1. AI to optimize the effectiveness and efficiency of public services

INTRODUCTION

With rapid digital technological change, it is inevitable for the government to innovate its traditional methods in order to achieve better citizen engagement, accountability, and interoperability (...). AI can help in freeing up the government labor by its implementation in automating the repetitive tasks resulting in increased transactions speed in the provision of government services and also accurately assessing the outcomes of policy options. AI has a huge potential in different government sectors such as, education, physical infrastructure, transportation, telecommunication, data security and management, finance, healthcare, research and development, policymaking, legal and justice system, etc. (Sharma, Yadav, & Chopra, 2020)

Technology and its multiple forms have played a key role in all civilizations. As Hannah Arendt (1958) argued, “[t]ools and instruments are so intensely worldly objects that we can classify whole civilizations using them as criteria” (p.144). Time and specific eras are often identified by the main or the newest technology innovation of that time: we refer to the “digital,” “stone,” or “steam” ages. Some nations are even characterized by their dominant technological artifacts, such as the Netherlands and windmills, or Japan and micro-electronics (Wyatt, 2008). The role of technology in a society illustrates well the experience of living in a specific era and place (Heilbroner, 1994a, 1994b). This propensity to designate an era or a nation with a technology may be due to the fact that the first scholars to examine in-depth technological change were anthropologists and archaeologists, who use technology as a marker to distinguish eras of development (Mumford, 1961). This role is also indicative of the perception of technology and the choices a nation makes at a given time and in a given place.

Governments have progressively adopted a number of technology innovations to respond to a growing demand to (1) digitalize public action and optimize its operations and services (de Feraudy, 2019), and (2) increase citizen engagement in the development, implementation, and evaluation of public policies (de Feraudy & Saujot, 2017). This is what the concept of e-gov or e-government refers to: using technology to achieve “higher levels of
effectiveness and efficiency in governmental tasks, improvement of processes and procedures, increases the quality of public services, also improves the use of information in the decision-making processes and allows for better communication among different governmental offices” (OAS, n.d.). The e-government efforts were mainly to take advantage of technological advances to (a) optimize the effectiveness and efficiency of government services, (b) put the citizen back at the center of the design of services rendered by organizations, and (c) increase trust in government (OECD, 2020). As Sidjanski (2000) argues, “[t]he emergence of the microcomputer reversed the trend by making it possible to develop horizontal organizations that could to a large extent replace vertical structures” (p.203).

Artificial intelligence (AI) is at the center of a stream of technological solutions, which are increasingly adopted by governments. For instance, it is used to process sensitive information for public health as illustrated by the many applications to combat the spread of the Covid-19 pandemic. AI applications can be considered useful for six types of government challenges: allocating resources, analyzing large datasets, overcoming the shortage of experts, predicting scenarios, managing procedural and repetitive tasks, and diverse data aggregation and summarization (Mehr, Ash, & Fellow, 2017). AI can also provide automated legal advice at lesser cost (Nissan, 2017). However, AI presents numerous challenges, whether they stem from technical or adversarial vulnerabilities (Mitchell, 2019). Vulnerability consists of weaknesses or flaws whether in the hardware, software or data security, which can enable an attacker to compromise its integrity (i.e. trustworthiness of a resource), availability (i.e. appropriate user is denied access to a resource), or confidentiality (somebody gains access to information that she should not have had access to) (see Bowen, Hash, & Wilson, 2006). Moreover, AI is often criticized for its black box characteristics: very few experts can understand how the most complex AI systems function, their lines of code evolve with the more data they are fed with (in the case of machine learning algorithms as discussed further), and they are challenging to audit.

This chapter first clarifies the terms artificial intelligence and discusses the conceptual challenges to define this technology. It argues that AI remains this blurry (i.e. conceptual challenges), variable (i.e. ongoing developments and applications), often opaque (i.e. black box phenomenon) agent in the citizen–government relation with various degrees of agency (i.e. capacity to observe its environment, learn from it, and take smart action or propose decisions). It then examines the tasks it can perform and the benefits and risks for governments and other stakeholders to govern with AI. Lastly, it looks at specific uses of AI for public action, and efforts to govern and regulate this technology.
Artificial intelligence is not new. AI has been researched for over 60 years. Its development has taken place over time and through different phases (Darlington, 2017). In his article published in 1950 on computing machines and intelligence, Alan Turing already asked the question of whether machines could think (Tulloch & Lupton, 2003). The Turing test, which is still used today, allows to test an AI when a human being has an interaction with another human being while he thinks he has an interaction with a machine. For many, the Dartmouth Summer Research Project that took place in the summer of 1956 is the birthplace of artificial intelligence (AI). It was during these discussions and exchanges between John McCarthy, Alan Newell, Arthur Samuel, Herbert Simon and Marvin Minsky that AI was conceptualized.

Since then, AI research has developed in stages. During the 1950s until the 1980s, AI research focused on the ambition to make machine think through the use of symbols. This first generation of AI is called symbolic AI, also called “classical AI.” John Haugeland (1989) coined the term GOFAI (“Good Old-Fashioned Artificial Intelligence”) for symbolic AI. In robotics, the analogous term is GOFR (“Good Old-Fashioned Robotics”). We, humans, make use of symbols to find a specific solution to a mathematical problem. We use the “+” symbol to represent an action, which is adding, and the “=” symbol represents the result of an equation. Similarly to mathematics, we also use symbols to identify the most basic things (e.g. house, or table) and to describe people (e.g. man, woman, doctor, lawyer). We also use symbols to define everyday actions such as walking, drinking, and writing. It is based on the idea that a machine could be trained to think through the use symbols, which represent specific things in the real world (Techslang, 2020). Developers first mapped human “reasoning” to identify rules and symbols we use to think. These symbols are articulated (linked with each other) through a set of rules (e.g. logic, causation). These rules and symbols represent a model of reality, which allow the machine to make a decision by deduction.

