtedly the goal of much AI research, it has yet to be achieved. Hence, while the first definition extends AI beyond recognition, the second defines it as an unfulfilled goal. In between these two extremes lie all kinds of definitions that describe specific features of intelligence, such as interacting, self-learning and predicting.

The reason why it is so difficult to define AI is that it is directed towards a goal that we do not fully understand: human intelligence. Consequently, any definition will always be in flux, insofar as our understanding of AI evolves alongside our understanding of human intelligence. Pamela McMurdock calls this the "AI effect": as soon as we understand how to do something, we cease to call it AI. In this report, the WRR uses the definition put forward by the EU's High-Level Expert Group on AI: "Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals." This definition is sufficiently broad to cope with the AI effect, while simultaneously being specific enough due to the reference to a "degree of autonomy", which constitutes a key feature of recent approaches to deep learning (which we will explain below).

AI in practice

Artificial intelligence plays an increased role in people's daily lives. Specifically, we distinguish five ways in which AI is applied in society at large (see figure 1).

Figure 1 Five types of AI



Machine learning refers to broader fields of AI as well as a specific domain that is also known as predictive analytics or advanced analytics, whereby datasets are analyzed to make specific predictions. Machine learning is used in this way in the financial sector to make predictions about creditworthiness, risk management and fraud detection. Similarly, many police forces across the globe use it for predictive policing, an application that is increasingly being criticized for its negative effect on specific minority groups. Social media platforms also use machine learning to personalize their services. For example, Netflix, YouTube, and Spotify all have 'recommender systems', as do companies like Amazon and Booking.com. Another controversial application of machine learning is the microtargeting technique as employed by Cambridge Analytica during several national elections.

Computer vision deals with the perception, analysis and interpretation of visual images like photos or videos. One example of this is facial recognition, which is used to tag people on social media, unlock mobile telephones, and is used by many security organizations throughout the world. Additionally, computer vision is essential for autonomous vehicles to recognize patterns on the road, as well as for analyzing medical imaging, or analyzing crops for precision agriculture.

Natural language processing deals with the reading, analyzing and generation of human language. Spam filters and search engines use it to determine the relevance of information. Many companies now provide messenger bots on their websites, which help to resolve customer issues. Governments also increasingly employ messenger bots to improve the services they provide for citizens, such as helping them with their tax returns or with applying for certain social services.

Speech recognition is the perception, analysis and interpretation of spoken language. The most famous examples of this are Siri (Apple), Alexa (Amazon), Google Assistant (Google) and Cortana (Microsoft). Spoken language is more challenging than written language because it involves things like tone of voice, accent and homophones. Natural language processing and speech recognition are also regularly combined to make translations from text-to-speech or speech-to-text.

Robotics is the branch of ai that is used in all kinds of robots. Although this often involves some of the aforementioned forms of ai, robotics adds a physical element, that is, the ability to manipulate objects. Robots are used in smart factories, but also to perform tasks in situations that are too dangerous for humans, such as working at the Fukushima nuclear disaster or on lunar missions. Robots can take manifold shapes, such as, amongst other things, autonomous vehicles, smart drones, care robots in healthcare as well as the logistical robots used in the distribution centres of e-commerce businesses like Amazon. Boston Dynamics is a pioneer in this field; indeed, videos of its robodogs and dancing robots have gone viral.

Momentum: AI moves from the lab to society at large

In recent years, there have been extensive news reports pertaining to both developments in AI and how it will affect society. The figure below shows some of the more prominent news items that have caught the public's imagination. AI has existed as a scientific discipline since 1956, when the Dartmouth Summer Research Project on Artificial Intelligence was held.[1] Over the course of its lifespan, AI has extended beyond the confines of the laboratory into society at large via the introduction of applications like decision trees, chess programs and expert systems. However, these are minor applications in comparison to some of the other ways in which AI has influenced society in recent years. Let's take a brief look at the recent momentum in AI.

First, there was a growth in the number of AI-related publications. Whereas the annual growth from 1996 to 2001 was 8%, this rose to 18% between 2002 and 2007. After 2015, it went up to 23%.[2] Out of all the patents in AI, 40% of them referred to machine learning. Within that category, deep learning has seen the strongest growth and has progressed the most in recent years.

As a result, businesses began to take notice and started to apply AI within their operations. In 2014, Google acquired the British AI lab DeepMind, with the subsequent improvements in AI helping to enhance the tool Google Translate from 2016 onwards. Facebook, Amazon, Microsoft, and Apple also routinely buy AI-related startups.[3] In their public announcements, the CEOs of these companies increasingly proclaim AI as a core driver of their business in the future.

During the next phase, governments also set their sights on AI. In March 2017, the Pan-Canadian Artificial Intelligence Strategy was published, which was during the same year that Singapore, Japan and the United Arab Emirates published their AI strategies. China published its New Generation Artificial Intelligence Development Plan, which outlined China's ambition to be the world leader by 2030. The United States (US), France, the United Kingdom (UK) and Germany followed suit as did the EU, which set in motion a range of AI policies. Currently, more than 60 countries have published AI strategies.

---

1 We provide a brief overview of the history of the discipline in the Appendix.
2 WIPO 2019.
3 CBS Insights 2021.

2016

Google's program AlphaGo defeats champion Lee Sedol at Go. When chess master Garry Kasparov was beaten by IBM's Deep Blue in the 1990s, the expectation was that it would take at least a hundred years before a computer would be capable of beating a human at the more complex game of Go.

Microsoft launches Tay, an AI bot that learns from human behaviour on social media. Within only a few hours, Tay transformed into an offensive and racist Twitter troll.

2017

Rumours circulate that Facebook's AI programs started to develop their own language, which was incomprehensible to humans. This triggered the association with uncontrollable AI and Facebook subsequently shut the programs down.

Robot Sophia, which was made by Hanson Robotics, gives a talk at a conference in Saudi-Arabia and is given honorary citizenship.

2018

Google CEO, Sundar Pichai, gives a demonstration of Google Duplex, an AI assistant that can, amongst other things, make dinner reservations, and is vocally indistinguishable from a human.

President Barack Obama shows up in a deep fake video which portrays him giving a speech, although this speech was actually recorded by the comedian Jordan Peele.

2019

IBM's Project Debater competes against one of the best debaters in the world, Harish Natarajan. After a debate about the financing of preschools, a jury chooses Natarajan as the winner.

2020

The Guardian publishes an essay written by GPT-3, a language generator developed by OpenAI. In this essay, GPT-3 argues that humans should not be scared of AI.

Boston Dynamics posts a video clip with its robot family dancing to The Contours' *Do you love me*?

Developments in the science of AI and its various applications by businesses and governments have occurred in parallel with emergent social interest in the technology. Indeed, a wide range of authors, including Brynjolfsson, McAfee, Bostrom and Floridi, have written books on the revolutionary potential of the technology. In response to ever-more concrete applications of AI, more and more publications have focused their attention on the already discernible impact that AI has had upon society. Cathy O'Neil's *Weapons of Math Destruction*[4] and *The Age of Surveillance Capitalism* by Shoshana Zuboff[5] serve to illustrate this trend. Furthermore, specific organizations have been founded for the express purpose of shedding light on the impact of AI on society. For example, the AI Now Institute, founded by Kate Crawford and Meredith Wittaker in 2017, publishes diverse studies as well as a yearly report on the major trends in AI usage. Algorithm Watch is a German NGO founded in 2020 that maps automated decision-making globally. In many countries, there has been a shift towards both the ethics and regulation of AI. Indeed, institutions like the EU, UNESCO and the OECD have published concept regulations and principles for the use of AI. The debate has shifted markedly from the rather broad and optimistic tone of a few years ago, to a focus on the concrete applications of AI and its various pitfalls.

Due to developments in science and patents, AI has recently been taken up by both businesses and governments, which, in turn, has led to considerable public debate. Given that AI has extended beyond the confines of the lab to society at large, the mission now is to determine the best way to embed AI into society; an initiative which requires a broad range of actors and governments in particular. What does that mission amount to exactly? To answer that question, we first need to take a step back and ask what kind of technology AI is.

AI as a system technology

Several prominent figures within the field of AI have drawn comparisons between AI and prior technologies. According to Andrew Ng, the impact of AI is comparable to that of electricity over a century ago[6], while other authors have compared it to the combustion engine.[7] Both Sundar Pichai and Eric Schmidt from Google have drawn comparisons with electricity, with the former even likening it to the impact of the invention of fire.[8] Although these comparisons are interesting, insofar as they point towards the special class of technologies to

4 O'Neil 2017.
5 Zuboff 2019.
6 Lynch 2017.
7 Horowitz et al. 2018.
8 Goode 2018; Morozov 2013.

which AI belongs, what precisely this amounts to is ultimately unclear, because the analogies are not explored any further.

There is an interesting strand of academic literature dedicated to exploring technologies that have a fundamental impact on society. For example, Joseph Schumpeter already spoke of an "evolutionary process of industrial mutation"[9], when outlining his famous concept of creative destruction. Similarly the Nobel Prize winning economist Simon Kuznets spoke of "epochal innovations" that drive an era of economic development, while innovation scientists Chris Freeman and Carlota Perez speak of "new technology systems" and "technological revolutions", respectively.[10]

One concept that is particularly relevant to understanding the nature of AI is that of 'general-purpose technologies' (GPTs), which was coined in the 1990s.[11] Such technologies are characterized by three features: (1) pervasiveness, which pertains to how they spread to sectors, production processes and products, (2) technical improvements, which ensures the continued improved performance of the technology, and (3) innovational complementarities, which lead to productivity growth by virtue of being connected with other technologies and processes.

If we apply these characteristics to AI, it becomes evident that we can classify it as a GPT for several reasons. First, this technology is pervasive throughout the economy, both across different sectors and product categories, which we will elaborate on in the next section. Secondly, there are continual technical improvements driven by Moore's Law in computation and a history of scientific improvement, which, in turn, drive the present widespread application of the technology. Finally, although applications of AI are still at the embryonic stage, there are several studies that quantify its impact on productivity.[12]

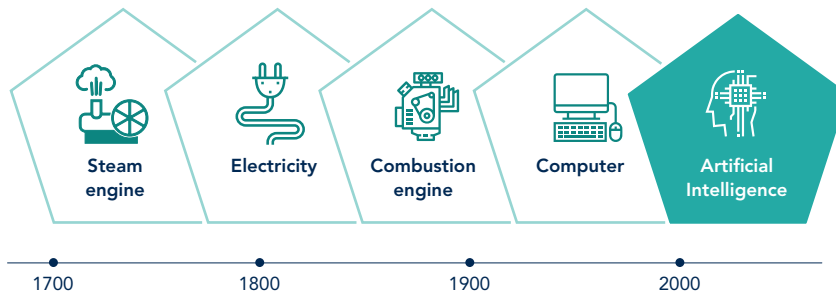
Several scholars have recently taken up this idea of understanding artificial intelligence as a GPT. A study from the American National Bureau of Economic Research, for example, seeks to understand it in precisely this way, while Jade Leung, in her PhD thesis at Oxford University, compares AI with several other 'strategic GPTs'.[13] The Canadian thinktank CIGI similarly describes AI as a GPT.

---

9 Juma 2016.
10 Freeman & Louca 2001.
11 Bresnahan & Trajtenberg 1995.
12 Rao & Verweij 2017; Loucks et al. 2019; Bughin et al. 2018; McKinsey & Company 2020.
13 Agrawal et al. 2019; Leung 2019.

The WRR finds both the lens and definition of GPTs useful for characterizing AI as a specific kind of technology. However, given that it primarily focuses on the technology itself – that is, its general-purpose character –, the term draws less attention to the process of co-evolution between such technologies and society. Hence, we coin the term 'system technology' to characterize AI. By using this term we want to shift the emphasis away from features of the technology towards a focus on its impact upon society, which is *systemic*. This implies concentrating on the complex and unpredictable nature of the technology. Furthermore, although we draw upon literature on GPTs, much of it focus on macroeconomic features and tries to quantify the impact on productivity growth for instance. Moreover, the economic modelling in this body of literature is complex, which, in turn, makes it difficult to draw strong conclusions. Conversely, our approach focuses on the qualitative changes that result from what we refer to as system technologies. Finally, alongside the macroeconomic emphasis of extant studies on GPTs, are attempts to model historical examples of GPTs, which highlight considerable disagreement over the amount of GPTs. In light of this disagreement, we pragmatically choose to look at four historical examples of system technologies: the steam engine (the railways), electricity, the combustion engine (the automobile), and the computer.

