Artificial Intelligence, real consequences
How AI is changing the way we live

Goncalves, J. E.

DOI
10.59490/mg.70

Publication date
2023

Document Version
Final published version

Citation (APA)
Goncalves, J. E. (2023). *Artificial Intelligence, real consequences: How AI is changing the way we live*. TU Delft OPEN Publishing. https://doi.org/10.59490/mg.70

AI*

ARTIFICIAL INTELLIGENCE REAL CONSEQUENCES

How AI is changing the way we live

Juliana E. Gonçalves

TU Delft Delft University of Technology

TU Delft OPEN

AI*

ARTIFICIAL INTELLIGENCE, REAL CONSEQUENCES
How AI is changing the way we live

Author:
Juliana E. Goncalves
Faculty of Architecture and the Built Environment, TU Delft, ORCID 0000-0001-6888-3362

Keywords:
Artificial Intelligence, Digitalisation, Computer Science, Ethics of Technology, Societal Impact

Published by:
TU Delft OPEN Publishing | Delft University of Technology, The Netherlands
ISBN/EAN: 978-94-6366-714-2
DOI: https://doi.org/10.59490/mg.70

Design and layout by Roberto Rocco.

Copyright clearance made by the TU Delft Library copyright team.

Foreword

Perhaps the first images that come to mind when thinking about artificial intelligence (AI) are machines and robots that look and behave like humans. Machines that are self-aware, conscious, and autonomous, just like us. This version of AI is undoubtedly the one most exploited in sci-fi movies and books: The "grand dream" of creating machines that have cognitive capacities as humans do. However, many scientists and critics challenge this grand dream. First, many believe such a dream is too grand to be feasible. Second, many argue that pursuing this grand dream is not desirable at all. Two questions then arise: What is feasible in AI? What is desirable for AI?

Although these questions remain open, AI is increasingly becoming more influential in our society. This influence may seem unnoticeable for most people, but it has profound societal consequences. Opportunities for AI are promising, but threats are looming.

On the one hand, AI brings with it a fairly untapped potential for positive societal impact. AI may help to ensure that operations run more efficiently, improving services for people and reducing operational costs. For example, AI has been an important tool in several efforts to tackle the impacts of the COVID-19 pandemic. Among other tasks, AI has been used in outbreak prediction and tracking, contact tracing, early diagnosis, and curbing the spread of misinformation.

On the other hand, the use of AI systems raises various ethical issues, particularly related to bias, discrimination, privacy, transparency and accountability. There are serious concerns that AI applications violate human rights and exacerbate existing structural inequalities and discrimination. Discrimination associated with face recognition systems is perhaps the most contentious issue around AI. Like many once-promising forms of new technology, AI also becomes less appealing when we look at the negative impacts caused by such systems, particularly when applied with negligence or bad intentions.

Despite the potential to facilitate several tasks and improve the collective well-being of humans, there are still many open questions about social, legal, ethical, and democratic impacts of AI that have to be addressed. As more complex AI methods are developed, uncertainty around these questions grows and more action is needed if we want to safely employ AI within society.

Understanding these dimensions of AI is essential to shape our future society with AI. This book clarifies AI definitions, describes AI methods, provides examples of AI applications, and highlights limitations and impacts of AI. This is not an academic publication; rather, it aims to inform and create awareness about the uses of AI, highlighting opportunities and exposing the threats associated with AI.

Motivation & Process

This project started within the Junior Council (JC) of the TPM AI Lab. The TPM AI Lab was established in 2021 at the Faculty of Technology, Policy and Management (TPM) of TU Delft to consolidate research on AI within and outside TU Delft.

At the TPM AI Lab, we believe the noise around AI can be overwhelming. It is for us, and we believe it is for others too. Recent developments in cutting-edge technology have fueled a new wave of interest in AI research and application, spanning multiple sectors and disciplines. Countless AI techniques, approaches, and tools are being employed or developed for various purposes. Everybody wants to see what AI has to offer. It is difficult to keep track of everything happening in the AI world. At the same time, the enthusiasm for AI is met with scepticism and scrutiny by critical scholars and experts.

To ensure that the book includes multiple angles on AI, we have reached out to several colleagues from the Faculty of Technology, Policy and Management (TPM) of TU Delft, working on AI across the different departments of the faculty. From this group of AI experts, we were able to gather insights from the perspectives of Engineering Systems and Services, Multi-Actor Systems, and Values, Technology and Innovation. Based on their input, relevant topics in current AI research were selected, delineating the scope of this book. From each colleague, we also collected research statements about AI. Insightful statements of those who contributed are presented in the last chapter. Last but not least, at the end of the book, a bonus: a must-read list for those who want to know more about AI.

Acknowledgements

I started this initiative during my postdoc at the Faculty of Technology, Policy and Management of TU Delft. I was then a member of the TPM AI Lab. This book project is close to my heart for two reasons. First, it aims at a broad audience, going beyond academia. Second, it has been published as a free open resource with the support of the TU Delft OPEN Textbooks. With this, I hope to take the research developments outside of TU Delft in the most accessible way, hopefully reaching people from diverse backgrounds.

A lot has happened between the start of the project and the publication of this book. Among others, I terminated my postdoc at TPM and membership at the TPM AI Lab earlier than expected to start a new position as an Assistant Professor at the Faculty of Architecture and the Built Environment, still at TU Delft. Although I moved to a building a few meters away from my previous location, the new position came with many other responsibilities and changing priorities. But, I am a very committed person and thus wanted to finish the project started at the TPM AI Lab. Not least because I deeply value the input and the time of many colleagues who contributed to this project. I wanted to give this project the end it deserves.

Initially, the project idea was to collect 'fact sheets' from TPM scholars and experts working on AI topics and publish them as a collection of sheets. However, after pitching the project to the AI experts, many of them challenged the idea of a 'fact' and highlighted the danger of publishing a list of so-called facts without any contextualisation. I reflected on their feedback and decided to reformulate the project. The result is the book you have in your hands.

Today, I am happy to see this project finished. I can finally thank all those involved for their priceless input and endless patience. Most importantly, this project would not have been possible without the contribution of the AI experts from the Faculty of Technology, Policy and Management of TU Delft, who shared their research insights and perspectives on AI. A huge thanks to each of you: Prof.dr.ir. Alexander Verbraeck, Emile J. L. Chappin, Prof.dr.ir. Ibo van de Poel, Jonne Maas, Prof.dr. Martijn Warnier, Nihit Goyal, Olya Kudina, Sander van Cranenburgh, Stefan Buijsman, Shuhong Li, and Wiebke Hutiri.