Because symbols are necessarily precise representations of reality, they do not allow for implicit knowledge, such as “[a] mother will necessarily be older than her daughter.” This is a major limitation in a world that makes extensive use of implicit knowledge. Furthermore, despite obvious reasoning capabilities, the researchers failed to develop the learning capabilities of symbolic AI. Consequently, limited results and less enthusiasm toward this symbolic approach led to the AI Winter: a certain disinterest and reduction in funding for research on this technology in the 1970s.

In the 1990s, computing power and data storage progressed to the point where some complex tasks were feasible for machines. With the emergence of
the internet, the web 2.0, smartphones and social media platforms, new sources of data were soon available. Combined with the increased computing capacity, another approach to AI was possible: statistical approach. These technological advances generated new interest in AI research and attracted new funding (UW, 2006). Statistical AI systems differ from symbolic AI in their inductive process: from a large dataset, they induce trends and create generalizations. For instance, developers can train an AI to recognize cats. To do so, one feeds an AI with a large number of pictures, in which are tagged pictures of cats. The AI will detect patterns and criteria to identify cats in pictures and then create its own definition of cats.

In 1995, Richard Wallace developed the artificial linguistic internet computing entity, which can hold basic conversations. During the same decade, IBM developed the Deep Blue computer that played against Garry Kasparov. In 1996, Deep Blue lost, but won the rematch against Kasparov a year later. Deep Blue had the ability to consider forward six or more steps and could compute 330 million positions per second (Somers, 2013). In 2015, Alphabet DeepMind launched a computer program that can play the game of Go against the best players in the world. AlphaGo is based on an artificial neural network that has been trained on thousands of games played between amateur and professional humans. In 2016, AlphaGo managed to beat Lee Sedol, the best player in the world at the time. Then, the developers let the program play against itself. The result was a new program, AlphaGo Zero, which, through trial and error, managed to beat the original program and all other versions of AlphaGo in 40 days without human intervention or historical data (Silver et al., 2017).

Many AI applications combine both approaches (symbolic and statistical). As an example, natural language processing (NLP) algorithms, which are particularly used by sentiment analysis tactics frequently use a combination of statistical AI (which rely on large amounts of data) and symbolic AI (which consider issues such as grammar rules) (OECD, 2019).

We are currently at the stage of Artificial Narrow Intelligence (ANI). ANI or “applied” AI is developed to solve a specific problem-solving or reasoning task. ANI corresponds to robotized systems and applications that can be considered “intelligent.” They cannot mimic human behavior, but they can modestly perform tasks that would require human intelligence, effort, and time to an unsustainable degree, either because of environmental conditions unfavorable to human work or the slowness with which our brains could perform large-scale data analysis (Misuraca & van Noordt, 2020). The most advanced AI systems available today, such as Google’s AlphaGo, are still “narrow.” Indeed, even if they can to some extent generalize pattern recognition, as for example by transferring the knowledge acquired in the field of image recognition to speech recognition, the human mind remains much more versatile (OECD, 2017).
Two other future steps in the development of AI are worth considering. Artificial superintelligence (ASI) refers to a situation where technology will outperform human intelligence at all times and places, under all conditions and situations. The “technological singularity” refers to that moment in history when human beings are no longer the most intelligent species on earth, but are overtaken by AI. This stage, which for some is more in the realm of science fiction, stirs up dreams and anxieties. Many researchers and ethicists are already trying to prepare our societies for this hypothetical situation, in order to avoid a scenario where AI could take control or even act against the interests of humanity. General AI (AGI) refers to ICT systems with forms of intelligence that are similar to those of humans. Research efforts for this stage are primarily focused on replicating the inner workings of the human brain and applying it to a machine. “AGI would have a strong associative memory and be capable of judgment and decision making. It could solve multifaceted problems, learn through reading or experience, create concepts, perceive the world and itself, invent and be creative, react to the unexpected in complex environments and anticipate” (OECD, 2017). It should be noted, however, that as with the superintelligence scenario, general AI is far from being realized and may still take decades (or more) to manifest itself (Misuraca & van Noordt, 2020).

This book focuses on the uses of AI that are currently present in our societies (ANI). More futuristic forms of AI (AGI and ASI) will not be discussed due to their future and hypothetical nature. The next section will discuss conceptual challenges to define artificial intelligence.

CONCEPTUAL CHALLENGES TO DEFINE AI

Despite the excitement around the uses of AI in much of academia, industry, and public institutions, experts fail to agree on one common definition. Artificial intelligence “presents a difficult case for studies of topic sentiment over time” (Fast & Horvitz, 2016, p.963), since this technology is still under development, and its applications are so vast and diverse that there is no general agreement on a common definition of this technology. As Monett and Lewis (2018) argue, “[t]heories of intelligence and the goal of Artificial Intelligence (A.I.) have been the source of much confusion both within the field and among the general public” (p.212). Indeed, AI presents a conceptual challenge that does not enable experts and policy makers to clearly identify its scope of application, its positive and negative consequences.

As Stone et al. (2016) argue, if society approaches AI with fear and suspicion, “missteps that slow AI’s development or drive it underground will result, impeding important work on ensuring the safety and reliability of AI technologies” (Stone et al., 2016, p.298). This is particularly problematic in a world where AI is increasingly used in everyday life, including processing
sensitive information for public health. Hence, politicians, regulators, and civil society must acquire a better understanding of this technology (Al-Amoudi & Latsis, 2019) and the associated hopes and concerns it triggers. One must also recommend applying the precautionary principle when the concerns and threats are not fully evaluated and addressed.