Figure 2 AI as the newest system technology



The five overarching tasks of embedding a system technology

When looking at the history of system technologies, one can distill patterns in the way that societies dealt with new technologies in the past and gradually embedded them into our everyday lives. Consequently, these prior experiences can directly inform the current task of embedding AI within society. In our study, we identify five tasks that are necessary to embed a new system technology when it proliferates beyond the lab into society. Although these tasks are interconnected and, as such, there is no general order in which they should be taken up, we analytically distinguish between the following tasks: demystification, contextualization, engagement, regulation, and international positioning. Each task answers a specific question. Before we apply these tasks to AI, we must first explain them through recourse to the history of system technologies.

Figure 3 Five overarching tasks for embedding AI in society

The first task we identify is **demystification**, which deals with the images and beliefs people have about a particular technology. It answers the question: *What are we talking about?* On the one hand, system technologies can engender overly optimistic beliefs. For example, electricity was held to be a 'defining

element of a great civilization' and produced several utopian visions.[14] Similarly, steam-powered railways[15], the telegraph[16] and industrial production[17] were all regarded as instruments for bringing about global harmony and world peace, while science-fiction author William Gibson described a wonderful new world when he coined the term 'cyberspace' in his novel Neuromancer. However, system technologies also bring forth apocalyptic visions of job destruction or of breaking with some form of natural order. The story of Frankenstein symbolized the latter in relation to the use of electricity, but continues to have contemporary relevance through the use of terms like Frankenfish and Frankenfoods to describe GMOs. To cite another example, Thomas Edison explicitly sought to link electricity with execution ("electro-cution") to discredit the technology of his opponents.

Both overly positive and negative imagery can hinder the effective adoption of a new technology, either by raising expectations too high and, in turn, leading to mistakes and disappointment, or by turning society wholly against the technology. Demystification thus requires developing a more realistic account of the technology and what it can do. Moreover, it is important to redirect the public's attention towards issues that are already at stake and that require genuine public debate, which fantastical visions only serve to detract attention from.

The second task of **contextualization** concerns putting a technology into practice and deals with the question: *How will it work?* The complexity of this task explains why the proper implementation of a new technology often takes longer than initially expected. We approach this task through the lens of a socio-technical ecosystem. The technical ecosystem deals with all kinds of supporting and emerging technologies that collectively enable a system technology to function in practice. For example, the important supporting technologies and facilities that spurred the development of the automobile were paved national roads, petrol stations and repair shops. Without them, automobiles were of little use in practice. Similarly, electricity required power stations, transmission cables and a power grid. Emerging technologies that stimulated further electrification were household appliances, such as, for example, irons and washing machines. Whereas the focus tends to be on the new technology itself, these surrounding technologies that 'envelop' it are of paramount importance for its actual use. The social ecosystem relates on a

14 Bakker & Korsten 2021.
15 Van der Vleuten et al. 2017: 27.
16 Gordon 2016: 178.
17 Edgerton 2008: 113-114.

macroeconomic level to the changes that are necessary in business practices and processes to truly reap the benefits from the new technology. It took time to adapt factories and work flows to electrical cables and engines. This is why one tends to see a productivity paradox.[18] On the microlevel, behavioural changes are required on the work floor in order to both build trust in the new technology and learn how to apply it successfully.

The third task of **engagement** pertains to the question: *Who should be involved?* This concerns civil society in particular. As a system technology extends beyond the laboratory into society, both businesses and governments alike typically have both the means and the motives to apply it. Although certain groups in civil societies ordinarily only get involved later, they nevertheless play a crucial role in terms of 'democratizing' the technology by bringing in different perspectives, values and concerns. For example, displaced workers protested against mechanization and steam engines, while a scandal related to electrocuted workers in New York ultimately led to a movement to make the technology safer at the end of the 19th century. Similarly, the automobile ignited a 'battle for the street' that pushed especially low-income groups out of certain public spaces and required pedestrians and children to learn traffic rules designed for the automobile. Authors like Upton Sinclair and Rachel Carson and concerned scientists like Bertrand Russell and Albert Einstein all played a key role in making the use of technologies more socially responsive. Overall, civil society has displayed a wide variety of forms of engagement, ranging from resisting the technology or certain uses of it to monitoring its usage and actively using the technology to pursue its own goals.

The fourth task of **regulation** concerns the question: *What kind of framework is necessary?* The Collingridge dilemma describes the difficulty of this task. In the initial phases, both the nature and impact of a new technology is difficult to assess, thus hindering any attempts to regulate it. When those things eventually become clear, the choices made in the past become incredibly expensive to change, while the power structures that developed in parallel with the technology become harder to challenge. What is clear from the history of system technologies is that the regulatory role of the government has grown over time. For instance, in many countries the government took an active role in railway services and the Rural Electrification Act of 1936, for example, saw to the democratization of electicity in the US. Moreover, although businesses protested, governments intervened in automobile technology through the introduction of safety standards, regulations, rules and drivers' licenses. One of

18 Robert Solow famously said in 1987 that "you can see the computer age everywhere but in the productivity statistics".

the reasons why the role of the government has grown over time is that power becomes concentrated in the hands of a few firms. The classic examples of this are GE, Westinghouse and Siemens in electricity, the Big Three (Ford, Chrysler, General Motors) in automobiles and IBM in computers. During the process of developing regulation for a system technology, the key issues concern whether regulation should be technology-neutral or not, what role private organizations and bodies should play, to what extent existing regulation can be applied and what the balance between flexible and fixed regulations should be.

The final task of international **positioning** concerns the question: *How to relate globally?* This deals with the two interrelated issues of competitiveness and security. Historically, new system technologies have prompted races amongst countries to improve their competitiveness. For example, countries developed a host of policies and strategies to cope with British dominance in steam engines and American dominance in the automobile industry. Such technologies also brought security concerns and helped to determine the outcome of wars. Railways played a central role in the Franco-Prussian War, for example, while the combustion engine was vital during both World Wars. As a result of these twin pressures, countries have often sought to become self-sufficient and 'nationalize' these technologies. However, history shows that no country ever managed to retain leadership within its national boundaries. Rather, the global nature of science and business has led to the ever-more distributed development of technologies. In this respect, the role played by the scientific community and global standardization bodies, such as the ITU and the ISO, not to mention regional cooperation in the EU, is often underestimated with respect to the positioning of countries in the field of new system technologies.

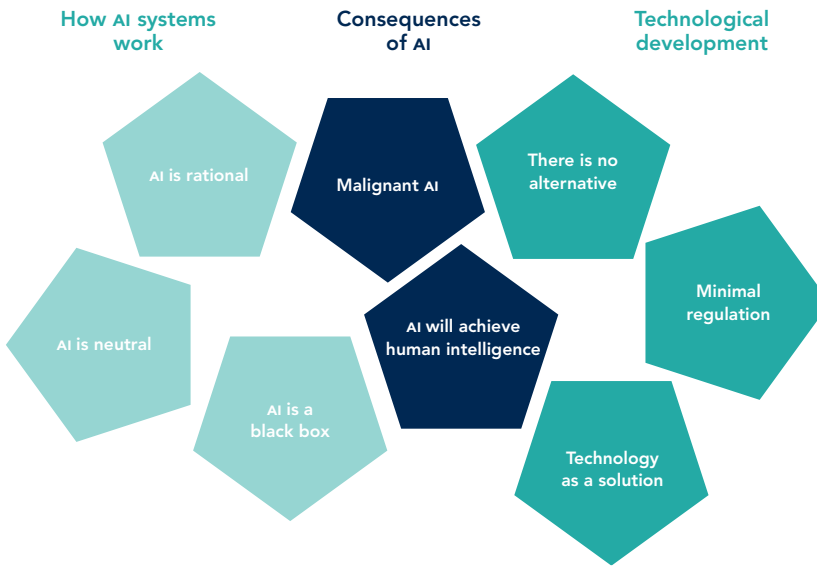
Now that we understand system technologies and the five overarching tasks of embedding AI within society, we can identify what needs to be done in relation to AI.

Task 1: Demystification

The first task deals with the question: *What are we talking about?* System technologies generate all kinds of myths that are either overly optimistic or too pessimistic. Both these viewpoints inhibit the societal integration of AI and detract from asking the right questions about the impact of the new technology.

There are quite a few powerful myths about AI that require demystification. The WRR addresses eight of these in its report, and we will also briefly discuss and criticize them here. The first three myths concern how AI systems work.

Figure 4 Contemporary myths about AI



The first myth is that AI is a neutral technology. This belief is derived from the idea that it lacks all kinds of human qualities, such as, for example, prejudices, fears, and ambitions. Although this is indeed the case, this is not to say that AI systems function neutrally. In fact, both the quality and character of the training data can lead to biases being coded into the algorithms. For example, Amazon's HR algorithm was trained using historical data from previous employees which led the AI system to be biased against women. Similarly, the worldview or interests of developers might also negatively impact certain groups of people. Even if great care is taken to avoid using certain characteristics like gender or race, proxies for these characteristics might still emerge. Postal codes, for instance, can still create a set of biases against certain demographic groups. Moreover, data are never neutral. Rather, data stem from a decision to capture certain aspects of reality, to the exclusion of other aspects. In light of this, some have argued that we should replace the word 'data', Latin for 'what is given', to 'capta', which means 'what is taken'.[19]

A second, related myth is that AI exhibits superhuman rationality. This is based on the notion that it draws upon vast computing power and that it can identify complex patterns in ways that the human brain cannot. One must

19 Greenfield 2017.

problematize this view by emphasizing that current methods of AI actually discern correlation, as opposed to determining causality. Exhibiting such blind faith in rational explanations is thus wholly unfounded. Moreover, the fact that numerous applications of AI are sold on the basis of being highly rational is driven strongly by commercial logic, rather than the notion being grounded in scientific veracity. The two archetypal examples of this are the field of emotion detection, which has no scientific basis[20], and the claim that sexual orientation can be derived from the analysis of pictures.

A third myth about the way AI systems work is that AI is some kind of black box. This is problematic in a number of respects, most notably the fact that the term itself is used in many different and confusing ways. Generally, it denotes either a lack of good documentation or a lack of legal access to how a system works. However, there are no inherent difficulties to open black boxes of these kinds. Truly incomprehensible systems are rare and new methods are continually being developed to make them more transparent. The idea that AI is a black box is a myth that inhibits efforts to bring greater control and transparency to all kinds of applications of the technology.

A second class of myths concerns the future consequences of the rise of AI. The first of these posits that AI is on the cusp of achieving human levels of intelligence ('artificial general intelligence') and eventually surpassing it ('artificial superintelligence'). Although great advancements have been made in this regard, these developments are not likely to come to fruition in the near future. In fact, a poll of leading AI scientists showed that they thought this reality was still eighty years or so away. Many high-profile demonstrations of robots, such as Hanson Robotics' Sophia, or Google's and Apple's digital assistants, hide the highly curated and controlled environments in which they operate. Moreover, intelligence in machines works differently from human intelligence. While a good chess computer may well impress us because chess is a hard thing to learn for humans, it is in fact relatively simple for computers. However, learning how to understand pictures (computer vision) is a much more difficult exercise altogether. This is called the Moravec paradox, which pertains to the phenomenon that AI might advance strongly in certain fields without being able to do things that are basic for humans.

Another myth about the future consequences of the technology is the idea of a malignant AI turning against humanity. Of course, this myth is fueled by popular cinema such as *The Terminator*, *The Matrix* or *Ex Machina*. Describing autonomous weapons as 'killer robots' also reinforces this myth. Like the

20 Crawford 2017.

aforementioned myth, this one projects human features onto machines. However, the fact that humans have intentions, desires or a will to power is not inherent to intelligence, but rather emanates from our biology. Thus, there is no reason to believe that machines will develop similar characteristics. A powerful analogy for this myth is the idea that because airplanes have flying abilities similar to birds, they will at some point feel the urge to build nests.