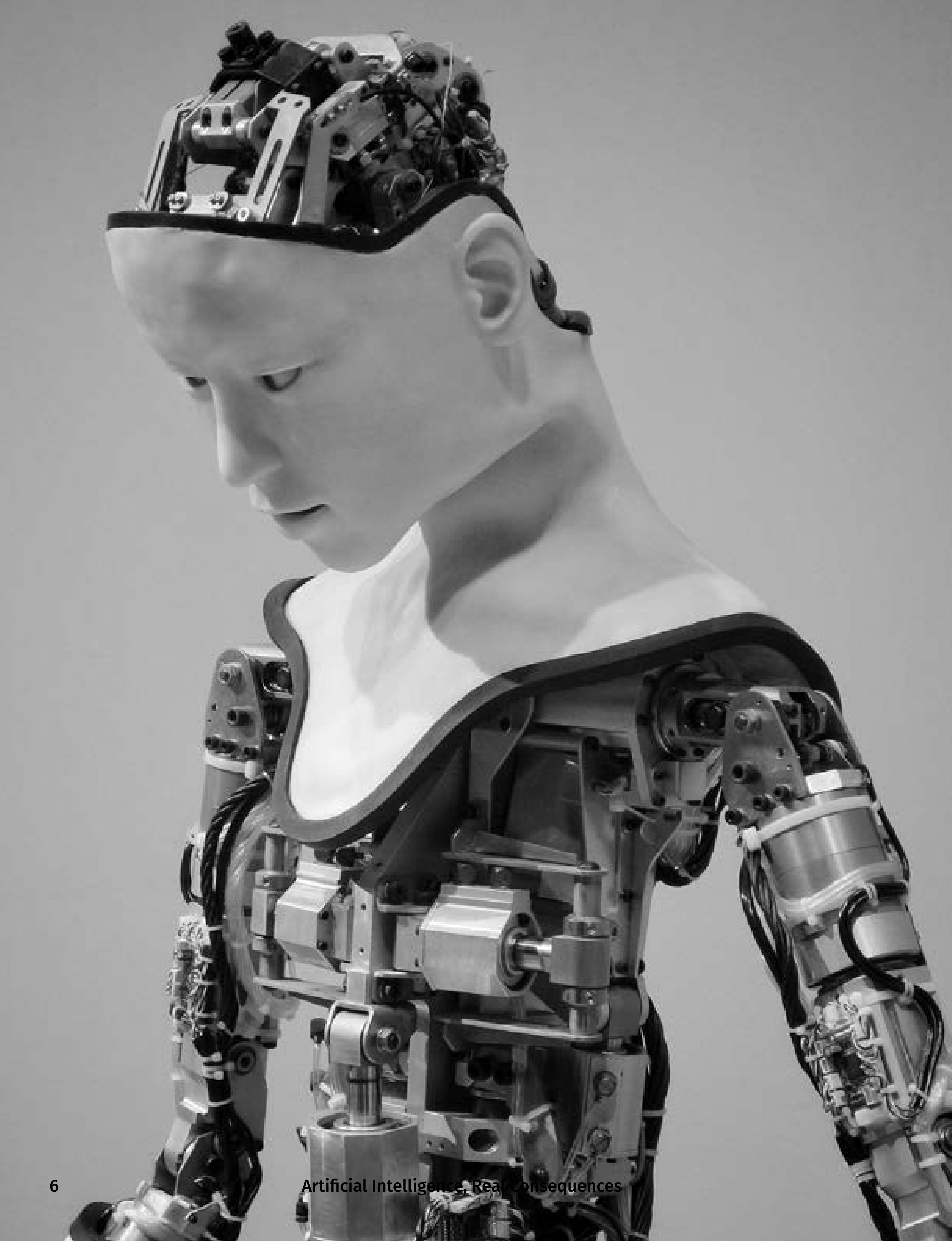
The collaboration within the TPM AI Lab was also crucial for the completion of this project. I would like to thank all the members of the lab for their contributions and support throughout the project: Aishwarya Suresh Iyer, Amineh Ghorbani, Amir Pooyan Afghari, Felicitas Reddel, Filippo Santoni De Sio, Francisco Garrido Valenzuela, Khushboo Sharma, Molood Aleebrahimdehkordi, Seda Gürses, Tanvi Patil, Trivik Verma, and Vanisha Jaggi.

A special thanks also goes to Roberto Rocco who kindly designed the cover and template of this book and to Trivik Verma who reviewed the text, providing insightful suggestions to improve the language style. I also would like to acknowledge the important contribution of Chiem Kraaijvanger and Mrinal Patil, who not only enthusiastically joined this initiative but also provided invaluable input by helping me to review the literature and collect interesting examples of AI applications. Finally, a big thanks goes to my Twitter friends who provided recommendations of books for the must-read list at the end of the book. Without all these people, this project would not have come to a successful ending.

Juliana E. Goncalves
Delft, 2023

Contents

Chapter 1 aims to answer the challenging question, “What is AI”. Next, Chapter 2 describes how AI works and provides a small AI glossary helpful to navigate this book and the AI field. Chapter 3 describes AI applications across multiple sectors to illustrate the potential of positive impact of AI. Chapter 4 sheds light on issues related to threats and negative implications of AI. Finally, Chapter 5 gives the word to the TPM AI experts with a selection of insightful research statements. At the very end of this book, a must-read list is provided for those who are interested in learning more about how AI is changing the way we live.



Chapter 1

WHAT IS AI?

What is AI?

Historical accounts of AI usually bring us back to the famous scientist Alan Turing. Turing (1912 - 1954) was an English mathematician often cited as the father of computer science and artificial intelligence. Among his many contributions to these fields, in 1950, Turing published a landmark research article in the British journal, Mind. In the article "Computing Machinery and Intelligence", Turing describes a test to verify whether a machine is intelligent. The test consists of a parlour game in which a man and a woman are in two separate rooms and communicate with an interrogator via chat. The interrogator may ask any question that can be communicated via chat to identify the man and the woman. The man tries to convince the interrogator that he is the woman, and the woman tries to convey her real identity. At some point, the man is replaced by a machine. The machine "passes the test" if the interrogator cannot distinguish the machine from the woman. Turing named his test the Imitation Game.

Turing's article has generated a lot of commentary and controversy throughout the years and remains relevant in the field. Today, the Imitation Game is called the Turing Test. It is conducted differently: a single room with either a person or a machine, and the interrogator determines whether they are communicating with a real person or a machine. Turing's article is indeed an important marker in the history of AI, but only possible thanks to previous developments. Arguably, AI has developed building upon previous work both in the field of computer science and in philosophic work reflecting on how humans think, bringing us back to names like Ada Lovelace, the first computer programmer, and Leibniz, Spinoza, Hobbes, Locke, Kant and Hume, famous Western philosophers.

When Turing published "Computing Machinery and Intelligence", the term "artificial intelligence" did not exist yet. It was only in 1956 that a group of American computer scientists coined the term for a conference organised at Dartmouth College. The conference is considered the "official birthday of artificial intelligence" and reunited scientists and students working in the field, who became the founders and leaders of AI research. It was at this conference that the first artificial intelligence computer program was presented to the scientific community by Allen Newell, Herbert A. Simon, and Cliff Shaw. From this point, the field received much attention and funding. Many important developments occurred between 1950-1970. However, progress slowed in the 1970s, when a series of disappointments in AI led to cutbacks in government funding. The generally held feeling was that AI did not deliver on the expectations around it and, the following years became a difficult period to obtain AI-related funding. This period marks the first AI winter.[1]

We can say that AI is a field of ups and downs, with both winters and summers. We are currently experiencing an AI summer, with major breakthroughs in the field and, again, a lot of expectations about future AI developments. However, some believe that another winter is just around the corner. Others argue that AI winters happen because we expect too much from AI: the higher the fly, the harder the fall. The hopes for AI have always been high, right from the beginning, as Turing himself believed computers would have consciousness by 2000. It is 2023, and we have clearly disappointed Turing. Further, AI experts believe we may never achieve Turing's expectations simply because it may not be possible to create machines that think. The Dutch computer scientist Edsger W. Dijkstra wrote, "The question of whether a computer can think is no more interesting than the question of whether a submarine can swim". The British-born American writer, Pamela McCorduck, had similar lines, comparing the question of "Can a machine think?" to whether a chicken has lips. In addition, critics claim that trying to create machines that think should not be AI's goal at all. And yet, this has not stopped efforts to create human-like machines[1].

Media often portrays AI as the "grand dream" - the idea of creating machines that can think and behave like humans - sometimes depicted as a utopia and other times as a dystopia, but always with strokes of exaggeration. These representations bring an unrealistic picture of both strengths and flaws of AI that do not match reality at all. In fact, many of the current applications of AI are based on techniques developed in the 20th century but are only now being implemented thanks to the increasing availability of data and computational power. Notwithstanding the media exaggeration and misrepresentation of AI, there are high hopes for AI and huge concerns about a future with AI.

If AI is not what we see in the media and movies, what is AI? Well, AI is notably difficult to define. Even researchers and engineers working on AI for many years have not agreed on a unique definition. AI encompasses intelligent systems and machines that emulate human thinking and behaviour and perform tasks that would otherwise require human intelligence. More officially, the European Union describes AI as "systems that display intelligent behaviour by analysing their environment and taking actions - with some degree of autonomy - to achieve specific goals"[2]. Under this definition, AI applications are multiple. They can be simple and already influencing our daily routines, like AI-powered systems that block spam. They can also be rather sophisticated, like AI-powered image analysis to support cancer diagnosis. Furthermore, AI can be fully integrated into machines in the form of advanced robots and autonomous cars, or operate mostly in the virtual world, like voice assistance, search engines, and speech and face recognition systems. Sometimes AI is equated with smartness, but a smart device does not necessarily have AI.