First, AI can be considered a field, a discipline, or a science. As McCarthy (1998) states, AI “is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable” (p.2). It can be considered an umbrella term to define a large number of scientific and technological advances. For instance, the Council of Europe (COE) considers AI as “a young discipline of about sixty years, which brings together sciences, theories and techniques (including mathematical logic, statistics, probabilities, computational neurobiology and computer science) and whose goal is to achieve the imitation by a machine of the cognitive abilities of a human being” (Council of Europe, n.d.). First, AI consists of a large area of study in the field of computer science (Fanni, Gabelloni, Alberich-Bayarri, & Neri, 2022). Second, AI is often compared to human capacity. Here, the discussion is about whether AI can mimic a human brain, where it is dedicated to the development of computers capable of engaging in human-like thought processes, including learning, reasoning, and self-correction (Kok, Boers, Kosters, Van der Putten, & Poel, 2009). The relation between AI and human capacity is also well illustrated by the Alan Turing test. This is considered a reliable first test to recognize AI, perhaps because everyone has the ability to conduct one:

> It is quite simple. We place something behind a curtain, and it speaks with us. If we can’t make difference between it and a human being, then it will be AI. However, this definition is not formal. Another problem is that this definition does not separate the knowledge from the intellect. (Dobrev, 2004)

AI is indeed often described in relation to human intelligence, or intelligence in general. Indeed, many definitions refer to machines that behave like humans or perform actions that require some form of intelligence (Russel & Norvig, 2010; McCarthy, 2007; Nilsson, 1998; Fogel, 1995; Albus, 1991; Salin & Winston, 1992; McCarthy, 1988; Gardner, 1987, 1983; Newell & Simon, 1976; Bellman, 1978; Minsky, 1969; McCarthy, Minsky, Rochester, & Shannon, 1955/2006). However, these definitions remain vague because of the difficulty of defining and measuring intelligence itself. Thus, this type of definition proposes an ideal target rather than a concrete and measurable research concept.
Third, AI performs a wide range of activities, including “verbal-linguistic, visual-spatial, logical-mathematical, naturalistic, and interpersonal intelligence” (Monett, Lewis, & Thórisson, 2020, p.19). Because AI can “assume some capabilities normally thought to be like human intelligence such as learning, adapting, self-correction” (Mitchell, 2019), it requires an understanding of it through the prism of multiple intelligences: “intelligence is the capacity of an agent to use computation, intended as the capacity to link perception to action in multiple possible sophisticated ways, to increase biological fitness or to accomplish goals” (Monett, Lewis, & Thórisson, 2020, p.19). This leads Wang (2019) to argue that “every working definition of AI corresponds to an abstraction of the human mind that describes the mind from a certain point of view, or at a certain level of abstraction, under the belief that it is what intelligence is really about” (Wang, 2019, p.19).

The many applications in all areas of private and professional life comes from the fact that this technology is a general or foundational technology, just like electricity. The issues associated with this technology are therefore very different from one field to another. It is necessary to distinguish between AI and computer systems that also support human intelligence. Wang (2019) defines AI as “the capacity of an information-processing system to adapt to its environment while operating with insufficient knowledge and resources” (p.17). In this working definition, he highlights the combination of two essentials that help distinguish between AI and computer systems: information processing and adaptation. He argues that the information processing capacity of AI consists of choosing and executing tasks, and adjusting its behavior according to its past experiences (Wang, 2019). This operational definition that focuses on technical aspects is close to the one published in February 2020 in the EU White Paper on AI, which describes AI as “a collection of technologies that combine data, algorithms and computing power” (European Commission, 2020). These definitions are consistent with the commonly used definition of AI as “the study of the computations that make it possible to perceive, reason, and act” (Winston, 1992, p.1).

A broader definition is offered by the OECD, which refers to AI as “A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments” (OECD, 2019, p.15). Although very useful, his definition focuses on these two aspects. Thus, it might be too limited to identify the promises and pitfalls of AI in the policy-making process.
In their analysis of the definitions of AI in scientific and gray literature, Samoili et al. (2020) identified four aspects of AI that could be considered as four main features of AI:

- Perception of the environment, which takes into account the complexity of the real world (HLEG, 2019; Nakashima, 1999; Nilsson, 1998; Poole, Mackworth, & Goebel, 1998; Fogel, 1995; Wang, 1995; Albus, 1991; Newell & Simon, 1976).
- Decision-making, which includes reasoning and learning: taking actions, performing tasks, as well as adapting and reacting to changes in the environment with some level of autonomy (HLEG, 2019; OECD, 2019; Kaplan & Haenlein, 2019; Nilsson, 1998; Poole, Mackworth, & Goebel, 1998; Fogel, 1995; Wang, 1995; Albus, 1991; Newell & Simon, 1976).
- Achieving specific goals: this is regarded as the ultimate reason for the existence of AI systems (HLEG, 2019; OECD, 2019; Kaplan & Haenlein, 2019; Poole, Mackworth, & Goebel, 1998; Fogel, 1995; Albus, 1991; Newell & Simon, 1976).

