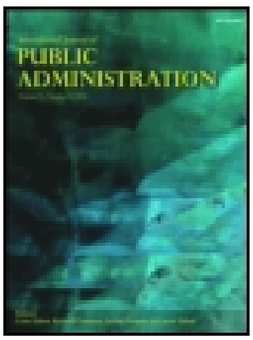

International Journal of Public Administration

ISSN: 0190-0692 (Print) 1532-4265 (Online) Journal homepage: http://www.tandfonline.com/loi/lpad20

Artificial Intelligence and the Public Sector—Applications and Challenges

Bernd W. Wirtz, Jan C. Weyerer & Carolin Geyer

To cite this article: Bernd W. Wirtz, Jan C. Weyerer & Carolin Geyer (2018): Artificial Intelligence and the Public Sector—Applications and Challenges, International Journal of Public Administration, DOI: 10.1080/01900692.2018.1498103

To link to this article: https://doi.org/10.1080/01900692.2018.1498103

Published online: 24 Jul 2018.

INTERNATIONAL JOURNAL OF PUBLIC ADMINISTRATION
https://doi.org/10.1080/01900692.2018.1498103

Routledge
Taylor & Francis Group

Artificial Intelligence and the Public Sector—Applications and Challenges

Bernd W. Wirtz, Jan C. Weyerer, and Carolin Geyer

German University of Administrative Sciences Speyer, Chair for Information and Communication Management, Speyer, Germany

ABSTRACT

Advances in artificial intelligence (AI) have attracted great attention from researchers and practitioners and have opened up a broad range of beneficial opportunities for AI usage in the public sector. Against this background, there is an emerging need for a holistic understanding of the range and impact of AI-based applications and associated challenges. However, previous research considers AI applications and challenges only in isolation and fragmentarily. Given the lack of a comprehensive overview of AI-based applications and challenges for the public sector, our conceptual approach analyzes and compiles relevant insights from scientific literature to provide an integrative overview of AI applications and related challenges. Our results suggest 10 AI application areas, describing their value creation and functioning as well as specific public use cases. In addition, we identify four major dimensions of AI challenges. We finally discuss our findings, deriving implications for theory and practice and providing suggestions for future research.

KEYWORDS

Artificial intelligence; public sector; AI applications; AI challenges

Introduction

Artificial intelligence (AI) is an interdisciplinary research field that has recently gained special importance in society, economics, and the public sector, opening up a variety of new opportunities (Boyd & Wilson, 2017). In particular, the self-learning algorithm serves as a basis for recent AI innovations (Holdren & Smith, 2016) and has a major impact in different sectors. Digitization is significantly changing the working environment, especially in the business sector, showing that the relevance of AI is evident (Castro & New, 2016).

The McKinsey Global Institute expects that "rapid advances in automation and artificial intelligence will have a significant impact on the way we work and our productivity" (Gaurav, Queirolo, & Santhanam, 2018, p. 4). Within the next 15 years, AI has the potential to increase the annual economic growth rates in the United States, Germany, and Japan by up to 2 percent points (Purdy & Daugherty, 2016). Accordingly, AI has the ability to bring change and benefits to the public and the private sector. More specifically, virtual workforces, so-called intelligent automation, as well as labor and capital augmentation where AI can complement the skills of existing workforces lead to cost efficiency and savings (Bataller & Harris, 2016). In this connection, Purdy and Daugherty (2016) state that AI can be seen "as a capital-labor hybrid [and] can replicate labor activities at much greater scale and speed, and to even perform some tasks beyond the capabilities of humans" (Purdy & Daugherty, 2016, p. 5).