1 McCorduck, P., & Cfe, C. (2004). Machines who think: A personal inquiry into the history and prospects of artificial intelligence. CRC Press.

2 European Commission. (2018). Coordinated Plan on Artificial Intelligence.

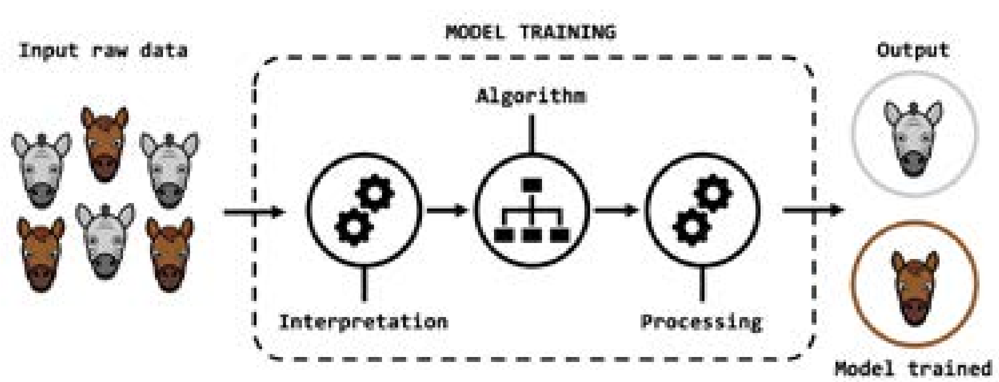
Chapter 2

HOW DOES AI WORK?

Generally speaking, the learning process of an AI system or machine is no different from a child being taught about the world. For example, children are only able to differentiate between a horse and a zebra once they know that a zebra has black and white patterns and a horse does not. Similarly, an AI system needs to learn what a horse is and what a zebra is to differentiate between them.

AI is an umbrella term for techniques computers use to perform multiple tasks, including cognitive tasks (such as learning or self-correction), which in the past have only been performed by humans. Many techniques and technologies belong to the AI domain, including machine learning, deep learning, natural language processing, among others. Most of these techniques work by analysing (huge amounts of) data to identify patterns and learn about the world. Generally speaking, the learning process of an AI system or machine is no different from a child being taught about the world. For example, children are only able to differentiate between a horse and a zebra once they know that a zebra has black and white patterns, and a horse does not. Similarly, an AI system needs to learn what a horse is and what a zebra is to differentiate between them. While humans are good at extrapolating and reasoning, learning from a relatively small dataset, AI systems require much more data to learn the same amount, or arguably less. Another difference is that AI can process highly complex data in a short amount of time, a task impossible for humans.

A fundamental step in developing AI systems is thus teaching the system to understand the environment in which it will operate so it can correctly perceive information and make accurate decisions based on the information provided. This is called training and requires training data. The training process teaches AI systems to differentiate a zebra from a horse, a bus from a car, or an intense cry from joyous laughter. In the first two cases, the training data has to contain multiple pictures of zebras, horses, buses, and cars and, in the third case, multiple voice samples of people crying and laughing, all needing humans to label the trained samples as horses or zebras, and so on. The quality of the training data highly influences the accuracy of the AI system. For example, if the training data only include front pictures of zebras and horses, the AI system will not be able to recognise side or back pictures. The same happens for voice samples of people yelling and laughing. Even humans sometimes have a hard time saying if some people are laughing or crying. If the training data is suitable, the AI system can be used for the task it has been trained for. Without training, an AI system will be, at best, pointless or inefficient and, at worst flawed and dangerous.



The training process teaches AI systems to perceive information and make decisions about it.

To differentiate between zebras and horses, the training data contains images of zebras and horses, properly labelled by humans.

After training, the AI system can be used to identify front pictures of zebras and horses.

A fundamental step in developing AI systems is thus teaching the system to understand the environment in which it will operate so it can correctly perceive information and make accurate decisions based on the information provided. This is called training and requires training data. The training process teaches AI systems to differentiate a zebra from a horse, a bus from a car, or an intense cry from joyous.

A small glossary

Advances in AI have brought a lot of new words in our daily life. A whole new vocabulary includes terms like algorithms and machine learning. Understanding AI jargon is key to facilitating discussion about the real-world applications of this technology. Before going deeper into the complex world of AI, we provide a small glossary below.

Algorithm

An algorithm is a series of step-by-step computer operations that can solve a problem in a finite number of steps. Algorithms are at the core of all AI systems.

Machine Learning

Machine Learning (ML) is a subfield of AI, although the terms are often used interchangeably. ML enables machines to learn from patterns in the data without being explicitly programmed. ML is often based on simpler mathematical methods, like linear regression. ML is everywhere: identifying spam in our mailboxes, providing song and movie recommendations in streaming services, and giving us warnings about suspicious transactions in our bank accounts.

Deep Learning

Deep Learning is a specific type of ML that allows AI to learn and improve by processing highly complex data. Deep Learning also works by identifying patterns in the data without being explicitly programmed for it. The difference is that Deep Learning uses artificial neural networks that mimic biological neural networks in the human brain to process information. Deep Learning systems are more complex and require less human intervention; higher complexity and less intervention means more data is required for training. Deep Learning is behind more complex and autonomous technologies, like self-driving cars or robots that perform advanced medical surgery.

Artificial Neural Networks

Artificial Neural Networks follow the operation principles of biological networks of neurons in the human brain, allowing AI systems to process large and complex datasets.

Natural Language Processing

Natural language processing (NLP) is also a type of ML. NLP allows computers to recognise, analyse, and interpret written or spoken languages. NLP is at the core of technologies that interact with humans in some way, either via text or audio inputs, including the famous virtual assistants Siri and Alexa and business chatbots.

Computer Vision

Computer Vision is one of the most prolific uses of AI technologies. It combines pattern recognition and deep learning to review and interpret the content of an image. Computer Vision is behind daily applications like face recognition and advanced image analysis in cancer treatment.

Cognitive Computing

Cognitive Computing refers to AI systems deployed to imitate the human thought process by simulating the human cognition process. To achieve this ambitious goal, Cognitive Computing combines multiple AI methods: self-learning algorithms, pattern recognition by neural networks, and natural language processing.

Artificial intelligence has already begun to transform many aspects of our lives and is expected to continue to do so in the future. In this chapter, we describe current applications of AI and opportunities for future developments across multiple sectors like healthcare and education. As AI technologies continue to develop, it is important to consider the potential benefits and opportunities it presents, but also the ethical considerations that must be taken into account to use these technologies in the right way and with the right purposes.

APPLICATIONS & OPPORTUNITIES

Supporting Human Decisions

1

Many AI applications have the ultimate goal of helping humans to make better decisions. These applications are known as Intelligent Decision Support Systems (IDSS). IDSS use artificial intelligence (AI) to help humans make decisions by providing them with relevant information and recommendations. The feature that makes recommendations in streaming services like Netflix or Spotify is an example of IDSS. IDSS help bridge the gap between analysing a situation (which movies or songs are available for the genres I like) and making a decision based on this analysis (which movie or song I will choose).

Of course, the application of IDSS goes beyond making movie and song recommendations. IDSS can be embedded in various other systems, like a chatbot offering advice on financial investments or a navigation app providing the fastest route by analysing all possible access points and monitoring traffic in real-time. IDSS are also used in healthcare to analyse patient history, medical images, and specific protocols, and provide suggestions for treatment.

One of the reasons why AI is attracting so much attention is the increasing use of AI to make decisions in the public realm. Governments are gradually integrating AI to deliver and improve their operations, with much-untapped potential, particularly concerning policymaking and monitoring. The development and monitoring of public policies require the ability to sense patterns, develop evidence-based programs, forecast outcomes, and measure effectiveness, all concentric on AI capabilities. By leveraging the increasing amounts of data being collected about daily life, AI can help governments to build a comprehensive evidence-based foundation that transcends the barriers between public sectors. Going beyond sectoral barriers, AI can reveal complex patterns and interlinkages between sectors and thus provide comprehensive evidence for decision-makers to design better public policies. However, introducing AI systems into public decision-making raises multiple questions, concerning liability and accountability, transparency, safety and security, impacts on jobs, and ultimately threats to fundamental rights and democracy[1].