These key aspects of AI can be found in the operational definition proposed by the High-Level Expert Group on Artificial Intelligence (HLEG):

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. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications) (HLEG, 2018, p.1)

This definition includes indeed the four aspects of AI mentioned previously and identified in the literature. However, it is quite technical. Since this book adopts a social science perspective, the following definitions are most adapted to the objective of this publication:

AI is a generic term that refers to any machine or algorithm that is capable of observing its environment, learning, and based on the knowledge and experience gained, taking intelligent action or proposing decisions. There are many different technologies that fall under this broad AI definition. At the moment, ML4 techniques are the most widely used. (Craglia et al., 2018, p.18)
As discussed in this section, AI is this foundational technology that remains difficult to define precisely. It remains an ongoing project with past, present and future developments. In the definition chosen as a reference for this book, Craglia et al. (2018) highlight key aspects of the agency of AI: its capacity to observe its environment, learn from it, and take smart action or propose decisions. However, this definition also highlights the fact that many technologies fall under the term “AI.” In other words, AI remains a conceptual challenge, which makes its understanding and adoption by policy-making stakeholders more challenging as well. Moreover, this conceptual challenge is combined with a level of agency unprecedented in other technologies as discussed in the next section.

ALGOCRATIC SYSTEM AND AUTONOMOUS TASKS PERFORMED BY AI

As mentioned previously, AI consists of three elements: data, algorithms and computing power. Said differently, an AI system performs three main tasks. First, it collects data from the environment through sensors: it perceives real and/or virtual environments. Second, it builds an abstract model of its environment. Lastly, it produces an output (e.g. recommendations, predictions or decisions).

The environment describes the space that the AI system can observe through perceptions (via sensors) and influence through actions (via actuators). At the core of an AI system lies an abstract representation of the external environment, whether it is a virtual or a real-world environment. This model consists of a set of algorithms (i.e. a set of rules) that represent the structure and/or dynamics of the environment (OECD, 2019). This AI model can be automatically built (also called model building), which means that new data it is fed with improves the precision of its representation of the world (e.g. ML algorithms). It can also be built by human operators and be based on expert knowledge. The model is built according to the type of output it is expected to generate (i.e. objective of the AI system) and performance measures (e.g. accuracy, resources for training, representativeness of the dataset).

This model allows humans and/or automated tools to derive an outcome such as recommendations, predictions or decisions (also called model inference). In this phase, the AI interprets the raw data collected in relation to the model it has of the environment. Sometimes, the interpretation process can be understood and explained by experts. In some cases, it cannot. This issue is called the black box phenomenon. According to the representativeness of the dataset and the accuracy of the model, the interpretation process will be precise and valid. It can produce one recommendation (e.g. deterministic rules), or several ones (e.g. probabilistic models) (OECD, 2019).
Autonomous driving systems illustrate well how an AI system functions. At its core, the AI model is built from large datasets (e.g. historical driving data, driving rules) and with pre-determined objectives (bring the car safely to a specific destination). Thanks to this abstract representation of the reality, the AI system can (1) perceive its environment (e.g. through sensors such as cameras), (2) make abstract this input data and incorporate it into its model (e.g. object recognition, and location data), and (3) make recommendations in terms of possible short-term futures (model inference). These recommendations have an impact of the environment (e.g. the car accelerates or stops).

Machine Learning (ML) is part of the statistical approach. It is a set of techniques that allow machines to learn in an automated way by detecting and deducing patterns in large datasets rather than by explicit rules and instructions created by a developer. ML approaches frequently teach machines to produce a result by showing them many examples of accurate results. ML can also set a collection of rules and let the machine learn by trial and error. ML can reveal helpful facts to build or adapt an AI model, but it can also be used to interpret the results of an AI model (see Figure 1.1).

ML includes many techniques such as neural networks. The main technology that has enabled the current wave of ML applications is a sophisticated statistical modeling technique called “neural networks.” Its deployment is made possible by the constant increase in computational power and the availability of large datasets (also known as “big data”). Neural networks involve the repeated interlinking of thousands or millions of simple transformations into a larger statistical machine, capable of learning complex relationships between inputs and outputs. In short, neural networks change their own code to identify and optimize relationships between inputs and outputs. Deep learning is a sub-category of neural networks. This term refers to especially large neural networks; there is no specific marker to define when a neural network becomes “deep” (OECD, 2019).

In their “AI Watch: Defining Artificial Intelligence” report, Samoili et al. (2020) from the EU Joint Research Centre built a useful AI taxonomy that takes into account political, research and industrial perspectives. It is divided into two main categories: core AI capabilities (e.g. computer vision) and transversal topics (e.g. ethical considerations). As mentioned earlier, AI capabilities can be grouped into two broad categories: (a) reasoning and decision-making, and (b) learning and perception (Table 1.1). The first group includes the transformation of data into knowledge (i.e. transforming data from the real world into information that is understandable and usable by machines), and in so doing, enabling them to make decisions. It includes the AI domains of Reasoning (often through symbolic AI) and Planning. The second group (statistical AI) includes the ability to learn – that is, the ability to extract information and solve problems from the perception of structured or unstructured
Artificial intelligence and democracy

Figure 1.1 Different approach to AI

Data (written and oral language, image, sound, etc.). It also includes the ability to adapt and react to changes and make behavioral predictions among others. It comprises of the domains of learning, communication and perception. In this taxonomy, the categories and sub-categories are related not disjunct. For instance, machine learning is used in computer vision, audio processing and natural language processing (Samoili et al., 2020).

ML and AI are both constituted of sets of algorithms. Algorithms are automated instructions, or step-by-step instructions to process inputs into outputs (Stone, 1972). Today, most algorithms consist of an aggregate of numerous algorithms that function as a computer program (Sandvig, 2014). As Osoba and Welser IV (2017) argue, an algorithm can be defined as “a computable function that can be implemented on computer systems. Machine learning algorithms can also update their behavior in response to experience (input data) and performance metrics” (Osoba and Welser IV, 2017 cited in European Commission, 2020).