In light of these great potential benefits, there is a huge increase in AI development through technology companies such as Google, Facebook, IBM, and Microsoft which make great investments to explore its potential (Grosz et al., 2016). Recently launched innovations like Google's AlphaGo (Byford, 2016), IBM's Watson, Apple's Siri, and Amazon's Echo with Alexa attract great attention in society (Power, 2016). For instance, AlphaGo Zero, the latest evolution of the AI technique AlphaGo, constitutes a technological milestone by achieving superhuman performance. Using self-play reinforcement learning, AlphaGo Zero is able to learn from simply playing against itself instead of being trained by humans and thus is no longer limited by human knowledge (Silver et al., 2017). These technological advancements in AI and the associated value potential are also increasingly gaining in importance in the context of government. In particular, China and the United States have recognized the value of AI for the public sector and their competitiveness in the global economy. For example, the government of China has shown an extraordinary commitment to catch up with the AI development of the Western world in the next 3 years (Knight, 2017a).

The State Council of the People's Republic of China has issued a guideline with a total investment of $147.8 billion to become a global innovator in the

CONTACT Bernd W. Wirtz Ls-wirtz@uni-speyer.de German University of Administrative Sciences Speyer, Chair for Information and Communication Management, Freiherr-vom-Stein-Str. 2, Speyer 67346, Germany.

field of AI by 2030 (State Council of the People's Republic of China, 2017). The United States is preparing for the AI future by promoting AI researchers and respective AI education programs. In 2016, for instance, they have spent approximately $1.2 billion on research and development of AI-related technologies (Holdren & Smith, 2016).

Likewise, Europe has spent up to €700 million on robotics and public–private partnerships in the context of AI (Ansip, 2017). Consequently, many AI-based applications for the public sector are emerging, promising great value with regard to workforce and productivity, but also bringing along significant challenges that are crucial to their successful implementation and use. In this connection, pressing challenges can be particularly observed with regard to AI responsibility and implementation as well as social and ethical issues (Purdy & Daugherty, 2016; Quraishi, Wajid, & Dhiman, 2017; Ransbotham, Kiron, Gerbert, & Reeves, 2017), which potentially threaten successful AI use and respective creation of value for the public sector and society as a whole.

Despite increasing investments in AI research and a growing number of research contributions, AI for public usage is still a young field of research that falls short in describing associated applications and challenges. This shortcoming also results from the lack of a common definition and understanding of AI in science (Scherer, 2016). Therefore, it is of greater importance to first clarify the meaning of AI, before identifying and describing AI application areas and related challenges in the public sector. This article aims to fill in this research gap by establishing a common definition of AI and providing an integrated overview of applications and challenges of AI in the public sector. In doing so, we first present the current state of research on AI in the public sector context, which serves as a basis for the subsequent elaboration of an integrated AI definition. Following this, we identify AI application areas and associated challenges within the public sector, finally deriving implications for research and public management.

Literature review of AI public sector research

In order to create a literature basis with articles from academic journals, we have focused on peer-reviewed English-language articles in. Against this background, we conducted a title and abstract search in the following EBSCOhost databases: Academic Search Complete Business Source Complete and EconLit with Full Text. In order to create a profound basis of AI research in a governmental context with a focus on applications and challenges, we used keywords like "artificial intelligence," "public AI," and "government AI." Accordingly, the literature review presents articles addressing AI in the public sector, particularly elaborating on those that focus on AI applications and challenges. Within this process, we could identify 14 studies that address AI applications and challenges with reference to the public sector.

To supplement this literature, we screened the references of these publications by means of a snowballing approach. In addition, we searched the Google Scholar database with the above-mentioned search terms. This procedure revealed 16 additional studies, which could not be found in the primary database search. The final relevant set contained 30 articles and studies, which then were evaluated and analyzed. Closer inspection of the AI research results reveals that the articles dealing with AI applications and AI challenges are equal in number while only few articles address both topics. Based on these research contributions, we could deduce the following categories to systematize the field of research and ensure a reasonable overview: **AI government service** (category 1), **working and social environment influenced by AI** (category 2), **public order and law related to AI** (category 3), **AI ethics** (category 4), **and AI government policy** (category 5). Articles in category one deal with public service-oriented applications and emerging challenges. This category mainly covers studies that address AI applications with a focus on improving workflows, AI forecasting models, data management, as well as decision and knowledge management (e.g. Chou, Lin, Pham, & Shao, 2015; Chun & Wai, 2007, 2008; Kouziokas, 2017; Kouziokas, Chatzigeorgiou, & Perakis, 2017; Metaxiotis, Ergazakis, Samouilidis, & Psarras, 2003; Sun & Medaglia, 2017; Zheng et al., 2018).