1 European Parliament (2022). [Artificial intelligence: threats and opportunities]().

Healthcare

2

AI is already a reality in healthcare, pharma, and life sciences industries to support diagnosis, medical image analysis (like X-ray images), and the development of new drugs. AI helps analyse, present, and understand medical and healthcare data. These data can include, for example, patient history and medical images. AI can learn from a huge set of medical images to diagnose diseases like cancer, a task impossible for humans. AI algorithms are already able to detect 95% of skin cancers using medical images[1]. Using patient data, AI can review their medical history to identify patterns and recommend interventions targeting unique symptoms and characteristics of people. For example, AI systems can suggest specific exercises patients should perform to improve injury recovery.

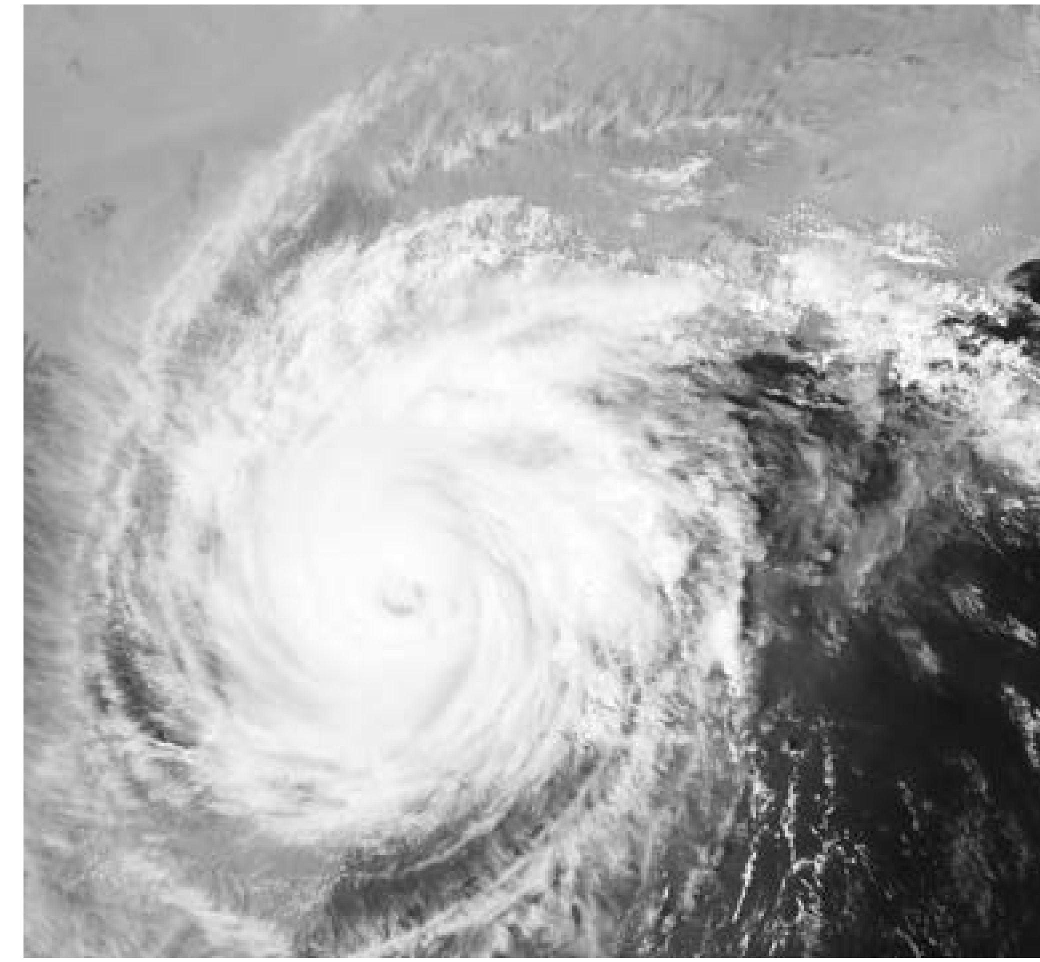
Advances in AI also bring the potential to boost patient care. One example is using so-called carebots - robots that support patient care. Images of nurse-like robots pop up when searching for carebots on the web. However, in reality, carebots look less like robots and more like invisible chunks of computer code connected to cameras. This way, carebots can monitor patients' health conditions, providing medications when needed or giving reminders about when patients need to take their medicine. They can also help lift and move patients and may serve as social companions. Carebots help human caregivers, alleviating significant manual or repetitive work. AI can also hold ground in places with a severe deficit of healthcare workers by taking over some diagnostic and care duties. Carebots thus offer many opportunities for the health and care support of vulnerable groups like the elderly and people with physical and mental impairment.

Despite the huge potential of AI applications in the healthcare sector, AI must be integrated into existing medical processes rather than substituting current practices. Doctors can achieve enhanced medical care by integrating AI insights into conventional medical practices that include both patient interaction and medical team discussions. This integrated approach can ensure encompassing diagnostic and healthcare processes, in a productive way that leverages both human and machine knowledge

1 https://www.karger.com/Article/FullText/504785

Environment & Climate Change

3



In today's world, environment preservation and combating climate change are of main concern for nearly all governments or public institutions. AI can provide helpful tools to identify patterns and correlations in large amounts of environmental data, providing insights that would otherwise would not have been able to be detected by humans. For example, concerning how wild species are disappearing and with them, the biodiversity that supports life on earth, an AI-powered system called WildTrack[1] is exploring the use of artificial intelligence in biodiversity preservation by analysing footprints the way indigenous trackers do and protecting these endangered animals from extinction. In this way AI can be used to monitor wildlife populations and their habitats. Additionally, it can track deforestation and desertification through analysis of satellite imagery or predict the occurrence of extreme climate-related events with higher accuracy. These applications can provide valuable information to government bodies to design required policy responses intended to cope with these challenges.

1 https://www.wildtrack.org

Energy Systems

4

As our energy systems are the largest contributors to carbon emissions globally, it comes as no surprise that we attempt to leverage AI technologies to minimise the carbon emissions of the energy sector. The amount of electricity generated from renewable energy sources such as wind and solar has increased from around 30 000 GWh in the year 2000 to almost 2 500 000 GWh in 2020, averaging a yearly growth rate of almost 25%[1]. This means that progress is happening, but we are still far from a sustainable energy system. When thinking about the energy transition, the first images that come to mind are probably not of AI. However, AI is already very much present in our energy systems, helping to accelerate the transition towards a sustainable energy system. AI has brought improvements in many aspects of the energy transition, including forecasting of solar and wind energy, grid operation and optimisation, and coordination and management of distributed energy assets, like solar panels. Many countries already use AI systems to optimise the system's performance and prevent power outages.

One of the main use cases for artificial intelligence in our current energy systems is within energy planning, demand and control. By analysing vast amounts of unstructured data on energy supply and consumption, AI technologies can enable more efficient operation of power companies. Identification of patterns of electricity demand and production results in more accurate predictions of future demand and necessary supply. These predictions can serve to shift flexible demand to reduce grid-strain, highlight optimal moments for increased levels of electricity storage or determine the optimal combination of electricity generators to reach the necessary supply[2].

Another important applications of AI in energy systems are grid fault detection, data security for smart grid and forecasting of electricity prices. As the amount of data produced by our energy systems increases along with general computational power, AI will continue to become more embedded in and vital to energy systems. Future advances in AI in supporting the energy transition look promising, but its full potential is yet to be realised. Examples of such advances are the prevention of cyber attacks on energy systems, which is of course crucial to national security, the integration of electric vehicles and to smart grids to serve as storage capacity, and improved optimisation of energy yield[2].