Algorithms are now used to govern many aspects of our society and economy (Janssen & Kuk, 2016) as argued by the Committee of Experts MSI-AUT in the 2018 Draft Recommendation of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems entitled “Addressing the Impacts of Algorithms on Human Rights” (Council of Europe, 2018):

Applications that, often using mathematical optimisation techniques, perform one or more tasks such as gathering, combining, cleaning, sorting, classifying and inferring (ed. personal) data, as well as selection, prioritisation, recommendation and decision-making. Relying on one or more algorithms to fulfil their requirements in the settings in which they are applied, algorithmic systems automate activities in a way that allows the creation of adaptive services at scale and in real time.
### Table 1.1 AI taxonomy

<table>
<thead>
<tr>
<th>Group 1: Transforming data into knowledge.</th>
<th>Reasoning</th>
<th>Ability to represent information about the world in a form that a computer system can employ to resolve complex tasks such as diagnosis of a medical condition or a human oral dialogue (also called Knowledge Representation and Reasoning or KR², KR&amp;R)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Planning</td>
<td>Ability to solve planning and scheduling problems (i.e. design and execute a sequence of actions where each action has its own set of preconditions to be satisfied before performing the action). Typically performed by intelligent agents, autonomous robots and vehicles.</td>
</tr>
<tr>
<td>Group 2: Learning, adapting and reacting to change.</td>
<td>Learning</td>
<td>Ability to automatically learn, decide, predict, adapt to changes and improve without being explicitly programmed.</td>
</tr>
<tr>
<td></td>
<td>Communication</td>
<td>Ability to identify, process, understand and/or produce information in human communications (e.g. text).</td>
</tr>
<tr>
<td></td>
<td>Perception</td>
<td>Ability to be aware of its environment through sensors.</td>
</tr>
</tbody>
</table>

**Source:** Adapted from AI Watch taxonomy, Samoili et al. (2020), p.17.

This form of delegation of authority to an algorithm that has a decision-making capacity and autonomy is well captured in the notions of “algorithmic regulation” (Yeung, 2017, 2018), “algorithmic governmentality” (Rouvroy, 2015) and algocracy, or algocratic system, originally coined by sociologist Aneesh (2006, 2009), which describes: “[a] governance system in which computer coded algorithms structure, constrain, incentivize, nudge, manipulate or encourage different types of human behavior” (Danaher, forthcoming). Hence, this form of governing with AI (i.e. using AI for public action) presents benefits and risks as discussed in the following section.

#### GOVERNING WITH AI: BENEFITS AND RISKS OF AI FOR PUBLIC ACTION

The development of new AI-based initiatives to improve public service delivery is part of an older research tradition. Already in the 1990s, the internet and computer technology helped transform paper-based processes to fully digi-
Artificial intelligence and democracy

tized processes and services available online 24/7. The rapid development of internet access and the fast adoption of social media platforms in many liberal democracies generated a growing demand among populations to digitize public action (de Feraudy, 2019). Consequently, governments developed new online services (i.e. e-government) and fostered participation of citizens in some decision-making processes through digital technologies (i.e. e-participation and civic tech).

More recently, governments used artificial intelligence in their relationship with citizens. As Sharma, Yadav, and Chopra (2020) argue, “[w]ith rapid digital technological change, it is inevitable for the government to innovate its traditional methods in order to achieve better citizen engagement, accountability, and interoperability (…).” AI are increasingly used in the fields of healthcare, education, social and cultural services. Moreover, AI can contribute to improving the efficiency and inclusiveness of the policy-making process through optimizing decision-making processes, data and opinion mining, game theory, and agent-based simulation (Milano, O’Sullivan, & Gavanelli, 2014).

These capabilities and applications could also play a significant role in various governmental tasks related to policy making. For example, and based on the evidence gathered from the case studies reported herein, an early data intelligence exercise can assist public decision makers in detecting emergent societal problems or citizens’ concerns much promptly, enabling more timely and accurate policy responses. (Misuraca & van Noordt, 2020, p.19)

Data plays an increasing role in the delivery of public services. Martens (2018) identified three phenomena to consider: the automation and lowering of data collection costs (price effect), the massive increase of available data (quantity effect), and the shift of many face-to-face human activities to the digital domain (substitution effect), have put data sharing at the heart of modern public services and allowed for more efficient and cost-effective delivery. There are both benefits and risks to sharing data. On the one hand, sharing allows for the discovery of new information through the linking of previously unconnected data. On the other hand, thanks to the data collected, it is easier to both know and respond to the needs of the population by adapting the services offered, and to evaluate them. On the other hand, data sharing has its dangers. These include the risk of losing some or all of the data; the possibility of identifying an individual through the combination of many data sources, despite the anonymization of the data; and some negative impacts from the reuse of the data in other contexts to which the owner did not want it disclosed (Involve UK, 2017).

AI research has mainly focused on the governance of AI and to a lesser extent governing with AI. AI differs from previous waves of technology transformation in governments and public organizations. Indeed, AI not only
has the ability to make information available due to its superior computational
capacity, but as mentioned previously, it also has the ability to make decisions
in place of humans (Latzer & Just, 2020). When AI is further deployed in
organizations, this decision-making power can then fundamentally influence
how governments and public administrations govern and provide services
to citizens (Engstrom, Ho, Sharkey, & Cuéllar, 2020; Mehr, Ash, & Fellow,
2017). Misuraca and van Noordt, in their AI Watch report for the EU Joint
Research Centre present a useful taxonomy of AI uses by government. It is
based on prior research including Wirtz, Weyerer, and Geyer (2019). This
taxonomy is particularly useful because it allowed the two researchers to
develop a mapping of AI uses by governments in Europe (EU, UK, Norway
and Switzerland).