Thirdly, there is a set of three myths that are part of broader beliefs about digital technologies that originate in Silicon Valley. The first is that such technology should either be completely unregulated or regulated as lightly as possible. One need not go very deep here to unmask this ideological claim. Indeed, globally, the tide is turning against this myth, with the EU in particular taking an active role in regulating technologies like AI. What is important to note is that there is an ideological strand of libertarianism running through Silicon Valley that is opposed to some of the key tenets of democratic governance.[21]

A second general myth about digital technology is that there is no alternative to its current organization and structure. That is to say, accepting the internet means accepting certain power structures, business models and predatory practices. Several authors have analyzed how 'the internet' serves as an ideological construct that inhibits discussion of some of its features.[22] Furthermore, there are various proposals, also by people who have been deeply involved in the development of the internet, such as Tim Berners-Lee, to change the current architecture of digital technology. Many of its current features, such as the rise of powerful private businesses, are the result of specific choices and policies that were made in the past, which means they can be changed.[23]

The eighth and final myth to be addressed can be called 'techno-solutionism' or 'techno-chauvinism'.[24] This refers to the uncritical belief that new technologies can solve all major problems society faces. Although technologies can certainly play an important role in addressing societal problems, this myth leads to an oversimplification and, at times, misrepresentation of social issues. Moreover, it diverts our attention from other 'low-tech' solutions that might work better.

Thus, just like with earlier system technologies, we can discern all kinds of myths about AI that either paint overly optimistic or pessimistic pictures of the technology and its subsequent impact on society. Such myths must be

21 Taplin 2017; Freeman 2001.
22 Morozov 2013.
23 Zuboff 2019.
24 Morozov 2013; Broussard 2019.

addressed head-on because they prevent the technology from being embedded within society effectively, while simultaneously distracting people from asking the right questions. For instance, while there is little reason to fear a malignant rise of the machines, AI systems can indeed be very dangerous for humans without any evil intentions on their behalf, due to the simple and rigid application of rules of automated decision-making that have been built into them by humans. The AI scientist Pedro Domingos formulated this eloquently: "People worry that computers will get too smart and take over the world, but the real problem is that they're too stupid and they've already taken over the world".[25] Therefore, what society needs to address are the ways in which such 'dumb' computers already affect people's daily lives.

Recommendations for demystification

To do so, the WRR makes two recommendations to governments. The first concerns how governments themselves operate and states:

1. Make learning about AI and its application an explicit goal of governmental policy.

On the one hand, many governments routinely fall victim to forms of techno-solutionism. For instance, the Dutch government quickly began to develop a 'corona app' in response to the pandemic without seriously considering whether this would be the right solution for this type of problem. The end result was that, although the app received notable attention, it ultimately proved to be of little value. On the other hand, numerous scandals related to surveillance and the targeting of certain social groups have made many governments understandably wary of using AI systems. To strike a balance between these two extremes, there must be a focus on learning about the technology. AI is not a simple tool that merely needs to be inserted into existing policy. Treating it as such will inevitably lead to one of the aforementioned extremes. To avoid this, more work must be done to educate public officials about what AI can and cannot do, particularly in relation to the proper management of data collection, archiving, and building enough internal capacity, so that the public sector does not become dependent on the private sector for key public provisions.

This recommendation also has implications for work processes in the public sector. Ordinarily, they follow long cycles in which plans are made for big and complex systems. By emphasizing learning, means that the public sector needs

25 Domingos 2017.

to instead develop its capacity to carry out smaller projects with fast evaluation cycles, before proceeding to scale up from these smaller projects. Applying a more iterative approach acts as a safeguard against the failure of large high-profile AI projects that have plagued so many governments around the world in recent years. Finally, the responsibility for AI projects should be delegate to the director and political levels, for the simple reason that experimenting with AI is difficult and thus mistakes are nearly impossible to avoid. Therefore, high-level officials should protect civil servants in their agencies and explain both the intention behind and nature of the projects that are being pursued.

The WRR's second recommendation to governments concerns the need for demystification in society:

2. Stimulate the development of 'AI wisdom' amongst the general public, beginning by setting up algorithm registers to facilitate public scrutiny.

Many parties play a role in demystifying AI in society: journalists, scientists and businesses can all help to either reproduce myths or dispel them. Governments need to do more to stimulate this process, as the prevailing sensationalistic and speculative nature of debates on AI is impeding the process of embedding it within society.

The first step in this process is to create algorithm registers that will explain to the broader public both where and how algorithms are being used in the public sector. Cities like Amsterdam and Helsinki have already begun to develop such registers, which are also being debated on a national level in both the Netherlands and the UK, and EU policies appear to be pointing in a similar direction. The purpose of such registers is to create awareness and initiate debate. In order for that to happen, however, the registers need to provide clear and readily understandable information, and their effect must be periodically evaluated.

Another way for the government to stimulate what we call 'AI wisdom' is to critically examine the message they convey when using the technology themselves. For example, AI is routinely used by governments to detect fraud, which, in turn, sends the message that AI is a technology that is used against citizens and, as such, is a tool for social control. However, the fact of the matter is that AI can and is used by governments in all kinds of other, more beneficial ways, ranging from facilitating public infrastructure, improving air quality,

tackling poverty and providing better healthcare. Governments need to both invest more in these types of positive examples and communicate about them more effectively.

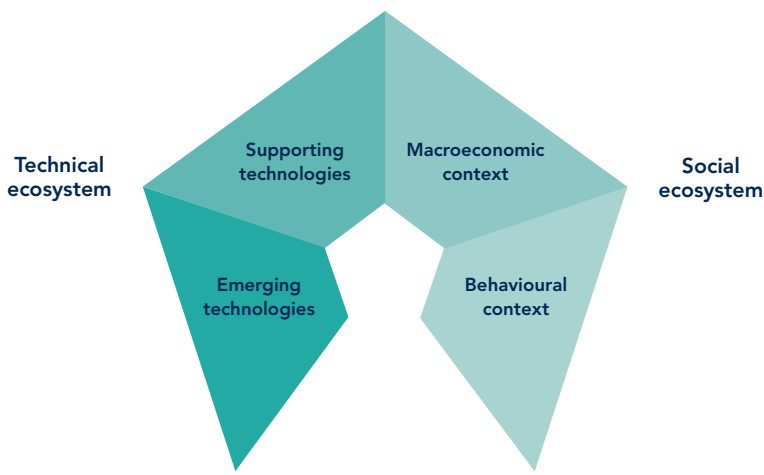
Indeed, the right use of words is important when communicating about AI. Terms like 'killer robots' and 'robo-judges' feed into myths that detract from sensible debate about AI.

A final way to stimulate AI wisdom is via educational and information campaigns. Finland pioneered the development of a national AI course, which many other countries have subsequently introduced. Such programmes need to be encouraged and framed as key features of the AI strategies of countries. Currently, many strategies stress competitiveness and national security, but neglect to pay attention to public attitudes towards AI. This neglect can lead to negative views of the technology.

Task 2: Contextualization

The second task of contextualization pertains to the question: *How to make it work?* This concerns how to make a new technology work in practice within a specific context. We approach this task by looking at the technical and social ecosystem surrounding a new technology. The technical ecosystem in turn consists of what we call supporting and emerging technologies. Let us first look at the supporting technologies surrounding AI.

Figure 5 The socio-technical ecosystem of AI



Although AI is strictly about certain algorithms, there are other technologies and technical facilities that are required to make an AI system work in practice. For example, it requires hardware such as good telecommunication networks. Certain advanced applications of AI, such as autonomous vehicles, require powerful networks with low latency to work. In fact, for any AI application to work, a basic telecommunication infrastructure must be in place. Other necessary hardware are the chips that provide the computing power for AI. Traditionally, central processing units (CPUs) were used, but as time passed the graphic processing units (GPUs) that were being used in the gaming sector proved to be better suited for machine learning applications. Large technology firms are even developing specialized types of chips like TPUs and FPGAs. Thus, making AI work requires having access to the necessary computing power of specific chips. Moreover, supercomputer constitute another form of computing power that functions as a supporting technology. Although not all AI applications require them, they are necessary for complex applications like scientific simulations.

In addition to these different forms of hardware, AI requires another technical facility which can be considered the fuel upon which it runs: data. This is especially true for the prevailing approaches to AI like deep learning, which requires vast amounts of data. This data must not only be vast, but also representative, commensurable and accessible. Advanced algorithms can do very little without such data to run on.

The history of system technologies shows that adaptations to the environment, a process referred to as 'enveloping'[26], are often crucial to making a technology work in practice. Examples of this would be the grid system for electricity and public roads for automobiles. The same holds for artificial intelligence. For instance, intelligent drones or robots are currently primarily deployed in specific environments such as warehouses or industrial production sites because these environments are relatively simple and predictable. Similarly, autonomous vehicles currently work best on highways and are unable to navigate the complex and unexpected environment of a busy inner city. Interestingly, while much of the public's interest is focused on the advanced features of the vehicles themselves, their future might depend just as much on the adaptations that are made to roads to make them work.

Alongside these forms of supporting technologies, we also identify emerging technologies. Whereas the former are necessary from the outset to make a specific technology work, the latter involve innovations that are initially

---

26 Floridi 2014.

separated, but subsequently become connected to the specific technology. 5G networks and the Internet of Things (IoT) are two examples of separately developed technologies that are already in the process of being connected to AI. While the future of other technologies like quantum computing and blockchain technology is less clear, they nevertheless have the potential to significantly affect the development of AI.

In addition to the technical ecosystem, we highlight the role the social ecosystem plays. In order to approach the question of how to make AI work in practice on a social level, we first need to adopt a macro perspective. One prominent issue that arises, is whether this new system technology will lead to massive unemployment. An early study by Carl Benedict Frey and Michael A. Osborne from 2013 suggests that within the next ten to twenty years, 47% of jobs could become automated. Although other studies came up with less dramatic numbers, the fear that AI will replicate human abilities and, thus, make human labour superfluous has, nevertheless, become widespread.

The history of system technologies shows that this is a recurrent fear that, at least until now, has never been realized. The argument that AI is different because it can displace all human cognitive abilities, can be countered by dispelling myths about achieving artificial general intelligence (see above) in the near future as we did in the previous section. Moreover, other factors such as globalization and political choices influence both the rate and shape of automation, whereas widespread unemployment as a result of technology has not yet posed a real threat to advanced economies.

This does not mean that AI will not profoundly affect the job market; it will. However, rather than necessarily leading to massive unemployment, it will likely lead to shifts in the types of jobs and the skills required to perform them. Therefore, instead of imagining a future of 'man versus machine', we should think of jobs requiring different forms of 'man with machine'; that is, in terms of AI augmenting rather than replacing human intelligence. The real challenge of AI for the labor market is figuring out how to make that work precisely.

Another issue from a macro perspective is the so-called productivity paradox. In 1987, Robert Solow famously stated that he could see computers everywhere, except in the productivity statistics. Interestingly, a paper in a recent book published by NBER describes the same phenomenon in relation to AI. Although some authors believe that current technologies will not significantly influence productivity[27], they argue that there is a lagging effect. Indeed, although AI

27 Gordon 2016.

undoubtedly carries great potential to increase productivity, unlocking that potential in practice requires changes in work processes, business models and training procedures. The history of system technologies consistently underscores this point. While electricity and the combustion engine radically changed and improved productivity, their revolutionary nature meant that it took a long time to figure out how to change the operation of factories in line with these new technologies.

This brings us to issues of contextualization on a micro level. Simply put, making AI work requires paying attention to the behaviour of individuals. This entails focusing on the routines, incentives and beliefs that people have. Too often, there will be instances of a technological push that aimed at increasing efficiency, while ignoring the behaviour of actual workers. As a result, this frequently leads to resistance or ignorance which, thereupon, results in the failure of many projects.