1 IEA (2022), Energy Statistics Data Browser, IEA, Paris https://www.iea.org/data-and-statistics/data-tools/energy-statistics-data-browser.

2 Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y., & Chen, H. (2021). Artificial intelligence in sustainable energy industry: Status Quo, challenges and opportunities. Journal of Cleaner Production, 289, 125834.

Transportation

5

The use of AI has exploded in the automotive industry. Perhaps the most evident application of AI in the mobility sector is the advent of autonomous vehicles (or self-driving cars[1]). AI is currently being used in voice recognition, voice search, image recognition and detection in autonomous vehicles. An autonomous vehicle can drive itself from a starting point to a predetermined destination in "autopilot" mode. This is possible thanks to the various sensors installed in these vehicles. Autonomous vehicles are fitted with cameras, sensors, and communication systems to generate massive amounts of data. When such data is used to train AI systems, it enables the vehicle to sense the environment and make decisions. In all cases, autonomous vehicles still require a human driver to override the system when necessary and are not truly autonomous.

AI also supports the management of transport systems at a larger scale. AI improves overall mobility by analysing various types of data relevant to transport operations, like weather forecast, topology of the roads, and information about past incidents and traffic bottlenecks. For example, if a road accident happens or a road gets flooded due to heavy rain, the system can direct the flow of cars to less affected roads. Furthermore, this type of AI application is not restricted to terrestrial transportation. In fact, AI has already reached our skies. One example is the use of AI to predict the arrival times of long-haul flights to optimise ground handling and shorten passenger queues. The AI system used by the Singapore's Changi Airport can make such predictions with almost 95% accuracy[2]. AI is also present in mobility systems in less sophisticated ways, for example, by providing recommendations of the fastest routes and counting the number of cars parked in a restricted area to prevent overcrowding. AI is also present in mobility systems in less sophisticated ways, for example, by providing recommendations of the fastest routes and counting the number of cars parked in a restricted area to prevent overcrowding.

1 https://medium.datadriveninvestor.com/artificial-intelligence-and-autonomous-vehicles-ae877feb6cd2

2 https://cmp.smu.edu.sg/ami/article/20190522/ai-gets-real-singapores-changi-airport-part-1

Water Management

6

Access to water is one of the most basic human needs for health and well-being. However, we are still far from universal coverage, and progress is slow. According to the United Nations, between 2015 and 2020, the population with access to safe drinking water services increased from 70 to 74%, and the population with access to handwashing facilities in the home increased from 67 to 71%. To achieve universal coverage in 2030, the rate of progress would need to quadruple. This is a big challenge given that water demand is rising due to population growth, urbanisation and increased water use in agriculture, industry, and energy sectors. In addition, countries are facing growing challenges linked to water scarcity caused by climate change[1].

The application of AI tools in the water sector contributes enormously to addressing multiple global challenges, in terms of water quality, management, and monitoring. For instance, acoustic sensors and gas tracers are being deployed to identify water leaks and detect contaminants in the water. These sensors and tracers are connected to AI systems to offer real-time monitoring of water quality, allowing for early detection of leaks and contamination, and thus preventing further damage. Furthermore, AI is increasingly being used in agriculture to optimise water use. AI can be used to monitor soil conditions and estimate water demand. Combined with complex systems to predict weather conditions, AI will help develop ideal crop management strategies and streamline the use of water used in agriculture making it more efficient. This is extremely important given that 70% of freshwater use today is in agriculture and the efficiency of use is less than 50%[2].

1 https://sdgs.un.org/goals/goal6

2 https://inc42.com/resources/how-ai-is-transforming-the-water-sector/

Food Systems

9

AI is turning out to be the secret ingredient for creating more efficient and sustainable food systems. According to the United Nations, the number of people going hungry and suffering from food insecurity has been gradually rising since 2014. The COVID-19 pandemic has led to a spike in this trend. In 2020, between 720 and 811 million persons worldwide were suffering from hunger - as many as 161 million more than in 2019. In the same year, over 30% of the world's population was moderately or severely food-insecure, lacking regular access to adequate food. The war in Ukraine has since aggravated this scenario, triggering food shortages that affect the most vulnerable.

AI-driven systems are already being deployed in various ways across the agricultural industry. The combination of AI systems with remote sensors, such as satellite imagery and drones, can help in tackling challenges such as pest management, nutrient deficiency, and irrigation issues. Other applications include the optimisation of irrigation schedules and the prediction of crop yields[1]. The application of AI to solve food challenges goes beyond the agricultural sector. AI is being used in food processing to analyse the quality of food and determine whether it is suitable for consumption, thereby reducing food waste.

1 R. Sharma, "Artificial Intelligence in Agriculture: A Review," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2021, pp. 937-942, doi: 10.1109/ICICCS51141.2021.9432187.

Culture, Education & Academia

8

AI has also found applications in arts and culture. One example is the AI system called DALL·E 2[1], which creates images and art from a simple combination of words. Starting from a simple set of words like "an astronaut", "riding a horse", and "photorealistic style", the system creates multiple realistic images that depict an astronaut riding a horse. By changing "riding a horse" to "playing basketball", a new set of images is generated, now with an astronaut playing basketball instead of riding a horse. Then, by changing the style from "photorealistic style" to "style of Andy Warhol", the system generates a new set of images of an astronaut riding a horse that resemble Andy Warhol's famous creations. All of this in a manner of seconds. AI systems working in similar ways can also generate other cultural and art pieces, from songs and guitar tabs to fiction stories and essays. All artistic material generated by AI is called synthetic media and covers audio, text, and video.

Like in the cultural sector, the potential applications and opportunities for AI within education and academia are plentiful[2].

In the education sector, AI can assist teachers in various ways to provide a more efficient learning process. Examples of possible tools are personalised learning algorithms that can focus on areas where a particular student is struggling. A popular platform making use of such algorithms is Duolingo[3], a company producing learning applications and providing language certification. Other examples are virtual tutors that can provide immediate feedback to individual students, or intelligent grading and assessments systems that can free up time for teachers to spend their time on higher level tasks like mentoring.

However, there are also serious drawbacks and threats to introducing AI into the classroom. AI can easily provide answers to students, stopping them from thinking critically and solving problems themselves. This is illustrated by the recent introduction of OpenAI's large language model ChatGPT[4], able to take over and perform many student assignments, like essays and reports. This aggravates increased dependence on technology, preventing students from thinking and learning independently. Additionally, increased use of AI-driven systems in education could reduce personal interaction and relationship-building, which is especially important for younger students in classrooms. New ways of teaching with percolating AI-based applications are urgently necessary.

In academia, too, AI can support research activities. Traditional academic ways of analysing acadtext cannot keep up with the massive amounts of textual data generated at unprecedented rates. AI technology can screen vast amounts of textual data faster and more efficiently than humans. Recently, AI started to be used more extensively in the analysis published articles, supporting researchers and scholars in drawing connections between different fields of research and finding new insights and hypotheses. A concrete example is the use of AI to search databases of molecules and find the ones with desired properties for a specific application[5].

1 https://openai.com/product/dall-e-2

2 Luckin, Rose; Holmes, Wayne; Griffiths, Mark and Forcier, Laurie B. (2016). Intelligence Unleashed: An argument for AI in Education. Pearson Education, London.

3 https://www.duolingo.com

4 https://openai.com/blog/chatgpt

5 https://www.iotforall.com/artificial-intelligence-in-research#:~:text=Researchers%20use%20AI%2Dbased%20algorithms,would%20take%20an%20expert%20scientist.

Communication

9

Over the last decade, voice assistants have really catapulted to fame, with Amazon Echo, Google Home, and Apple's Siri. These applications work through Natural Language Processing (NLP), a subdomain of AI to teach computers to comprehend natural language or human speech. Fundamentally, NLP applications take a clip of verbal audio in a particular language and transcribe it into text. Applications like Siri also extract the text and even interpret the semantic meaning of the spoken words, responding or taking action on what was said to them.