As shown in Table 1.2, AI presents many benefits for governments to
increase the efficiency and effectiveness of their operations and services, such
as:

- Improving the knowledge management capacity (e.g. assist in the browsing
  and finding of relevant data in Slovakia);
- Mapping and predicting risks (e.g. predicts burglaries in Switzerland);
- Automatizing data collection and analysis (e.g. process satellite imagery
  in Estonia);
- Automatizing data collection and analysis (e.g. process satellite imagery
  in Estonia); some services (e.g. self-driving snowploughs in Norway),
decision-making (e.g. nursery child recruitment system used in Warsaw),
and the communication with citizens (e.g. Chatbot to answer frequently
asked questions in Latvia).

Although this chapter is not focusing on e-participation and the use of tech-
nology to foster the inclusion of civil society in policy making, some of these
initiatives also contribute to putting the citizen back at the center of the design
of services rendered by the government, such as Natural Language Processing
(e.g. AI system to detect the most pressing concerns on Twitter in Ireland). In
addition, initiatives that improve the effectiveness of government action may
also have the side effect of increasing trust in the capacity of government.

Evidently, citizens benefit strongly from more efficient and effective public
action. However, prior research has questioned the real benefits of digitization
of government operations and services. As Bannister and Connolly (2020)
argue, the promises of digital technologies far exceed the reality and expect-
tations of users (Bannister and Connolly, 2020). Misuraca, Codagnone, and
Rossel (2013) and Savoldelli, Codagnone, and Misuraca (2014) have even
questioned the merits and real impact of the massive investments in digital that
governments have made in recent decades. To what degree do they improve
Table 1.2  Current and prospective technologies and uses

<table>
<thead>
<tr>
<th>Type of application</th>
<th>Tasks / Objectives</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KNOWLEDGE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI-empowered Knowledge Management</td>
<td>To create a searchable collection of case descriptions, texts and other insights to be shared with experts for further analysis.</td>
<td>In Slovakia, an AI system is used in the government to assist in the browsing and finding of relevant semantic data.</td>
</tr>
<tr>
<td>Machine Learning, Deep Learning</td>
<td>While almost all the other categories of AI use some form of Machine Learning, this residual category refers to AI solutions which are not suitable for the other classifications.</td>
<td>In Czechia, AI is used in social services to facilitate 17 citizens to stay in their natural environment for as long as possible.</td>
</tr>
<tr>
<td>Predictive Analytics, Simulation and Data Visualization</td>
<td>To identify patterns in data that are consequently used to visualize, simulate or predict new configurations.</td>
<td>Since 2012, the Zurich City Police have been using software that predicts burglaries. Based on these predictions, police could be forwarded to check these areas and limit burglaries from happening.</td>
</tr>
<tr>
<td><strong>LEARNING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer Vision and Identity Recognition</td>
<td>Image, video or facial recognition to gain information on the external environment and/or the identity of persons and objects.</td>
<td>In Estonia, the SATIKAS system is in use which is capable of detecting mowed (or the lack of mowed) grasslands on satellite imagery.</td>
</tr>
<tr>
<td>Audio Processing</td>
<td>To detect and recognize sound, voices, music and other audio inputs, thus enabling the transcription of spoken words.</td>
<td>Corti in Denmark is used to process the audio of emergency calls in order to detect whether the caller could have a cardiac arrest.</td>
</tr>
<tr>
<td><strong>PERCEPTION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Security Analytics and Threat Intelligence</td>
<td>To analyze and monitoring security information and to prevent or detect malicious activities.</td>
<td>In the Norwegian National Security Authority a new system is used based on machine learning. It is enabling the automatic analysis of any malware detected to improve cybersecurity.</td>
</tr>
</tbody>
</table>
### Type of application

<table>
<thead>
<tr>
<th>Tasks / Objectives</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chatbots, Intelligent Digital Assistants, Virtual Agents and Recommendation Systems</td>
<td>To provide generic advice and behavior-related recommendations to users. In Latvia, the Chatbot UNA is used to help answer frequently asked questions regarding the process of registering a company.</td>
</tr>
<tr>
<td>Natural Language Processing, Text Mining and Speech Analytics</td>
<td>To recognize and analyze speech, written text and communicate back. In Dublin, an AI system analyzes citizen opinions in the Dublin Region for an overview of their most pressing concerns by analyzing local Twitter tweets with various algorithms.</td>
</tr>
<tr>
<td>Expert and Rule-based Systems, Algorithmic Decision-Making</td>
<td>To facilitate or fully automate decision-making processes of potential relevance. Nursery child recruitment system used in Warsaw. The algorithm considers data provided by parents during the registration, calculates the score and automatically assigns children into individual nurseries.</td>
</tr>
<tr>
<td>Cognitive Robotics, Process Automation and Connected and Automated Vehicles</td>
<td>To automatize a process, which can be achieved through robotized hardware or software. The use of self-driving snowploughs in an airport in Norway in order to improve the clearing of snow on runways.</td>
</tr>
</tbody>
</table>

**Source:** Adapted from AI Watch (Misuraca & van Noordt, 2020).

The efficiency and effectiveness of government operations and services? Hence, there is a need for auditing the benefits of these burgeoning AI uses in terms of efficiency and effectiveness. Due to the early stages of AI adoption, it is still a difficult task to endeavor.

As mentioned previously, AI presents a conceptual challenge. It remains challenging for many citizens and policy makers to grasp what is AI, what it does, and what its potential benefits and risks are (Duberry & Hamidi, 2021). This is true for AI and for digital technologies at large. The digital divide remains an important issue even in Europe. Indeed, a part of the population is less computerized and less connected than others, in particular the elderly, the countryside, or women. It is both about access to digital technologies and digital skills. In this context, governments’ investments to digitize their services can also lead to the exclusion of that part of the population that does not have access or digital skills.