One way to think of the right mode of human-machine interaction is the model that distinguishes between three modalities: *human-in-the-loop*, *human-on-the-loop*, and *human-out-of-the-loop*. In the first modality, a system can only perform certain actions or make certain decisions when a human actively does something. For example, while a system may recommend whom to reject for a mortgage, a human makes the final decision. In the second modality, the system can make decisions autonomously, but a human can intervene or change the outcome. Finally, the system is wholly autonomous when a human is 'out of the loop'. The idea behind the model is to design modalities depending on the setting in which AI is applied. Article 15 of the GDPR gives citizens the right to let a human decide in fields that may "significantly effect" their lives.

The model certainly provides a useful approach through which to understand human-machine interaction, but we also must be cognizant of the challenges that remain. For instance, although humans might be kept in the loop, they might still defer to the suggestions of the AI system due to cognitive dynamics like automation bias, thus making their decisions less meaningful. Moreover, increased reliance on automated systems in the long-term may undermine the ability of professionals to question such systems, especially if it is accompanied by an increase in the speed with which the work is carried out. Furthermore, the increased speed and complexity made possible by such systems may eventually make it impossible for humans to follow the reasoning processes behind certain decisions. John Danaher refers to this in terms of the rise of an 'algocracy'.[28]

28 Danaher 2016.

Recommendations for contextualization

The WRR's first recommendation regarding the task of contextualization pertains to the technical ecosystem:

3. Explicitly choose an 'AI identity' and investigate in which domains changes in the technical environment are required to realize this.

Governments must pay attention to all the different components of the technical ecosystem we identified in order for AI to work in a specific context. Given that the focus is often on the development of AI systems themselves, our recommendation instead concentrates on the process we described as 'enveloping'; that is, the changes in the environment of a new system technology that are crucial for its actual use. As we have seen with electricity and automobiles, such changes can considerably stimulate the use of AI.

Since governments cannot and need not implement such changes across every domain, we advise that governments explicitly choose an 'AI identity'. This could include domains that are important building blocks of a country's economy, such as the automobile industry in Germany or agriculture in the Netherlands. It can also include domains that embody important public values in a given society, like healthcare, ecology or governmental services.

Certain governments have chosen to focus on specific domains in their AI strategies. For instance, the French AI strategy emphasizes four domains, while the German strategy is closely aligned with its policy of Industrie 4.0. The WRR recommends that governments carry out thorough assessments of those domains that should be given priority. Once these domains have been chosen, governments should then look at missing components within the technical ecosystem and the adaptations to the environment that are required. Focusing on autonomous driving for instance entails looking at adaptations in roads, designing unambiguous streets signs and the placement of sensors, but might also involve the development of dedicated routes for autonomous vehicles where the complexity of phenomena on the road is greatly reduced—just like highways made fast travelling by car possible in the twentieth century. If a country chooses to focus on the domain of healthcare, then this might entail investing in a uniform and safe system of data collection.

An often-overlooked instrument in stimulating an AI identity is public procurement. Governments are large actors in most economies; their demand for products and services can stimulate or even create certain markets. Consequently, governments should think strategically about the use of this

instrument. For instance, they could stipulate the use of new technologies like AI in the criteria for procurement, as well as shifting the focus to the identified domains via this instrument. In fact, the US government, and especially its military, has already played a large role in the development of AI. Moreover, the EU is also increasingly paying attention to this instrument to stimulate innovation.[29]

The WRR's second recommendation relating to contextualisation addresses the social ecosystem:

4. Enhance the skills and critical abilities of individuals working with AI, and establish educational training and forms of certification to qualify people.

Extant literature on AI places notable emphasis on technical features, such as transparency and explainability, or on ethical principles. However, the concrete interaction between individuals and AI systems requires a greater level of attention. People who work with AI must be trained to understand what such systems can and cannot do, to understand the margins of error, and to distinguish correlation from causality. AI systems can undoubtedly do certain things better than human beings, but perform much worse than humans in other areas. What is required, then, is an understanding of how to deal with the fallibility of both humans and machines, after which we can focus on devising optimal combinations of both.

In time, a system of educational training and forms of certification need to be set up for the use of AI. Think of quality certifications, licenses, and other requirements for using the technology. The history of system technologies again provides instructive analogies. For instance, the rise of the industrial production of food and medicine brought all kinds of new and hazardous products to consumers. In time, the production of these products was organized through an entire system to ensure safety. Food now requires expiry dates and information on ingredients, not to mention that there are many supervisory bodies for the sector. Furthermore, pharmaceutical companies need to acquire licenses to bring medicines to the market, and must also provide medical information for users. There are many ways to organize certification. It can be attached to certain products or to organizations themselves.

29 European Commission 2021b.

It is also important to establish a system of certification for the individuals that are actually using the new technology in certain contexts, in an analogous fashion to the standards that electricians and medical professionals must meet in order to perform certain procedures. In many fields, people are required to have certain licenses or certificates if they want to do something that is potentially dangerous, such as deep sea diving, sailing a boat, or driving a car.

To be clear, we are not proposing that everyone involved with AI must acquire some form of licence. Although all of us use electricity, only those people who have special responsibilities like electricians are required to have a licence. Similarly, only those people who use AI in such a way that it affects the lives of citizens should be required to obtain some form of licence. Although further study would be required to determine the exact form and organization of this 'AI license', two things must be emphasized. Firstly, a licence can be beneficial because it is an enabling measure. That is to say, while many measures are designed to restrict access or prevent particular activities, a licence also serves to grant knowledge and confidence to its holders. In that sense, it can help to stimulate responsible use, as opposed to merely banning irresponsible use. Secondly, educational training should emphasize practical knowledge. While there are already a lot of AI courses globally, which are useful for better acquainting the general public with AI, the emphasis is primarily on theoretical knowledge. However, similar to passing a driving test, it is critical to extensively practice and develop specific capacities, such as a tacit understanding of a technology and the ability to recognize when something is wrong.

Task 3: Engagement

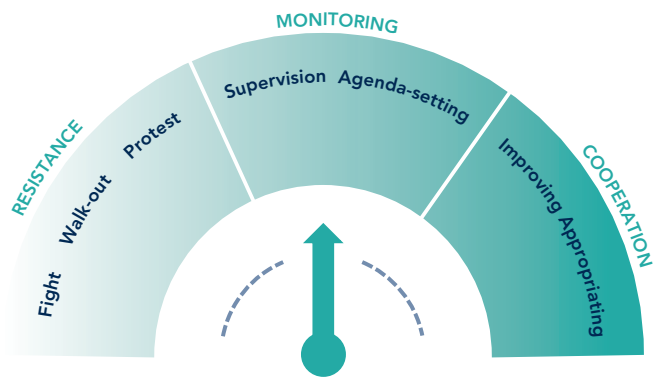
The third task of engagement concerns the question: *Who should be involved?* With the introduction of new system technologies, large companies are ordinarily the first parties to reap their benefits. States also generally have the means and motives to get involved early in the development. The opposite tends to be the case for certain members of civil society. As a result, new system technologies usually deepen existing inequalities and create new ones. The task of engagement, then, is ultimately about bringing in a more diverse range of concerns, values and interests, and is primarily targeted at civil society. It is only through the inclusion of civil society that a new technology can be increasingly democratized.

AI is already deepening existing social differences and inequities. There is an abundance of examples of algorithms that discriminate against women in fields like HR (Amazon) and finance (Apple), and numerous facial recognition algorithms have been proven to perform less effectively for individuals of colour. Moreover, several authors have shown that digital technologies like

AI give new form to existing inequities. For example, Ruha Benjamin speaks of the 'New Jim Code' as a digital version of the old Jim Crow laws of racial segregation.[30] Similarly, Virginia Eubanks speaks of the 'digital poorhouse', which operates as a version of the old exclusionary practice against the poor.[31] Finally, it has been argued that digital technology is making it possible to create 'open air prisons' in China.[32]

So, what does the task of engagement with regard to AI amount to? We can understand engagement by placing its different forms on a spectrum ranging from resistant and antagonistic to more embracing and symbiotic relations to AI.

Figure 6 A spectrum of engagement



At one end of the spectrum, resistance can take the form of outright violence against the new technology. Perhaps because of its intangible character, direct violence against AI is relatively rare. However, an anarchist movement, The Counterforce, did violently target people in Silicon Valley based on their development of autonomous vehicles.

A less violent and more common mode of resistance towards AI are walk-outs, which is when people refuse to work for businesses that develop certain AI applications. In recent years, there have been several walk-outs at many large firms, such as when Google employees protested the work on Project Maven, which developed AI for use in drones in the US military. Similarly, employees of

30 Benjamin 2019.
31 Eubanks 2016.
32 Morgus 2019.

companies like Palantir, Microsoft and Salesforce wrote an open letter speaking out against working for the American Immigration and Customs Enforcement (ICE).

A third form of resistance to AI is protest, which is when people campaign to outlaw certain uses of AI. Several applications in particular have generated such forms of protest. The first being the applications used by police forces across the US to enable predictive policing, which inspired various forms of protest, such as the 'Stop the LAPD Spying Coalition'. Facial recognition is another application of AI which has generated significant protest. In the US and many European countries, people have especially protested against its use by public authorities, which has resulted in the ban of this application in certain places. A third AI application which has garnered notable protest concerns autonomous weapons. In 2012, the 'Campaign to Stop Killer Robots' was launched, and in 2015 an open letter was written against the use of these weapons which was signed by such luminaries as Steven Hawking, Steve Wozniak, Elon Musk and Noam Chomsky.

Alongside these three forms of resistance engagement, we identify a second, more neutral form of engagement, which we categorize as 'monitoring' in the sense of John Keane's monitory democracy.[33] The first form of this type of engagement is supervision. This entails all kinds of organizations that investigate how AI is used and subsequently take action in the event of abuses. For instance, the New York-based AI Now Institute publishes a yearly report that describes trends in the field as well as presenting recommendations to counter bad usage of the technology. In Europe, the German organization Algorithm Watch also publishes a yearly report on how automated decision-making is applied throughout different countries. In the Netherlands, both a range of organizations focused on human rights and privacy and several prominent authors went to court to fight against Systeem Risico Indicatie (SyRI), which was a project of the Ministry and Social Affairs and Employment and several municipalities to detect fraud. The judge eventually banned the project.

A second form of monitoring concerns putting issues on the public agenda. Besides addressing abuse, the aforementioned research organizations also play a pivotal role in informing the general public. Artists like Trevor Paglen and writers like Ian McEwan, for example, use creative tools designed to raise awareness. This form of engagement is expedient in terms of ensuring that politicians and policy-makers have the required knowledge with which to act upon AI-related issues.

---

33 Keane 2009.

The third and final category of engagement we discern is cooperation. Here, the relation to AI can be described as a symbiotic one. One form in which this expresses itself is through improvement. Consider experts that mobilize their knowledge to improve the use of AI. An example of this was the Asilomar Conference for Beneficial AI in 2017, where hundreds of experts developed 23 principles for the beneficial use of AI. Montreal University also mobilized several hundred people to both write about and discuss AI, which ultimately resulted in the Montreal Declaration of Responsible AI. The Partnership on AI and OpenAI are organizations that also involve for-profit organizations in order to improve the usage of AI within society.

The last and final form of engagement we identify is appropriating AI. Here, organizations in civil society deploy AI to serve their own goals, values and concerns. This can be done by local communities or specific social groups. There are currently all sorts of international organizations, such as Women in AI, Black in AI and Queer in AI. Moreover, professional organizations for lawyers, doctors and teachers that have started to use the technology in their field of work come to mind. All these different groups and organizations have their own expertise through which they can help to democratize the use of AI.[34]

In our review of different forms of engagement with AI, we found that forms of resistance like walk-outs and protests are already well-developed, while there is also an emergent field of monitoring the technology. On the right end of the spectrum, engagement is in a more embryonic stage which has to be developed further.

Recommendations for engagement

The first recommendation of the WRR for the task of engagement is:

5. Strengthen the capacity of organizations in civil society to expand their work to the digital domain, in particular with regard to AI.