Voice assistants are mostly used to help people with daily tasks, like reading the weather forecast, checking calendar events, and playing songs. However, they also find application beyond that. Voice assistants can help children in the autism spectrum to practice communication in a safe environment. People who are visually impaired or have dyslexia also find a great help in voice assistants. Last but not least, voice assistants can help to reduce the feeling of loneliness, particularly among elderly people who live alone[1].

1 https://www.tudelft.nl/en/stories/articles/this-is-how-digital-voice-assistants-influence-your-life

Speech recognition also becomes important for people who have face or speech impairments since it provides an alternative way for a conversation. Scientists are exploiting AI for these purposes, since sign language may not be a long-term solution. Some challenges still remain: Lip reading is difficult even for AI, and some solutions require voice aid implants. If AI manages to analyse emotions by looking at facial expressions and hand gestures, it will revolutionise communication for impaired people.

Virtual Reality

10

AI has changed the way we use our smartphones. The camera and image sensors have become accessible and very cost-effective making them the locus of a lot of information. We can now try out clothes and eyewear on the internet thanks to AI. We can also design an entire room using just the camera on our smartphones. Creating these virtual environments in our smartphones is possible due to two AI features: augmented reality (AR) and virtual reality (VR). AR augments our surroundings by adding digital elements to a live view - it is like an extension of reality. VR, in contrast, provides a completely immersive experience that replaces a real-life environment with a fully simulated one.

VR and AR applications are not limited to what we can do using our smartphones. They are extensively used in the gaming industry: there is a huge difference in terms of virtual experience between games of the 90s and games of today.

But, beyond entertainment, VR and AR technologies are being used in various fields with significant societal impact. For example, safety simulations are being carried out in the civil engineering sector to develop better construction practices[1]. Similarly, VR safety simulators are being used to train professionals across aviation, automotive, medical, and military sectors. In education, AR and VR can make schooling more interactive, engaging, and efficient[2]. For instance, students can observe a 3D galaxy on their tablets or even make discoveries in a virtual science lab without leaving the classroom.

1 https://www.scratchie.com/post/vr-training-the-secret-weapon-to-immersive-construction-safety-education

2 Khan, T., Johnston, K., & Ophoff, J. (2019). The impact of an augmented reality application on learning motivation of students. Advances in human-computer interaction, 2019.

Artificial intelligence has the potential to revolutionise many aspects of our lives, but it also presents numerous risks and challenges that can have serious consequences for social justice, safety, and other critical aspects of our daily lives. In this chapter, we explore the various threats and negative impacts of AI, including the potential for AI to be used for malicious purposes, the risk of social biases and discrimination in the use of AI, the lack of transparency and the complexity of accountability, and challenges we face to preserve privacy while exchanging increasing amounts of personal data.

Chapter 4

THREATS & NEGATIVE IMPACTS

Bias & Discrimination

1

One of the most negative impacts of AI is undeniably the reproduction and perpetuation of social biases. AI systems are not neutral because they embody the biases held by their designers and developers (bias by design) and reproduce the biases existing in the data that feeds them (garbage in, garbage out).

Biased by design

A concrete example of how social biases are embodied in AI systems from its design phase is observed in digital voice assistants, like Siri and Alexa. These devices are used for tasks like setting reminders and scheduling appointments. The way people interact with voice assistants is often top-down, via commands. It does not help that in many cases the command has to be repeated several times, which makes people agitated. People command, and voice assistants must obey. Not surprisingly, often female names and voices are attributed to them. By deliberately assigning female attributes to voice assistants and robots, AI developers are introducing their own biases into AI systems and helping to perpetuate gender stereotypes[1]. This is also seen in other sectors like healthcare and security, where robots identified as female are perceived as more trustworthy and reliable by people.

1 https://www.tudelft.nl/en/stories/articles/this-is-how-digital-voice-assistants-influence-your-life

Garbage in, garbage out

AI systems work by learning patterns in the training dataset (as explained in Chapter 2). The quality of this dataset thus determines the quality of the outputs of AI. “Garbage in, garbage out” is one of the rules of computer science. It means that the output of an AI system is as “good” as the data that enters the system. In other words, if a biased dataset is used as input into the AI system, the output will be biased as well.

A biased dataset is a set that represents only a small part of the society. A recent study conducted at Stanford University found that speech-to-text services used by tech giants like Amazon misidentified the words of Black speakers far more than white speakers[2]. The internet is also flooded with examples in which Alexa and Siri struggle with Scottish accents[3]. These situations happen due to a lack of diversity in the data. If AI systems continue to feed on limited datasets that do not incorporate a diverse set of language variants, voice assistants will continue to struggle to capture language nuances like accents and dialects. This might sound like a minor issue, but what if they are used in autonomous vehicles and fail to recognise commands from the driver? What if they are used in emergency call centers and fail to recognise requests for help? In these cases, failure to interpret human speech due to gender, race, or accent, could have tragic endings.

2 https://news.stanford.edu/2020/03/23/automated-speech-recognition-less-accurate-blacks/
3 https://www.heraldscotland.com/opinion/16219100.agenda-scots-struggle-alexa-siri/

There are many reasons why datasets are biased. In some cases, data are biased due to an inadequate data collection process that excludes a portion of the population (sometimes intentionally). In other cases, data are inherently biased simply because they reflect historical biases in our society. Other times is a combination of both, biased data emerges from a juxtaposition of bias from both historical context and collection processes. Regardless of what makes a dataset biased, the issue is that when biased data go unchecked, AI systems will make biased decisions and continue to reproduce social disparities.

1

The consequences are profound

The reproduction and reinforcement of existing biases are highly problematic as AI is increasingly being deployed across both public and private institutions, influencing decisions that affect people’s lives. With existing social bias woven into AI systems, their outcomes are bound to be unrepresentative for specific social groups, including women, people of colour, or ethnic minorities. There is ample evidence that AI causes discriminatory harm against already vulnerable groups.

Facial recognition is one of the AI applications that has already raised issues concerning bias and discrimination[12]. Recent research has shown that facial recognition software performs much worse at identifying certain demographics. An incorrect match between two faces is called a false positive. False positives are more likely to occur for women and people of colour compared to white men. This means that women and people of colour are more likely to be wrongly identified by facial recognition programs. The recognition inaccuracy is a reflection of the imbalance in the composition of the training data, a typical case of garbage in, garbage out.

1 https://www.nature.com/articles/d41586-020-03186-4
2 https://www.nature.com/articles/d41586-020-03187-3

Another major concern related to bias in AI is its use to replace or automate decision-making. AI systems are increasingly being used to screen and select people for housing, mortgage, and social benefits. For example, to calculate loan terms, the system uses data about the applicant’s credit score, income, total assets, debt-to-income ratio, among others. In addition to financial data, these systems also use historical data including criminal and eviction data, which reflect long-standing racial disparities in both legal and housing systems[3]. This means that an applicant can be denied a loan if their profile falls under a category deemed ineligible by the AI system, even if their financial situation is favourable. Growing evidence shows that AI-powered systems perpetuate housing discrimination during tenant selection and mortgage qualification[4]. Members of minorities have flagrantly been denied housing despite having sufficient financial resources[5].