Moreover, AI is characterized by an unprecedented level of agency by structuring, constraining, nudging, and encouraging different types of human behavior (i.e. algocracy, cf. Danaher, forthcoming). To ensure the trust of citizens in this new technology, governments need to ensure that AI-mediated governmental processes and decision-making are transparent and accountable. In other words, citizens need to be able to understand how the decision was made.
made, and to be able to appeal to the decision. The black box phenomenon describes the difficulty, even for programmers, of dissecting the precise operation of an AI system and more specifically how it arrives at a decision or choice. Since it is constantly adapt its code according to the data provided, its decision process can be particularly difficult to decipher, and just as difficult to audit. This is all the truer since an AI system can be composed of several algorithms. This poses great challenges in terms of legitimacy, transparency, and accountability for decisions made using an AI (Annoni et al., 2018).

A second major concern is the issue of bias in the data used by AI. ML-based applications learn from data. If bias exists in that data, the algorithm will replicate or even reinforce it (Wirtz, Weyerer, & Geyer, 2019). This is particularly the case for historical data, which would lead the algorithm to base itself on a period of history where certain discriminations were widespread, and thus ultimately contribute to reinforcing them. In terms of data, AI can put privacy at risk, especially when the data collected is not voluntarily shared by citizens. This is particularly the case for metadata. This is also the case for sensitive information (e.g. sexual orientation, a health condition) that is inferred from public, non-sensitive data, potentially leading to discriminatory treatment (Floridi, 2017).

Hence, governments face a dilemma. On the one hand, they are asked to improve the performance of their processes and services. To do this, AI presents many opportunities as discussed above. However, they also have the role to protect citizens against the risks that AI presents. It is thus a matter of government responding to two simultaneous demands: governing with AI and governing AI. The following section will discuss some of the main propositions to govern and regulate AI in the context of governmental use.

EFFORTS AND CHALLENGES TO REGULATE AND GOVERN AI

In recent decades, the progressive digitization of internal processes and public services, as well as the gradual privatization of certain activities previously handled exclusively by the public sector, have required the development of standards and principles of good governance specific to the public sector. The use of AI by the public sector requires a framework of specific standards, principles and values. OECD-SIGMA, in collaboration with the European Commission, has developed a set of principles and a methodological framework for assessing good governance in public administrations.

The values promoted by the European Union (EU) in terms of public service are found in various documents. First, Article 2 of the Treaty on European Union specifies the fundamental values on which the Union is founded. It describes “a society in which pluralism, non-discrimination, tolerance, justice,
solidarity and equality between women and men prevail” (EU, 2008). In addition, there are the rights and freedoms defined by the Charter of Fundamental Rights of the European Union, which only applies when Member States directly implement European regulations or transpose them into their national legislation. In the context of this discussion, we can consider the right to data protection and the right to good administration: “Every person has the right to have his or her affairs handled impartially, fairly and within a reasonable time by the institutions, bodies, offices and agencies of the Union” (Article 41 of the EU Charter of Fundamental Rights).

Specifically for AI, guidelines and other documents developed by universities, think tanks and governmental and non-governmental organizations also provide a number of principles to guide the adoption of AI by governments and public administrations, such as the EU White Paper on AI or the OECD AI principles. The latter proposes five principles that should guide the adoption of AI (OECD, 2020):

- Inclusive growth, sustainable development and well-being
- Human-centered values and fairness
- Transparency and explainability
- Robustness, security and safety
- Accountability.

The OECD AI principles also include five recommendations for policy makers:

- Investing in AI research and development
- Fostering a digital ecosystem for AI
- Shaping an enabling policy environment for AI
- Building human capacity and preparing for labor market transformation
- International cooperation for trustworthy AI.

There are also negative requirements to frame the use of AI in the public sector. The objective of these requirements is indeed to reduce the risk of negative consequences of certain AI applications in public service provision. For example, they require that the AIs used do not have a bias and do not discriminate against a part of the population or a category of citizens. It can also mean requiring a maximum error rate when a government or public administration has delegated a decision-making process to an AI system. These requirements respond to the precautionary principle, in particular for known or anticipated risks. An additional risk with AI comes from its great diversity of applications, and its constant evolution. In other words, it is very difficult today to envisage all the potentially negative consequences, direct and indirect, of using AI in the public sector. This requires principles and standards that are flexible enough to adapt to future situations that are still unknown.
The legal and regulatory framework is of paramount importance to understanding the use of AI in public services. Over the past few decades, a large number of legal and policy tools have been developed to address the growing prominence of AI in the lives of citizens, and in particular the labor market (Frey & Osborne, 2017), health (Jiang et al., 2017), and human rights protection (Eubanks, 2018). AI governance can be defined as “rule-making around algorithms that process data” (Misuraca & van Noordt, 2020, p.49).

The responsibilities associated with the processing of personal data do not fall solely on public sector organizations. When public sector organizations use technologies developed by companies, they make themselves vulnerable to the risk of abuse by those companies, either intentionally or through negligence. Calls for further regulation of online platforms, even after the adoption of the GDPR, demonstrate the growing concern of citizens and civil society organizations about the management and processing of their personal data by technology companies.

Selbst, Boyd, Friedler, Venkatasubramanian, and Vertesi (2019) have shown that efforts to make machine learning algorithms fair (i.e. to ensure that there is no bias or hidden discrimination in the algorithm) tend to “render technical interventions ineffective, inaccurate, and sometimes dangerously misguided when they enter the societal context that surrounds decision-making systems.” They identified five different traps or “failing to properly account for or understand the interactions between technical systems and social worlds” (p.59).