Based on our review of different forms of engagement, it is evident that there is an increased focus on AI within organizations that work on digital themes like privacy and transparency. However, this is not the case within more traditional organizations that represent groups like patients, teachers, people from low-income backgrounds, workers or people of colour. For democracies to function effectively, it is critical that such groups have a say in the policies that concern

34 Pasquale 2020.

them. Moreover, these traditional organizations have extensive expertise in communities or professional practices that are crucial to strengthening public values regarding the use of AI. Democratic governments thus play a pivotal role in terms of strengthening the capacities of such organizations. Our earlier recommendation to develop algorithm registers also contributes towards this aim. Furthermore, governments can stimulate knowledge of AI-related issues through the financing schemes that they provide to such organizations. Additionally, they can facilitate training and forms of cooperation amongst organizations, such as between 'digital native' organizations and more traditional organizations.

Furthermore, there are already all kinds of existing mechanisms for pluralism and democratization that can aid engagement. For instance, many countries already have rules and laws in place concerning the representation of employees within large organizations, which could be applied to the introduction of AI technologies on the work floor. Finally, governments should ensure that civil society organizations are involved throughout the entire process of creating the policies and regulating AI. Their expertise and input is vital for evaluating how the policies are working.

Our second recommendation for the task of engagement concerns the process of feedback surrounding AI systems. Currently, much emphasis is being placed on the quality and reliability of such systems, but too little on the question of whether these systems actually do what they are supposed to do. Therefore, the WRR recommends:

6. Ensure strong feedback loops between the developers of AI, its users, and the people that experience its consequences.

A lack of good feedback can derive from several factors. One emergent trend is the use of 'real-life experiments' which, after an initial test phase, are deployed without further evaluation. Moreover, the use of aggregated data frequently implies that no personal data is being used, which means that citizens are not required to give consent. This is deeply problematic because it can result in group discrimination, and citizens have few means at their disposal to object to such practices.

Furthermore, although feedback and evaluation are integral components of scientific procedure, the messy world of everyday life can reduce the emphasis on maintaining these rigorous procedures, as AI expands beyond the confines of the laboratory.

Another reason for a lack of feedback is that in certain cases it is simply hard to come by. An example of this are the AI systems used to advise students on further educational programmes. The feedback that is used to evaluate whether the advice was correct is only received after several years, and even then the complexity of all kinds of other factors relating to educational achievement makes it hard to truly judge the adequacy of such systems. Finally, contractual confidentiality or the inherent need for secrecy in fields such as criminal investigation complicate developing effective feedback systems.

Regardless of these dynamics, adequate feedback on AI systems is absolutely crucial. In this respect, we see that two gaps need to be bridged in particular. The first is between developers and users. This could be between the developer of a predictive policing algorithm and the police officer in the field, or the developer of educational and healthcare algorithms and teachers and doctors, respectively. Generally speaking, these two parties have little contact with one another and tend to work separately. However, AI developers need the expertise of their users and need their feedback on how it effects their field of work. At the same time, the users need to have a greater understanding of both how these systems work and what they can do. A second gap exists between users of such systems and the people who are affected by these systems. In light of the examples provided above, citizens that come into contact with the police, as well as students or patients, can be considered people who are affected by these systems. This group also possesses relevant expertise and feedback on the operation of these systems which also must be taken into account.

Therefore, there needs to be more of a dialogue between these three groups of people (developers, users and the people affected by the technology) and governments should encourage this dialogue. Furthermore, feedback mechanisms should be mandated for AI systems that are used in the public sector in domains that have serious consequences for people, such as relating to subsidies, housing, or fraud detection. If there are concerns regarding secrecy, then a tiered system of supervisory organizations should be established to organize the feedback.

Task 4: Regulation

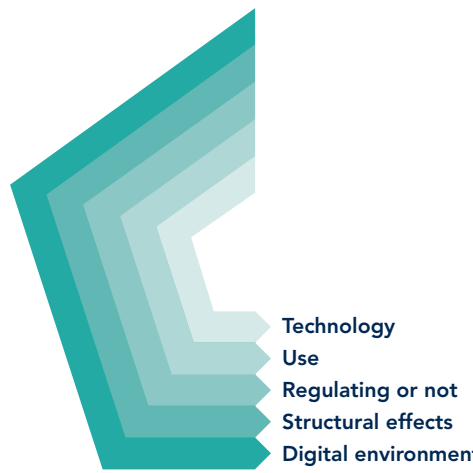
The fourth task for embedding a system technology like AI within society pertains to regulation. This concerns the question: *What kind of framework is necessary?* More so than the other tasks, governments are the key players here, even though other societal actors also play an important role in regulation.

History teaches us that, as a system technology becomes more embedded into society, the role of the government grows. During that process, technology affects an increasing number of public values for which the government is

responsible. For instance, steam engines and automobiles led to air pollution and a host of other environmental issues, which, in turn, led governments to develop regulation. Historically, guaranteeing equal access to the benefits of a system technology, such as electricity, for example, has also been another concern of governments.

One challenge to regulation is formulated by the aforementioned Collingridge dilemma, which posits that in the early phases of a new technology much remains open, which, in turn, makes it difficult to regulate because of a lack of clarity over both the extent and impact of the new technology. Although these things become clearer over time, prior decisions and the attendant power structures that developed make it more difficult to regulate. In this section we outline the key dimensions of regulating a system technology like AI.

Figure 7 Different levels of regulation



The first key question is how the new technology relates to existing regulation. Lyria Bennett Moses identifies four types of issues. Firstly, technologies can create new types of activity that require regulation. Secondly, it might be necessary to clarify existing regulations in light of the new technology. Thirdly, the unregulated use of risky applications can require regulation. Finally, the issue might arise that the assumption behind existing regulation no longer holds in light of the new technology.[35]

35 Bennet Moses 2007.

A further issue revolves around the question of whether regulation should be generic or specific. In light of its broad applicability, generic regulation might appear wholly sensible for a system technology. However, when we look at issues relating to AI, such as the need for transparency or explainability, then it becomes clear that they require specific knowledge of both the context and goals involved in the area of application. For example, explainability in healthcare requires different things than in consumer-based applications or environmental safety. Consequently, the right balance between generic and specific regulation thus needs to be struck in such a way that considers the protection of public values, while, simultaneously, leaving space for innovation.

The task of regulation also raises other important questions that we can mention only briefly within the scope of this summary. One is whether regulation should be specific to certain technologies or whether it should be technology-neutral. Another question arises around what precisely constitutes the right level of regulation. Although certain regulation should take place at the national level, the EU is also an important actor when it comes to regulating AI, as well as the relevant fora for regulation on the global level. A final question pertains to the role of different actors in the process of regulation. Companies have a role to play through self-regulation, albeit this has serious limitations. Governments and businesses, for instance meet in public-private settings for standard setting. Furthermore, scientists and parties in civil society also play a role in regulation, by bringing their expertise to the table, for example.

We already mentioned how the Collingridge dilemma poses a challenge for regulation. That is to say, a lack of knowledge, path dependency and power structures make it difficult to regulate new technologies and, as a result, governments can become reluctant to intervene and instead focus on reacting to immediately pressing issues. As aforementioned however, over time the role of government expands to encompass a more active and directive role. Hence, the focus needs to shift from acute issues related to the technology itself and its direct effects, towards adopting a broader perspective on how the technology can be societally embedded.

Initially, the generation of electricity led to a host of issues, such as the safety of wires, and over time planning and overview were required to develop national grid systems. Similarly, after cars immediately caused issues, the focus shifted towards the organization of the physical environment. Decisions had to made about what role the car would play in cities, who would be allowed on which roads and what the national road system should look like. In other words, a broad regulatory agenda was required that took the broader effects of the new technology into account and embedded these effects within the society's public values. Governments today should approach AI in the same way. Spring 2021

European Commission set an important step towards a broader framework with the proposed AI Act. However, numerous issues remain to be addressed. What is more, we find ourselves at a moment in time where an overall perspective needs to be developed on how society wants to relate to the digital world, among others AI. Inspiration can be found in the type of initiatives taken by governments as regards our physical living environment. In the past decades numerous countries have based short term policy choices on where to build what (housing, roads, forest, etc.) on their strategic plans containing a long-term perspective on how society relates to the physical environment. The WRR recommends that similar strategic plans need to be developed for the digital world and its interaction with society at large.

Recommendations for regulation

Concerning the task of regulation, the first recommendation of the WRR is:

7. Connect the regulatory agenda on AI to debates on the principles and organization of the 'digital environment' and develop a broad strategic regulatory agenda.

There is currently a lot of activity surrounding the regulation of AI. This involves specific uses of AI, such as facial recognition and autonomous weapons, but also questions about data usage and the levels of risk involved in different applications. The EU's draft Artificial Intelligence Act[36] lays down a framework for the latter, which thus shows that regulators are responding to the acute dangers and opportunities that AI provides.

As the technology becomes more broadly used in society, its impact will become more general and diffuse. Moreover, secondary and tertiary effects will emerge. Many of these effects will be determined less by the technology itself than by the conditions under which it is used and the broader economic dynamics to which it is connected. This will require governments to adopt a broader and more directive role.

Regarding the physical environment, governments have all kinds of programmes, regulations and planning agencies. We suggest adopting a similar approach for the digital environment, in order to shape and safeguard public values. Failure to adopt a more active and broad approach can lead to governments losing grip over the digital environment, which is increasingly

36 European Commission 2021a.

controlled by several large multinational corporations. However, digitalization has well and truly entered the daily lives of citizens and, as such, its design requires the oversight of democratic governance.

Of course, this will be a long-term process that will take decades to take shape, and which will generate much uncertainty in the interim. Governments can, however, take concrete actions towards developing a regulatory agenda for the digital environment. First and foremost, they can make a list of legal provisions in which the effects of AI can already be made explicit. For example, considering automated decision-making, liability, archiving and the legal status of autonomously acting systems. Secondly, governments can improve their knowledge of broader societal effects by focusing more on the signals that come from institutions that are closer to the field than policy-makers themselves. For instance, supervisory bodies have valuable knowledge about what is happening in specific domains or markets, while jurisprudence also has an important signalling function. Finally, governments can consider conducting annual assessments of AI's effect on society in order to gain some perspective on developments in the digital environment.

Thus, a strategic regulatory agenda that is aimed at the long-term effects of AI on society requires a shift in focus from ad hoc issues, to investigating the more structural effects of AI. With this in mind, the WRR's second recommendation for regulation addresses several of these structural effects:

8. Use regulation to actively steer developments of surveillance and data collection, the concentration of power, and the widening gap between the public and private sector in the digital domain.

With respect to the broader societal effects of AI, we identify three structural issues that require regulatory action. The first being the steady growth of data collection and surveillance throughout society. Currently, data collection is judged on a case-by-case basis, for example by looking at what is permissible for police investigations or traffic management. However, such an approach misses the real structural issue at stake, namely that forms of surveillance are steadily increasing. In part, this is due to the fact that the collection and analysis of data is central to the business model of many digital firms. The proliferation of sensors in the physical environment and inside household goods gives further impetus to this development. Behaviour, biometric information and even emotions are now monitored and influenced by means of AI. Even if in specific cases the use of these technologies is wholly justified, the broader question

of how deeply surveillance should penetrate the different spheres of society remains. Ultimately, this is a political question.

The second structural issue is the concentration of power. The largest players in AI are the business titans of the internet, such as American companies like Google, Facebook, Amazon, Microsoft and Apple. The corona pandemic led to ever-more digitalization (like video-conferencing) within the spheres of education, work and healthcare, wich further increased our dependence on a few foreign companies. Globally, resistance to this concentration of power is increasing. The European Commission, as well as the American Department of Justice and the British government have all spoken out against the effects of this concentration of power. The growing influence of these companies on society implies that private commercial concerns increasingly trump democratic public concerns. It is not yet clear how this concentration of power should best be addressed. There are calls to split up these companies or to regulate them as utilities. The European Commission has made the most headway in terms of devising concrete measures.