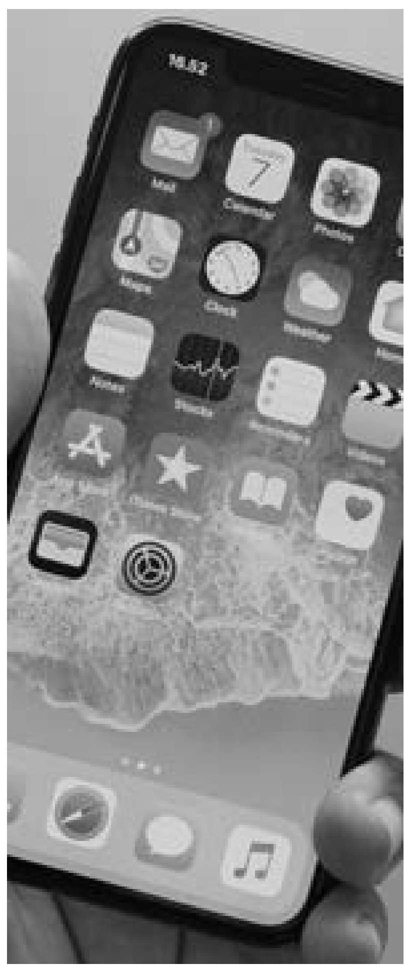
3 https://www.aclu.org/news/privacy-technology/how-artificial-intelligence-can-deepen-racial-and-economic-inequities
4 https://archive.curbed.com/2019/12/17/21026311/mortgage-apartment-housing-algorithm-discrimination
5 https://www.cbsnews.com/news/mortgage-discrimination-black-and-latino-paying-millions-more-in-interest-study-shows/

In The Netherlands, a similar example of how AI systems have had severe and detrimental impacts on human life is the “toeslagenaffaire” (child benefits scandal). The child benefits scandal refers to a situation involving the erroneous suspicion of fraud regarding childcare and other types of social benefits and the subsequent strict recovery of these benefits by the Dutch tax agency. By wrongfully having to repay the often substantial amounts of received social benefits, numerous families had to take on up to tens or even hundreds of thousands of euros of debt. This financial hardship led to the disruption of families, out-of-home placement of children, loss of jobs, and serious psychological problems. The use of AI algorithms in this context was meant to improve the efficiency and accuracy of the benefits’ calculation process; however, biases in AI algorithms led to certain groups being disproportionately affected by errors in the calculation process. These were typically groups with lower income and ethnic minority groups.

Personal Data & Privacy

2

Many of our daily activities generate data: Our smartphones count the number of steps we take, our transportation cards track our travel habits, and supermarket loyalty cards know what we like to eat. The massive data currently being generated by human activity is often called Big Data. The rise of Big Data is arguably one of the reasons why AI has become so intertwined in our society. AI thrives on large datasets provided by Big Data: the more, the better. This combination has a profound impact on questions of privacy.



A recent episode involving data privacy is the Cambridge Analytica scandal. Cambridge Analytica was a British consulting company behind the collection of data from millions of Facebook users without their consent. The data was then (mis)used for political advertising during the 2016 presidential elections in the USA. After the scandal was reported in the media around 2017-2018, the co-founder of Facebook, Mark Zuckerberg, published a letter in various news outlets apologising on behalf of the company. Facebook also decided to implement the EU's General Data Protection Regulation (GDPR)[1] beyond the EU to all their areas of operation. In 2019, the scandal was reported in a Netflix documentary called The Great Hack[2]. Despite the gravity of the scandal, Facebook continues to grow and generate revenues.

Another privacy concern related to AI-powered systems is the widespread use of intelligent personal assistants like Apple's Siri, Amazon's Echo, and Google Assistant. They are designed to assist people with most trivial tasks like describing the weather forecast, playing music, and checking calendar events, among others. In principle, personal assistants act upon being called. However, there are concerns about their constant listening and data collection[3]. These questions are bounded to become more problematic with the increased use of care and companion robots working inside people's homes.

1 https://gdpr.eu/what-is-gdpr/
2 https://en.wikipedia.org/wiki/The_Great_Hack
3 https://www.tudelft.nl/en/stories/articles/this-is-how-digital-voice-assistants-influence-your-life

Fake news

3

The term fake news refers to false stories that appear to be news but are created to deliberately misinform or deceive readers. This is hardly a new phenomenon. Hoaxes and misinformation have long been part of strategies to manipulate public opinion. They have also been used during wartime to deceive opponents. However, in a world increasingly dominated by digital technologies, the magnitude of fake news has increased dramatically, to a level that is difficult to grasp. Misinformation campaigns now take advantage of social media to quickly spread fake news. The consequences of fake news and misinformation are wild, including greater political polarisation, and eroding trust in the media and governments, all ultimately constituting a threat to social cohesion and democracy.

AI has been playing an important role in the spread and creation of fake news, particularly on social media like Twitter and Facebook.

Fake news spread

One of the most negative impacts of AI is its role in spreading misinformation at a pace and scale not seen before. Many advertising companies have been using AI to identify and target groups of consumers that are likely to buy their products. In a similar way, AI has been used to identify and target groups of people with specific fake news. A very extreme example is that AI can identify individuals who often engage with content related to theories of conspiracy and then target them with more similar content. This is problematic because it blinds those individuals from having contact with other sources of information that may provide an alternate view or evidence to counter conspiracy theories and false beliefs.

Social media bots

Another way in which AI can help propagate fake news is via so-called bots. Bots are fake social media users. Bots simulate the behaviour of humans in a social network: they can share information and messages and interact with other users. They are basically computer code created by someone with computer skills. When AI algorithms are introduced into social media bots, they can learn to respond to certain situations by analysing how people interact on social media. Social media bots can be programmed to keep posting fake news and to commenting pieces of fake information on the postings of real social media users. To make the information look credible, often a network of bots is put together to work incessantly to circulate large numbers of fake news-related comments and content.

AI-created fakes

So-called "deep fakes" exploded in 2017 and are the most recognised form of AI-generated news media. Deep fake news includes tools that are capable of generating photos, videos and audio messages that replace one person's face and voice with another to create fake situations. Needless to say, fake video and audio recordings can have disastrous consequences. Just imagine an election in which we cannot distinguish whether campaigning videos and audios are fake or true. Moreover, this type of AI application is under scrutiny due to fraud and security risks, as it can be used to bypass biometric security tools like facial recognition or voice recognition software.

3

Transparency

4

Another major issue with AI systems is their lack of transparency and interpretability. Due to the lack of transparency, sometimes AI systems are referred to as “black-boxes” (sic): Whatever happens in the AI “black-box” is not clear to users and developers. In other words, it can be very hard to understand how AI systems come to a certain result or decision, particularly advanced AI systems that make use of deep learning. The mechanisms through which AI systems arrive at a specific answer can be unintelligible even to the most brilliant scientists.

The lack of interpretability and transparency leads to systems that are hard to control, monitor, let alone correct, making it very difficult to predict failures and establish accountability. Although the lack of transparency and interpretability in AI has been a concern for over 40 years[1], these issues are expected to become more pronounced as AI techniques develop further and become more complex. The question that remains open is: If we cannot fully understand AI systems, how can we deploy them safely?

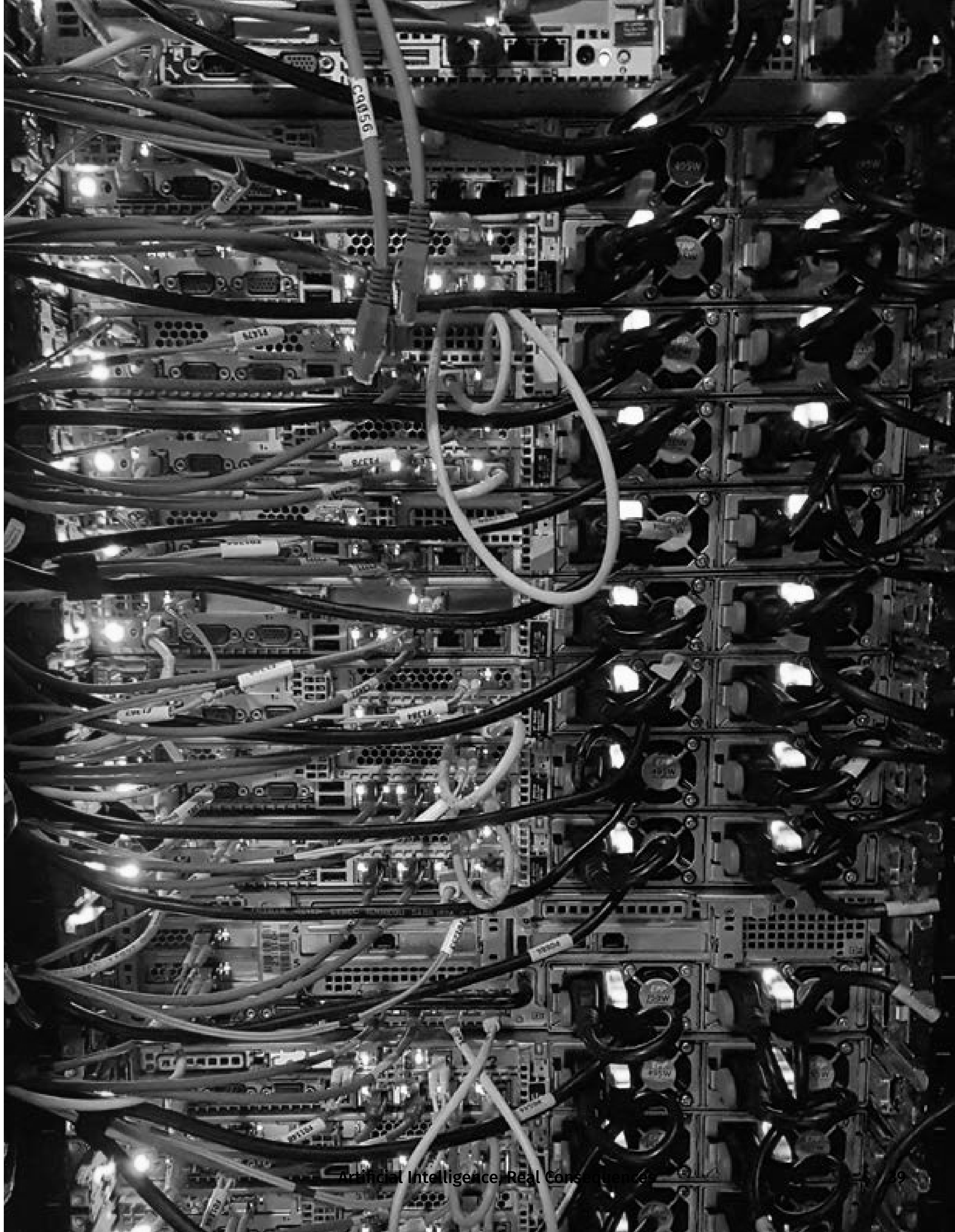
1 https://link.springer.com/chapter/10.1007/978-3-030-32236-6_51