- The Framing Trap: “Failure to model the entire system over which a social criterion, such as fairness, will be enforced” (p.60);
- The Portability Trap: “Failure to understand how repurposing algorithmic solutions designed for one social context may be misleading, inaccurate, or otherwise do harm when applied to a different context” (p.61);
- The Formalism Trap: “Failure to account for the full meaning of social concepts such as fairness, which can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms” (p.61);
- The Ripple Effect Trap: “Failure to understand how the insertion of technology into an existing social system changes the behaviors and embedded values of the pre-existing system” (p.62);
- The Solutionism Trap: “Failure to recognize the possibility that the best solution to a problem may not involve technology” (p.63).

In terms of AI regulation, the current trend is to treat AI as a technology so specific and unique that it does not fit into existing governance structures, public policies, and laws. A number of organizations and governments have therefore seen the need to produce specific recommendations, strategies and other guidelines for this technology. Their approach, for the most part, shows that they
perceive current norms and governance as inadequate for AI. However, this siloed approach is risky. As Misuraca and van Noordt (2020) argue:

it would make an enormous difference to think of AI governance as an extension of data protection and competition regulations, acting hand in hand to reduce harms and secure human dignity. Such effort – instead of happening in a vacuum – would help update major existing regulations (i.e. GDPR) to make them work where they do not: by addressing massive imbalances in power, advancing data portability and privacy by design or securing EU wide, public digital infrastructure (p.49).

For example, many existing regulatory frameworks and standards could be applied to AI and its externalities, such as antitrust and consumer protection measures, ethics guidelines, data protection enforcement, intellectual property (IP) protection standards and rules to name just a few. Similarly, both the German Bundeskartellamt and the French Competition Authority in 2019 deemed existing competition laws sufficient to address the challenges posed by the widespread use of AI, and in particular pricing algorithms (Bundeskartellamt & Autorité de la concurrence, 2019).

AI governance is primarily composed of voluntary ethical codes and guidelines. Fjeld, Achten, Hilligoss, Nagy, and Srikumar (2020) mapped these ethical codes developed around the world and identified eight major cross-cutting themes present in most documents they analyzed:

1. Privacy: respect citizens’ privacy, both in terms of what type of data is being processed, and in terms of ensuring citizens agency over their personal data;
2. Accountability: existence of accountability mechanisms for the externalities of AI systems;
3. Safety and Security: ensuring that the AI system does not present any vulnerability;
4. Transparency and Explainability: the AI system allows for audit and oversight, including how decisions are made, and where, when, and how they are being used;
5. Fairness and Non-discrimination: ensuring that the design of AI systems and their usage is done according to fairness and inclusivity principles;
6. Human Control of Technology: all important decisions taken by the AI system stay under human review;
7. Professional Responsibility: all professionals and experts involved in the design and maintenance of AI systems follow principles of professionalism and integrity, including the involvement of the stakeholders potentially affected by the AI system;
8. Promotion of Human Values: The purposes to which AI is dedicated, and the means by which it is deployed, must be consistent with our fun-
Artificial intelligence and democracy

When adopting new technologies, and especially when the new technology is not mature in its development, early adopters may face mistakes, which then may jeopardize the confidence of later adopters in the technology (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). It is therefore essential to ensure that AI is not adopted too prematurely for tasks associated with policy making.

**CONCLUDING REMARKS**

As discussed in this chapter, governments have progressively adopted a number of technology innovations to respond to a growing demand to (1) digitalize public action and optimize its operations and services (de Feraudy, 2019), and (2) increase citizen engagement in the development, implementation, and evaluation of public policies (de Feraudy & Saujot, 2017).

AI is increasingly deployed by governments to automate and analyze large datasets, enabling the optimization and support of existing processes and services. And yet, this technology remains unregulated and with specific characteristics, which imply a higher degree of uncertainty and risk in the citizen–government relation. AI is indeed this blurry (i.e. conceptual challenges), variable (i.e. ongoing developments and applications), often opaque (i.e. black box phenomenon) agent in the citizen–government relation with various degrees of agency (i.e. capacity to observe its environment, learn from it, and take smart action or propose decisions). In order to ensure that the benefits of this new technology are shared equally, inclusively and transparently among all parts of the population, the need to adopt a human-centric approach to AI as defined by the EU is all the more crucial:

The Commission has developed key principles to guide the European approach to AI that take into account the social and environmental impact of AI technologies. They include a human-centric way of developing and using AI, the protection of EU values and fundamental rights such as non-discrimination, privacy and data protection, and the sustainable and efficient use of resources. (EU, 2021)

As discussed in the next chapter, conventional forms of citizen participation tend to be in decline, whereas non-conventional forms of participation (i.e. social movements and street protests) have grown significantly in the last decades. Scholars and experts raise a flag about the increase of distrust among citizens toward different forms of public authority, including governments. This is to say that the citizen–government relationship is fragile, and a foundation of any liberal democracy. In this context, one can only recommend...
adopting forms of AI where principles of equality, freedom, human rights, and the notion of popular sovereignty are integrated in the technology “by design”.

REFERENCES


Artificial intelligence and democracy

.deciderensemble.com/articles/47158-etude-cartographie-de-la-civic-tech-en-france
[Accessed 21 August 2021].


AI to optimize the effectiveness and efficiency of public services


Sandvig, C. (2014). Seeing the sort: The aesthetic and industrial defense of “the algorithm.” *Journal of the New Media Caucus*, ISSN, 017X.