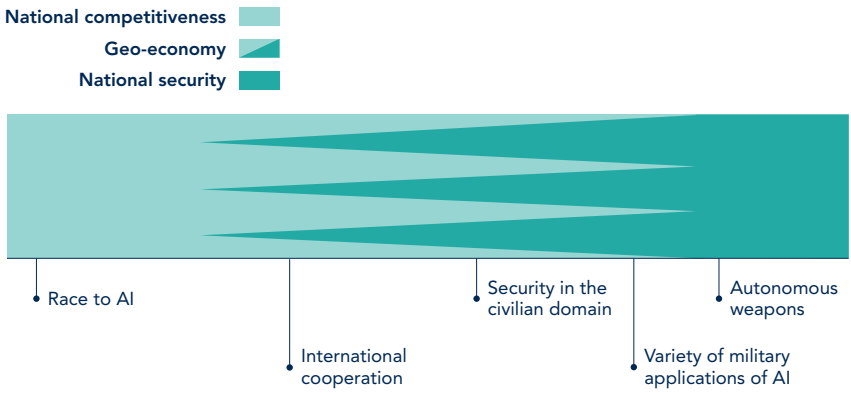
The third issue we discern is the gap in AI capacities between the private and public sector. Skill levels in the public sector are currently lagging behind, and concerns about the risks of AI currently limit its use and development in the public sector. This gap can lead to all kinds of problems. Amongst other things, it could mean that public institutions have less of an understanding and grip over their domains of responsibility. For instance, private parties are increasingly gathering valuable information on air quality and traffic that is valuable for the maintenance of societal infrastructure, which means that the public sector could become increasingly dependent on them. Valuable information would no longer be in public hands, which could lead to more illegitimate practices due to the lack of public oversight and involvement. Consequently, there is an urgent need for the public sector to bridge the knowledge gap with the private sector.

Task 5: Positioning

The final task for society that we identify, is positioning AI within an international context. It deals with the following question: *How do we relate to others globally?* All of the previous four tasks also have an international dimension. For instance, we saw that engagement by scientists takes place through international fora and organizations. Regulation also occurs at the international level, as we saw when we looked at the EU's draft Artificial Intelligence Act. Yet it is important to discuss the task of positioning explicitly, because it involves two specific issues: national competitiveness and security. The concept

of 'geo-economics' sheds light on how both types of issues are becoming increasingly intertwined.[37] However, we will first discuss them separately.

Figure 8 Issues regarding national competitiveness, national security and geo-economy



When we look at national competitiveness, the question of how to assess a country's AI capacities arises. Although there are several AI indices, this query is complex and needs to take several variables into account. Based on the work of several authors, we discern five variables[38]:

1. The quality of fundamental research
2. The availability of data
3. The required hardware
4. A dynamic private sector to commercialize the technology
5. An enabling government

Based upon these variables, two countries appear to be strong across all dimensions: The US and China. Depending on how much weight is placed on the various variables, different estimates of their relative strengths are obtained. For example, given that Jeffrey Ding focuses on hardware, he believes that the US has an edge over China in terms of the production of advanced chips. Conversely, Kai Fu Lee places more emphasis on data availability, which is why he believes China will have the upper hand.

37 Luttwak 1990; Scholvin & Wigell 2018.
38 Ding 2018; Lee 2018.

The government's New Generation Artificial Intelligence Development Plan, the most ambitious of its kind in the world, also gives China an edge.

The EU as a whole also has strong AI capacities. In fact, its fundamental research is ahead of China. However, it scores much lower with respect to having a dynamic private sector that uses AI. The UK is also strong in AI, particularly in terms of fundamental science, which dates back to the work of Alan Turing. Another country that is strong in fundamental science is Canada. Many pioneers in the field of AI work here, which led the country to publish the first AI strategy.

Many other countries have specialized in certain applications of AI. For example, because Japan, Germany, and France have large automotive industries, they focus on industrial applications, particularly in relation to mobility. In line with its security policy, Russia is strong in fields like facial recognition.

Since 2017, dozens of countries have published AI strategies. One discernible pattern in these strategies is that many countries have chosen specific domains to focus on, thus creating something we refer to as an 'AI identity'. This can involve sectors that a country's economy is already competitive in, like the aforementioned countries that focus on AI in mobility, but also pertains to the specific challenges countries face. For instance, Japan, France and India emphasize healthcare, sustainability and inclusiveness, respectively.

When one considers these capacities and strategies, it may give the impression that countries around the world are engaged in an 'AI race' to the top. Indeed, this frame is prevalent in a vast number of government documents and the news media. In certain domains that may well be an adequate description of events. Countries do indeed compete over scarce talent, while military applications of AI can yield decisive advantages to certain countries. What we want to emphasize, however, is that framing the problem in this manner has a number of shortcomings. First and foremost, it suggests a zero-sum situation between countries, which is wholly misleading. In fact, the benefits of system technologies are always experienced widely. For example, electricity and the combustion engine stimulated economic growth globally and benefited consumers worldwide. Although commentators tend to focus on a few advanced AI laboratories, AI is also diffused throughout the economy and can be used by different actors for a wide range of goals.

A second shortcoming of framing the situation as a race is that it suggests that all countries are racing towards the same goal. What that goal is, however, is unclear. As we have already seen, countries can have very different profiles in

terms of fundamental science, applications, or their choice excel in particular industries. Thus, there is no reason to assume that all countries are moving towards the same goal.

The third problem is that it suggests that there is a tension between competitiveness and protecting important public values. The idea of the race is often used to describe privacy protection and regulation as being impediments to innovation. However, the history of system technologies shows that this is factually incorrect. Safeguards and regulations have made many automobiles better, thereby improving their usage, rather than hindering it. The EU explicitly emphasizes this connection between competitiveness and trust.[39]

Finally, the AI-race frame implies that a country can win by constraining AI development to its own borders. While countries have indeed tried to do this across history, it has consistently proven to be counterproductive. The development of new system technologies has always been an international affair, and attempts to nationalize it have weakened countries' competitiveness.[40]

Alongside competitiveness, the task of positioning also relates to national security. In the military domain, the idea of a zero-sum race carries more weight than in the economy. Indeed, there is much talk about an 'AI arms race.' Even Vladimir Putin famously stated that the country that takes the lead in AI will rule the world.

The most frequently discussed form of military AI concerns autonomous weaponry. This is probably because it captivates the imagination, conjuring images of 'the rise of the machines' or Terminator-like scenarios. There is a wide range of military technologies with varying amounts of autonomy, ranging from Israel's Harpy drone and South Korean robot weapons at the North Korean border to the American AEGIS and Patriot systems. China is a large exporter of drones with varying degrees of autonomy.

The phenomenon of autonomous weapons deserves a considerable amount of attention, and numerous campaigns have already been launched to enact a ban on the technology. Although it is difficult to predict whether this will occur, we highlight three challenges associated with implementing such a ban. First of all, it is really hard to define what an autonomous weapon exactly is. Autonomy can be defined across multiple dimensions, including the sort of

39 Smuha 2019.
40 Leung 2019.

activity conducted autonomously, the role of humans, and the required level of intelligence.[41] This raises all kinds of challenging questions. For example, if a human selects a target and the weapon then duly proceeds to shoot at it until it is destroyed, is that considered as an autonomous action? Is a defensive system that automatically launches missiles out of the air an autonomous weapon, and, if so, do we want to ban those types of systems? A landmine also acts autonomously, but considering the low complexity of the action, are we able to define a minimum level of required intelligence?

A second challenge related to the regulation or banning of autonomous weapons concerns the dual-use character of AI: The same technology that is used in consumer applications can also be put to dangerous military use. The ability to recognize and follow moving objects can be used to both record the route of a wedding car and to destroy a vehicle. Drones can use facial recognition to identify the buyer of a package they are delivering, but they can also use it to identify and kill that same person.

The final challenge pertains to the motivation of powerful countries. Although a growing number of countries support the prohibition of autonomous weapons, the majority of these countries have limited military capabilities in this area, such as Pakistan, Ecuador, Ghana, and the Vatican. Conversely, countries like the US, China, Israel and Russia are far more reluctant to curb their ambitions in the field of autonomous weapons.

These are three important challenges related to the regulation of autonomous weapons, but what is important to emphasize here is that the impact of AI on security goes far beyond just these three challenges. Another important field which AI impacts upon is both the quality and speed of decision-making in the military.[42] By virtue of its ability to analyze more sources of information, the use of AI in decision-making also poses new challenges. Parties might, amongst other things, seek to intentionally mislead the systems of their opponents. Which is not necessarily uncommon, as the manipulation of only a few pixels previously led a neural network to misidentify a car for an elephant.[43] Hence, combatants might use such technologies to cloak threats or to trigger automatic attacks against the wrong targets.[44] Another way in which AI can affect the military is by supporting internal processes like logistics or communication.

---

41 Scharre 2018.
42 Tonin 2019.
43 Libicki 2019.
44 Lin 2019.

These examples are all still within the military domain. However, security is a much broader issue, and thus more attention should be directed towards security threats in the civilian domain. Cybersecurity and the threats it poses to the digital infrastructure are already receiving increased attention.[45] The growing amount of information flowing through that infrastructure makes it a target for both nations and criminal, non-state actors. Furthermore, social networks and open (government) databases can provide adversaries with sensitive information, and AI systems can discern complex patterns from seemingly innocent information. For example, anonymized traffic data from a New York taxi company was quickly deanonymized and used to calculate the salaries of specific drivers, to identify the customers they drove to strip clubs, and even which traffic drivers were Muslim, based on the time of stops during prayer time.[46]

Microtargeting and sentiment analysis are other ways in which AI is being used for what has been called 'information warfare'.[47] One particularly dangerous use of AI that is on the rise are so-called 'deepfakes'. AI makes it increasingly easy to create artificial messages, audio and videoclips, which eventually may become just as readily available as photoshop.[48] Technology expert Aviv Ovadya foresees the coming of an 'Infocalypse'.[49] The final trend we identify is the rise of 'digital authoritarianism'. In particular, China and Russia have developed advanced tools to control and manipulate the internet and are increasingly exporting these tools to other countries, including democratic countries in Europe.[50]

Recommendations for positioning

With regard to the first issue of positioning, national competitiveness, the WRR makes the following recommendation:

9. Bolster national competitiveness through a form of 'AI diplomacy' that is focused on international cooperation, specifically within the European Union.

45 WRR 2019.
46 Crawford 2021.
47 Singer & Brooking 2018.
48 Schick 2020.
49 Warzel 2018.
50 Wright 2019.

Previously, we discussed the prevailing frame of a competitive 'AI race', as well as its four shortcomings. Rather than emphasizing a zero-sum competition, we argue that countries can increase their competitiveness through international cooperation in several domains. The first domain is in fundamental research in AI. Within Europe, there are already networks like CLAIRE and ELLIS. Secondly, countries can cooperate in developing AI applications or supporting services. An example of the latter is the European project to develop a cloud and data infrastructure, Gaia-X. Cooperation can also take place in relation to strengthening and protecting existing companies. The Sino-US trade war caused significant disruption to European technology firms and greater cooperation within the EU could serve to strengthen these businesses. The final field of cooperation comprises regulation and standardization. Regarding the former, the EU has greater international weight than is commonly recognized. The latter is often overlooked, but it is a highly important domain for new technologies.[51] There is currently a 'geopoliticization' of standards going on and countries need to be aware of what this implies.

What we coin as 'AI diplomacy' consists of multiple aspects. Firstly, governments should formulate goals for the aforementioned domains, but also look at the synergies between them. Secondly, they should be sensitive to the geopolitical goals underlying the positions of other countries in international fora. In short, countries must integrate their AI activities into the international arena.

With regard to the second issue, security, the WRR recommends:

10. Know how to defend yourself in the AI era; strengthen national capacities to combat both 'information warfare' and the export of digital authoritarianism.

We saw above that with respect to security, a lot of attention is paid to autonomous weapons. We also highlighted some of the challenges involved with regulating or banning such weapons. While such sustained international attention in this field is promising, too little attention is paid to the ways in which AI can threaten security within the civilian domain. There is a brewing 'information war' that is partly being fought manually, by people in click farms for instance, but which also involves AI-powered technologies like microtargeting, natural language processing, sentiment analysis and deepfakes. These technologies can infringe upon the freedom and rights of individual

51 See Bradford 2020 on the 'Brussels Effect'.

citizens, but can also threaten society as a whole, through the manipulation of elections or by eroding trust in institutions by spreading fake news and conspiracy theories.