Environmental Impact

5

It is clear that AI applications provide a lot of opportunities across multiple sectors of society, including the energy transition and climate change mitigation. However, the infrastructure around AI systems is highly energy consuming. Data and training centres demand huge amounts of energy, as computer systems have to run continuously to process data streams. Some data centres in the US require more energy than 80.000 households yearly[1]. Moreover, these centres also require millions of gallons of water for cooling. The environmental impact of AI infrastructure is not limited to the energy and water required during operation but also the energy and materials used in the construction of supercomputers, data collection, and storage. This quashes notions about the digital space being green. We are far from that.

1 https://www.weforum.org/agenda/2022/08/sustainable-data-centre-heating/



In this chapter, we give the word to the experts with a selection of insightful critical statements about AI.

THE WORD OF EXPERTS

"For AI to support or replace human decision making, one must understand human decision making first."

Prof.dr.ir. Alexander Verbraeck,
Department of Multi-Actor Systems

"AI systems, however technically robust, may do more harm than good if implemented without considering the broader sociocultural setting of an organization."

Olya Kudina,
Department of Values Technology and Innovation

"AI system may have unintended consequences. These need to be monitored during their lifetime and may require embedding additional values in the AI system over time."

Prof.dr.ir. Ibo van de Poel,
Department of Values Technology and Innovation

"The entire lifecycle of an AI system must be trustworthy."

Jonne Maas,
Department of Values Technology and Innovation

"Artificial Intelligence quickly consists of millions to trillions of calculations. Moreover, those calculations are adjusted based on massive datasets, and so it's practically impossible to track everything the AI does to reach an output."

Stefan Buijsman,
Department of Values Technology and Innovation

"Models of human behaviour that are trained with large datasets can replicate human behaviour but it will be hard to explain "why" the behaviour occurs."

Prof.dr.ir. Alexander Verbraeck,
Department of Multi-Actor Systems

"Without help, it's incredibly difficult to tell why modern AI gives a specific answer."

Stefan Buijsman,
Department of Values Technology and Innovation

"Failure to properly interpret human speech due to e.g. gender, age or accent bias, can lead to serious and costly mistakes."

Olya Kudina,
Department of Values Technology and Innovation

"AI is not the same as machine learning. Other forms of AI, including more symbolic ones, remain very relevant in technological innovations that can help in addressing today's grand societal challenge."

Prof.dr. Martijn Warnier,
Department of Multi-Actor Systems

"To unlock major breakthroughs in science, progress, scholars must transcend the boundaries between theory-driven and data-driven modelling paradigms."

Sander van Cranenburgh,
Department of Engineering Systems and Services

"With [AI methods for] automated text analysis, we can utilize novel data sources to build, test, and refine theories of public policy to explain why and how (some) policies succeed."

Nihit Goyal,
Department of Multi-Actor Systems

"The energy transition is a term used to capture desired long-term changes throughout all aspects of society in light of climate change, among others. Only part of the challenge is developing technology. Also important are issues related to social dynamics, policy, and implementation. AI methods help to tackle these challenges."

Emile J. L. Chappin,
Department of Engineering Systems and Services

"Ethically designed care robots can improve the quality of care through various functions, such as physical assistance and emotional companionship."

Shuhong Li,
Department of Values Technology and Innovation

"Traditional coding requires programmers to define rules and logic that a computer can interpret and execute. For some tasks, e.g. vision, it is impossible to define rules that enable a computer to see. Deep neural networks learn from data to expand what computers can do. Deep neural networks are powerful because they can help computers do tasks that are difficult or impossible to code: e.g. seeing, hearing. Even though deep neural networks are used in many products, they can fail in unexpected ways and discriminate against users."

Wiebke Hutiri,
Department of Engineering Systems and Services

"We need a stronger focus on democratic legitimacy regarding the development and deployment of AI systems."

Jonne Maas,
Department of Values Technology and Innovation

Want to know more? A must-read list

Non-fiction

Aalbers, M., Alizadeh, T., Ash, J., Graham, M., Rose, G., & Shaw, J. (2019). How to run a city like Amazon, and other fables. Meatspace Press.

Benjamin, R. (2020). Race after technology: Abolitionist tools for the new Jim code. John Wiley & Sons.

Bostrom, N. (2014) Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

Brevini, B. (2021). Is AI good for the planet?. John Wiley & Sons.

Bridle, J. (2018). New dark age: Technology and the end of the future. Verso Books.

Broussard, M. (2018). Artificial unintelligence: How computers misunderstand the world. mit Press.

Coeckelbergh, M. (2022). The political philosophy of AI: an introduction. John Wiley & Sons.

Crawford, K. (2021). The atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Greenfield, A. (2017). Radical technologies: The design of everyday life. Verso Books.

Joque, J. (2022). Revolutionary Mathematics: Artificial Intelligence, Statistics and the Logic of Capitalism. Verso Books.

Kaltheuner, F. (Ed.). (2021). Fake AI. Meatspace press. Available at https://fakeaibook.com.

Kearns, M., & Roth, A. (2019). The ethical algorithm: The science of socially aware algorithm design. Oxford University Press.

Marcus, G., & Davis, E. (2019). Rebooting AI: Building artificial intelligence we can trust. Vintage.

Mattern, S. (2020). A city is not a computer. In The Routledge Companion to Smart Cities. Routledge.

Mitchell, M. (2019). Artificial intelligence: A guide for thinking humans. Penguin UK.

Murgia, M. (2023). Sci-fi writer Ted Chiang: 'The machines we have now are not conscious'. Financial Times.

Noble, S. U. (2018). Algorithms of oppression. In Algorithms of oppression. New York University Press.

O'neil, C. (2017). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.

Russell, S. (2021). Human-compatible artificial intelligence. Human-like machine intelligence, 3-23.

Williams, A., & Gilbert, J. (2022). Hegemony Now: How Big Tech and Wall Street Won the World (And How We Win it Back). Verso Books.

Fiction

Nayler, R. (2022). The Mountain in the Sea: A Novel. MCD Books.

Newitz, A. (2018). Autonomous. Tor Books.

Maughan, T. (2019). Infinite Detail. MCD X FSG Originals.

List of images

AI*