Moreover, certain countries are increasingly using and exporting technologies that are undemocratic, which are enabled by AI. While the surveillance capabilities of totalitarian regimes in countries like East Germany were ultimately limited by the percentage of the labour force that they could devote to surveillance, AI makes surveillance scalable, centralized and relatively cheap. Countries need to have more awareness of the implications of surveillance technology and other digital threats to security. The infrastructure of Huawei is currently facing considerable scrutiny, however there is scarce debate about cameras with facial recognition, monitoring of smart city infrastructure and the software used in public services. Furthermore, democratic countries are not only on the receiving end of digital authoritarianism. Rather, European firms are also involved in exporting technology that is subsequently used for authoritarian ends. Hence, their governments should develop and enforce policy to prohibit such exports.

The brewing information war is complex and diverse. No one knows all the means through which it will be fought or won. However, governments need to first and foremost increase their awareness of it and structurally monitor it, so that they can start to devise effective policy measures against information warfare.

Final recommendation: Towards a policy infrastructure for AI

Across all of the five tasks, AI requires work from society as a whole to become embedded within our daily lives. The aforementioned recommendations aim to provide governments with a starting point for taking up the tasks of demystification, contextualization, engagement, regulation and positioning. This process can be supported by developing an institutional infrastructure that facilitates the exchange of knowledge and the coordination of state activities. Ultimately, this will allow governments to both address the opportunities and risks of AI in a structural way and to develop an integrative approach to embedding this system technology in society.

As aforesaid, history shows that every new system technology created new tasks for the government to ensure that its development and use was in alignment with public values at that juncture. Consider the establishment of new authorities to determine the criteria for automobiles to be allowed on public roads, authorities to certify people as legal drivers, adjustments to the physical infrastructure and the development of traffic regulations, such as a speed limit. This analogy from the history of cars not only teaches us that the

process of embedding a system technology is ever-evolving, but also that the role of the state in this process cannot be reduced to that of a single authority or ministerial department. As a system technology potentially affects every domain and requires efforts on all five fronts of the process of embedding it, the full apparatus of the state is deeply involved. This is why we see that the introduction of new system technologies was historically accompanied by the development of a new 'policy infrastructure'.

The WRR expects that the same process will be necessary for embedding AI. Here too, the responsibility cannot be assigned to a single ministerial department or authority due to its system technological character, and, hence, the state will have to do more than merely developing specific instruments to push AI in the desired direction. Rather, it will have to work on developing an institutional or *systemic* answer to the new system technology that AI is. When extrapolating the historical patterns of prior system technologies, the WRR deems it necessary for governments to work on a new policy infrastructure and therefore recommends the following:

11. Establish a policymaking infrastructure for AI, starting with an AI coordination centre that is embedded into the political process.

With this recommendation, the WRR aims to structurally integrate the gathering and sharing of knowledge into the state's AI-related activities. At this stage, it remains unknown how AI will evolve in society and how society will evolve alongside. By structurally bringing together new insights, governments can boost their learning capacity and develop insights into the more fundamental issues relating to the organization of society that emerge as a result of the growing use of AI.

In Europe, there is an additional factor that requires the development of a policy infrastructure. The EU's draft Artificial Intelligence Act states that every Member State has to designate one or more national authorities to supervise the application and implementation of the regulation. Moreover, Member States will have to appoint a national supervisory authority that functions as an official point of contact for the public and other societal actors. Therefore, European governments should at least for this reason work on developing a policy infrastructure for AI.

Globally, there are preliminary signs that such policy infrastructures are starting to be developed with respect to AI. Indeed, several governments are currently sorting out how they can institutionalize their intelligence on AI,

in a variety of forms. Although most of these initiatives began as temporary committees, such as the European Commission's High-Level Expert Group on AI and the Ad Hoc Committee on AI (Council of Europe), governments are now looking for permanent ways to embed AI institutionally. There are examples of countries that have established a specific ministerial department to cover AI-related issues, such as the Ministry of Artificial Intelligence in the United Arab Emirates, the former Ministry for Digital Affairs in Poland and the Ministry of Technological Innovation and Digital Transition in Italy. An alternative approach to this is structural cooperation between existing ministerial departments and governmental organizations, such as the inter-ministerial office for AI in the UK, the interdepartmental working group on AI in the Netherlands, and the National Initiative on Robotics in the US. Other forms focus on the institutionalization of expertise, for example, by establishing an AI Council (UK, Spain), a Digital Council (Germany, New Zealand), or even an international Council on AI (as proposed by Colombia). Alternative options are permanent advisory boards on AI, which Austria and Singapore have set up, or AI taskforces like those launched in Kenya and India. Previously, the European Parliament also suggested that it would consider establishing a European agency for robotics and AI, "to provide technical, ethical and regulatory expertise on AI needed to support the relevant public actors, at both Union and Member State level, in their efforts to ensure a timely, ethical and well-informed response to the new opportunities and challenges."[52]

These efforts to embed AI institutionally show that we are still in the explorative phase of developing a new policy infrastructure. Although there is no blueprint for an AI policy infrastructure, the WRR argues that the first step for a country like the Netherlands is to establish an *AI coordination centre*. This centre can help to develop expertise about AI in society, consult with governments to identify and address relevant issues, and play a coordinating role in developing an integrative view of precisely what an AI-savvy society should look like. By giving a structural character to the way states learn about AI, the coordination centre would provide a solid base for political debate. In the future, both the experience and activities of the coordination centre could prove to be an invaluable starting point for policy development and implementation. An AI coordination centre could thus serve as a hub for knowledge, a platform to align state activities and a partner in policy-making.

First and foremost, an AI coordination centre is relevant at a national level, insofar as it enables governments to develop an integrative approach to the process of embedding AI within society. However, the same function could

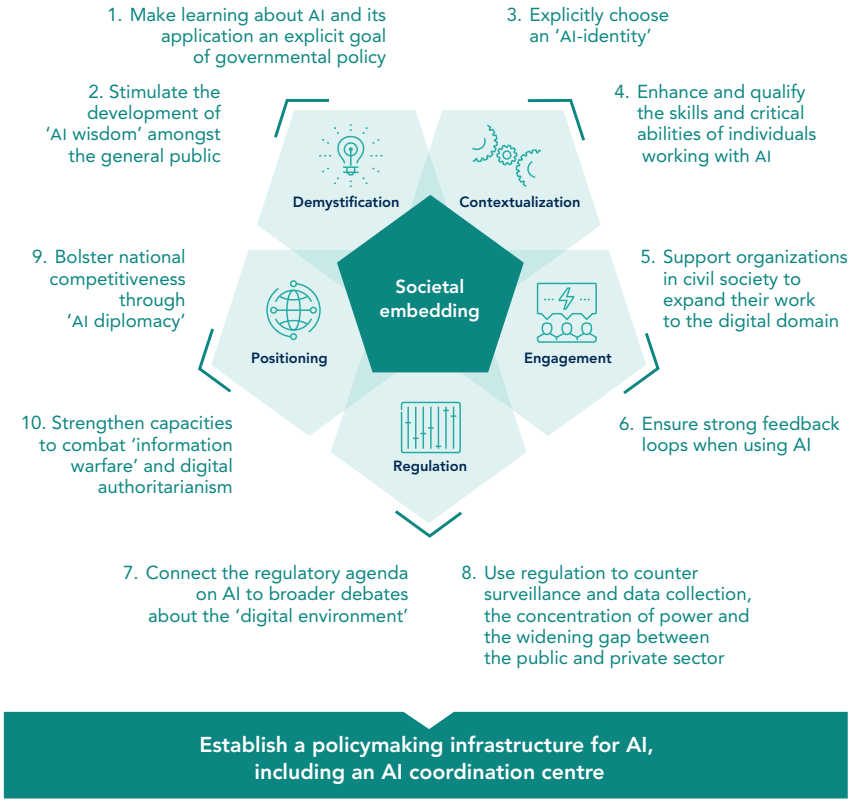
52 European Parliament, 16 February 2017.

also work at an international level in the context of international cooperation. Here too, a coordination centre could facilitate the sharing of knowledge between countries, connect their AI activities, and aid in the development of international policy.

Given that the coordination centre could help to shed light on the systemic impact of AI and the fundamental governance questions that it brings about, the centre could also play an important role in terms of both setting the political agenda and enhancing the quality of political debate. The WRR argues that, given the fundamental role that digital technologies already play and AI is expected to play in our modern society, digitalization should be considered a more pressing political issue than it currently is. This not only applies to the Netherlands, but to many other countries as well. The WRR, therefore, recommends to meaningfully position the AI coordination centre within the political arena, debate and decision-making process, thereby consolidating its function at that level as well.

Depending on states' organization, different institutional forms could be considered. In the Netherlands, the WRR proposes to set up a cabinet committee on digital affairs, which would also encompass AI-related issues. This committee, which would consist of a subset of ministers, would structurally address issues relating to digital technologies that demand an integrative approach. Alongside this, comparable levels of political integration in other countries could ensure that the expertise of the coordination centre translates into international political debate about how to embed AI within our societies.

Figure 9 An overview of the recommendations for governments to address the five tasks of embedding AI in society



Conclusion

Over the last several years, AI has moved beyond the confines of the lab to proliferate amongst society at large. This brings with it a mission to effectively embed this new technology within society. There is an extensive body of literature on AI and its societal implications. The WRR sought to add to this literature by conducting a systematic investigation into the type of technology that AI belongs to. We conclude that it can best be understood as a *system technology*. The history of system technologies teaches us that their impact is complex, unpredictable and that it unfolds over many decades. Moreover, it also shows us that embedding this kind of technologies in society requires a range of processes to incrementally adjust technology and society to one another.

In order to effectively embed a system technology, we identify five overarching tasks for society: *demystification*, *contextualization*, *engagement*, *regulation* and *positioning*. These tasks all require efforts from a wide range of actors, such as businesses, civil society, science and individual citizens. Governments in particular have a key role to play, and, once again, history shows that this role increases over time. To aid governments in this process, we provide recommendations to take up the five tasks of embedding AI within society. As mentioned earlier, these recommendations were initially addressed to the Dutch government. However, our general framework of system technologies, the five tasks of embedding AI in society and our analysis of how to take up these tasks, can be of value to governments across the globe. In this way, the WRR aims to contribute to the exciting mission of embedding AI, the 'combustion engine of the twenty- first century'.

Appendix

Textbox – A very brief history of AI

Prior to entering the lab, AI went through three different phases. Myths of artificial forms of intelligence have existed for many centuries. For example, Greek mythology spoke of the robot warrior Talos (after which an American army suit is named), the mechanical helpers of the smith god Hephaestos, Galata, the statue that turned to life as well as the many machines designed by the human engineer Daedalos. Estonian myths refer to the Kratt, after which an AI-related law in the country is named, while similar concepts can be found in ancient India and China.

With the advent of the Enlightenment, a next phase of speculation on AI emerged. As calculators began to be built, people came to imagine all kinds of automata, that, in practice, were impossible to realize. For a while, people believed that an autonomous chess machine had actually been developed, the so-called mechanical Turk, but it was subsequently revealed that there was a human lurking inside the machine.

From the nineteenth century onwards, a third phase was set into motion as people developed the theoretical antecedents for computers and AI. Ada Lovelace and Charles Babbage made notable contributions to the field, as did researchers in World War II, who were working on ballistic trajectories and code breaking. The most notable of these figures was Alan Turing, of course, who is considered the father of both the computer and the field of AI.

The field of AI was launched with the Dartmouth Summer Research Project on Artificial Intelligence in 1956, where the world's leading scientists in the field gathered. Since then, AI entered the lab and has developed through three so-called waves, two of which have been followed by so-called AI winters, periods of reduced funding of and interest in AI research.

In the first wave, all kinds of algorithms were developed to solve or prove things in the fields of mathematics and logic. The first winter set in over the course of the 1970s as the initial high hopes were not realized, which led many governments to scale back their funding.

The second wave began in the 1980s, boosted by the rise of Japan and the attendant fear that its investments in the field would give it a competitive edge. Governments in the US and Europe also increased their funding for the field. This wave also brought the rise of expert systems, dubbed the first truly commercial application of AI. These were systems in which the knowledge of specific experts like doctors or scientists were coded into algorithms to assist them with their work. This wave also ended in disappointment as it failed to live up to expectations.

Since its inception, the field of AI has been distinguished into two broad approaches (although there are more). The first of these is called symbolic AI or logical systems, and deals with pre-coded rules of the character if X, then Y. This approach dominated the field from the outset, with the aforementioned mathematical programs and expert systems being the classic examples of this approach. The second approach is known as connectionism, neural networks or machine learning, and does not follow pre-coded rules. Instead, an algorithm is fed data from which it derives patterns.

It is this second approach that strongly drives the current, third wave of AI. There are three factors underlying its development. Firstly, the expansion in computing power (Moore's Law) made it possible to perform complex calculations of large amounts of data. Secondly, vast amounts of data became available due to the rise of the internet as well as the pictures, messages, and transactions that people placed online and on which algorithms could subsequently be trained. Finally, scientific breakthroughs occurred that made it possible to discern patterns in different layers of the data. For example, pictures of a human face could not merely be broken up into noses and eyes, but rather also into segments of them, edges and curves, all the way down to individual pixels. By learning from these deeper layers (which is why it is called 'deep learning') and assigning value to them, scientists paved the way for current applications of AI.

Bibliography

Agrawal, A., Gans, J., and Goldfarb, A. (2019) *The Economics of Artificial Intelligence: An Agenda*, National Bureau of Economic Research, Chicago: University of Chicago Press.

Bakker, S. and P. Korsten (2021) *Artificiële Intelligentie als een general purpose technology: Strategische belangen en verantwoorde inzet in historisch perspectief*, WRR Working Paper nr. 41, Den Haag: Wetenschappelijke Raad voor het Regeringsbeleid. Retrieved from: https://www.wrr.nl/publicaties/working-papers/2021/02/16/artificiele-intelligentie-als-een-general-purpose-technology

Benjamin, R. (2019) *The Race After Technology: Abolitionist Tools For the New Jim Code*, Cambridge: Polity Press.

Bennett Moses, L. (2007) 'Recuring Dilemmas: The Law's Race to Keep Up With Technological Change', *University of Illinois Journal of Law, Technology and Policy*, Vol. Fall: 239-285.

Bradford, A. (2020) *The Brussels Effect: How the European Union Rules the World*, Oxford: Oxford University Press.

Bresnahan, T. and Trajtenberg M. (1995) 'General Purpose Technologies 'Engines of Growth'?', *Journal of Econometrics*, 65(1): 83-108.

Broussard, M. (2019) *Artificial Unintelligence: How Computers Misunderstand the World*, Cambridge: MIT Press.

Bughin, J., J. Seong, J. Manyika, M. Chui en R. Joshi (2018) *Notes From the AI Frontier: Modeling the Impact of AI on the World Economy*, McKinsey Global Institute. Beschikbaar op: https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20frontier%20Modeling%20the%20impact%20of%20AI%20on%20the%20world%20economy/MGI-Notes-from-the-AI-frontier-Modeling-the-impact-of-AI-on-the-world-economy-September-2018.

CB Insights (2021) *Despite A Pandemic Slump, The AI Sector Remains Hot For Acquirers*, CB Insights Research Briefs. Retrieved from: https://www.cbinsights.com/research/artificial-acquisitionstrends-annual-deals/

Crawford, K. (2021) *The Atlas of AI*, New Haven: Yale University Press.

Danaher, J. (2016) 'The Threat of Algocracy: Reality, Resistance and Accommodation', *Philosophy & Technology*, 29(3): 245-268.

Ding, J. (2018) *Deciphering China's AI Dream*, Future of Humanity Institute Technical Report, Oxford: University of Oxford.

Domingos, P. (2017) *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, London: Penguin Random House.

Edgerton, D. (2008) *The Shock of the Old: Technology and global history since 1900*, London: Profile books.

Eubanks, V. (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, New York: St. Martin's Press.

European Commission (2021a) *Regulation Of the European Parliament and of the Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM(2021) 206 Final. Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

European Commission (2021b) *The Strategic Use of Public Procurement For Innovation In the Digital Economy: Final Report*, Luxembourg: Publications Office of the European Union. Retrieved from: https://op.europa.eu/o/opportal-service/download-handler?identifier=7f5a67ae-8b8e-11eb-b85c-01aa75ed71a1&format=pdf&language=en&productionSystem=cellar&part=

European Parliament (2017) *European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics* (2015/2103(INL)). Retrieved from: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html

Floridi, L. (2014) *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*, Oxford: Oxford University Press.

Freeman, S. (2001) 'Illiberal Libertarians', *Philosophy & Public Affairs*, 30(2): 105-151.

Freeman, C. and F. Louçã (2001) *As Time Goes By: From the Industrial Revolutions to the Information Revolution*, Oxford: Oxford University Press.

Frey, C.B. and Osborne, M.A. (2013) *The Future of Employment: How Susceptible Are Jobs To Computerisation?* Oxford: Oxford Martin Programme on Technology and Employment.

Goode, L. (2018) *Google CEO Says AI Will Be More Important To Humanity Than Electricity or Fire*, The Verge, 19 January 2018. Retrieved from: https://www.theverge.com/2018/1/19/16911354/google-ceo-sundar-pichai-ai-artificialintelligence-fire-electricity-jobs-cancer

Gordon, R. (2016) *The Rise and Fall of American Growth: The U.S. Standard of Living Since the Civil War*, Princeton: Princeton University Press.

Greenfield, A. (2017) *Radical Technologies: The Design of Everyday Life*, New York: Verso Books.

Horowitz, M., Allen, G., Kania, E. and Scharre, P. (2018) *Strategic Competition in an Era of Artificial Intelligence*. Washington: Center For a New American Security. Retrieved from: https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNAS-Strategic-Competition-in-an-Era-of-AIJuly-2018_v2.pdf?mtime=20180716122000enfocal=none

Juma, C. (2016) *Innovation and Its Enemies: Why People Resist New Technologies*, Oxford: Oxford University Press.

Keane, J. (2009) *Life and Death of Democracy*, Toronto: Simon & Schuster.

Lee, K.F. (2018) *AI Superpowers: China, Silicon Valley, and the New World Order*, Boston: Houghton Mifflin Harcourt.

Leung, J. (2019) *Who Will Govern Artificial Intelligence? Learning From the History of Strategic Politics In Emerging Technologies*, Oxford: Oxford University. Retrieved from: https://ora.ox.ac.uk/objects/uuid:ea3c7cb8-2464-45f1-a47c-c7b568f27665

Libicki, M. (2019) 'A Hacker Way of Warfare', in N. Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, 137-142, Montgomery: Air University Press.

Lin, H. (2019) 'Escalation Risk in an Artificial Intelligence-Infused World', 143-152 in N. Wright (red.) Artificial Intelligence, China, Russia and the Global Order, Montgomery: Air University Press.

Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, 143-152, Montgomery: Air University Press.

Loucks, J., S. Hupfer, D. Jarvis en T. Murphy (2019) *Future in the balance? How countries are pursuing an AI advantage*, Deloitte Center for Technology, Media & Telecommunications. Retrieved from: https://www2.deloitte.com/content/dam/Deloitte/lu/Documents/public-sector/lu-global-ai-survey.pdf

Luttwak, E. (1990) 'From Geopolitics To Geo-economics: Logic of Conflict, Grammar of Commerce', *The National Interest*, 20: 17-23.

Lynch, S. (2021) *Andrew Ng: Why AI Is the New Electricity*, Stanford Graduate School of Business, 4 May 2021. Retrieved from: https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity

McKinsey & Company (2020) *How nine digital front-runners can lead on AI in Europe*, McKinsey & Company. Retrieved from: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/how%20nine%20digital%20frontrunners%20can%20lead%20on%20ai%20in%20europe/how-nine-digital-frontrunners-can-lead-on-ai-in-europe.pdf

Morgus, R. (2019) 'The Spread of Russia's Digital Authoritarianism', in N. Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, 89-97, Montgomery: Air University Press.

Morozov, E. (2013) *To Save Everything, Click Here: The Folly of Technological Solutionism*, London: Penguin.

O'Neil, C. (2016) *Weapons of Math destruction: How Big Data Increases Inequality and Threatens Democracy*, London: Penguin.

Pasquale, F. (2020) *New Laws of Robotics: Defending Human Expertise in the Age of AI*, Cambridge: Harvard University Press.

Rao, A. en G. Verweij (2017) *Sizing the Prize: What's the Real Value of AI for Your Business and How Can You Capitalise?*, PricewaterhouseCoopers. Retrieved from: https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

Scharre, P. (2018) *Army of None: Autonomous Weapons and the Future of War*, New York: WW Norton & Company.

Schick, N. (2020) *Deep Fakes and the Infocalypse: What You Urgently Need to Know*, London: Octopus Publishing Group.

Scholvin, S. and Wigell, M. (2018) *Geo-economics As a Concept and Practice In International Relations: Surveying the State of the Art*, Working Paper nr. 102, Helsinki: Finnish Institute of International Affairs.

Singer, P. and Brooking, E. (2018) *LikeWar: The Weaponization of Social Media*, Boston: Mariner Books.

Smuha, N. (2019) 'From a 'Race To AI To a 'Race To AI Regulation': Regulatory Competition For Artificial Intelligence', *Law, Innovation and Technology*, 13(1): 57-84.

Taplin, J. (2017) *Move Fast and Break Things: How Facebook, Google, and Amazon Have Cornered Culture and What It Means For All of Us*, New York: Pan Macmillan.

Tonin, M. (2019) 'Artificial Intelligence: Implications for NATO's Armed Forces', *149 STCTTS 19 E rev. 1 fin.*

Vleuten, E. van der, R. Oldenziel en M. Davids (2017) *Engineering the future, understanding the past: A social history of technology*, Amsterdam: Amsterdam University Press.

Warzel, C. (2018) *Believable: The Terrifying Future of Fake News*, Buzzfeed News, 11 February 2018. Retrieved from: buzzfeednews.com/article/charliewarzel/the-terrifying-future-of-fake-news

WIPO (2019) *WIPO Technology Trends 2019: Artificial Intelligence*, Geneva: World Intellectual Property Organization.

Wright, N. (2019) 'Global Competition', in N. Wright (red.) *Artificial Intelligence, China, Russia and the Global Order*, 35-41, Montgomery: Air University Press.

WRR (2019) *Voorbereiden op Digitale Ontwrichting*, The Hague: The Netherlands Scientific Council for Government Policy (the WRR).

WRR (2020) *Het Betere Werk. De Nieuwe Maatschappelijke Opdracht*, The Hague: The Netherlands Scientific Council for Government Policy (the WRR).

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight For a Human Future At the New Frontier of Power*, London: Profile books.

Mission AI

The New System Technology

Artificial intelligence (AI) is the combustion engine of the twenty-first century. The technology is currently moving out of the lab and into society, which raises the issue of its impact upon public values. In the publication *Mission AI – The New System Technology*, the Netherlands Scientific Council for Government Policy (WRR) offers a new perspective on this theme. AI can best be compared with the steam engine, electricity, the combustion engine and the computer. Such "system technologies" are ubiquitous, can be used for all kinds of purposes and change the economy and society in profound and unpredictable ways. We are currently at a turning point: AI needs to be embedded within society. The government in particular faces several tasks in this respect, including tackling unrealistic images of AI (demystification), creating a good environment for it to work in (contextualization), involving societal actors (engagement), drafting a broad regulatory agenda for AI (regulation) and reflecting on the Netherlands' relationship with international parties in this domain (positioning).

WRR

THE NETHERLANDS SCIENTIFIC COUNCIL FOR GOVERNMENT POLICY