Most important to our study is the approach of Zheng et al. (2018) who investigate AI service provision by the government, highlighting the bilateral relationship between the needs of the public sector and the solutions provided by AI applications. In doing so, the authors show that supporting e-government tools with AI technology increases efficiency and improves government service provision.

Moreover, the articles of Chun and Wai (2007, 2008)) address service-oriented applications in public administration that focus on optimizing immigration forms with the aid of AI technologies. These AI-based services support e-government and help to reduce processing times, minimize the workload and improve the workflow, thus increasing efficiency and driving economic growth (Chun & Wai, 2007, 2008).

Category two contains studies that emphasize AI and its impact on the working and social environment. Studies in this category mainly focus on changes

through innovative AI technology in the labor market and the resulting impact on the social environment (e.g. Aguilera and Ramos Barrera 2016; Bartlett, 2017; Chen, 2009). Thereby, some articles analyze AI application in advanced and disadvantaged economies, indicating consequences caused by AI technology. While advanced economies may face high unemployment rates due to technological progress and human replacement, economically disadvantaged countries may not afford AI technology or may have poor network access and low educational background to respond to changes through AI technology (Aguilera and Ramos Barrera 2016; Chen, 2009). Moreover, Bartlett (2017) carves out several human limitations, drawing the conclusion that people should contribute their best skills and abilities to an AI-based computer system to compensate their shortcomings. In this context, AI technologies are able to enhance human capabilities, thus seeking a positive impact on the social environment.

Furthermore, articles assigned to category three address public order and law-related issues of AI technology. On the one hand, AI technology is able to support forecasting models to minimize damages and casualties from natural disasters (Tang & Wen, 2009). On the other, AI technology is able to support surveillance through government agencies and helps to protect the people (e.g. Power, 2016). Moreover, the study by Kouziokas (2017) investigates public security issues related to AI, focusing on AI for risk prevention strategies in transportation management. By using computer-based artificial neural networks, the author connects security-related issues with the quality of transportation services to identify regions with a high crime rate (Kouziokas, 2017).

As can be seen, AI in general has become a popular field of research and the impact of AI applications on daily life is rising. In this context, ethical issues of AI technology become increasingly important (Quraishi et al., 2017). For this reason, category four, AI ethics, was formed. In general, changes may lead occasionally to misunderstanding and anxiety. Especially the loss of control associated with the transition from human-administrated AI technology to independent AI technology, which may affect human beings and society, leads to insecurity (Johnson & Verdicchio, 2017). Against this background, the articles in category four address ethical issues with regard to AI applications and their impact on society, including possible benefits and prospective risks (e.g. Müller, 2014; Johnson 2015; Massaro, Norton & Kaminski, 2017; Quraishi et al., 2017). In this connection, Quraishi et al. (2017), for instance, discuss ethical decisions on whether AI technology and robotics should be promoted or rejected, placing special emphasis on the issue of how to protect humans from AI technology risks. In consideration of the ethical issues mentioned, category five contains articles that focus on AI government policies, elaborating restrictions according to laws and policies on research, as well as the development and usage of AI technologies (e.g. Boyd & Wilson, 2017; Scherer, 2016; Thierer, O'Sullivan Castillo, & Russell, 2017). Boyd and Wilson (2017), for instance, develop a practical approach for public authorities to implement AI technologies in New Zealand, demanding to create national and global norms for AI.

The literature review and the categories identified show that previous research has scarcely considered AI applications and associated challenges in the public sector. The overview of AI literature by means of categories indicates that prior studies lack an integrated approach to AI applications and their associated challenges due to their specific focus on sub-areas.

Against the background of emerging applications for the public sector that promise great public value, major challenges are arising with regard to AI responsibility, as well as social and ethical issues, which potentially threaten value creation for public service providers and agencies (Quraishi et al., 2017; Ransbotham et al., 2017). In this connection, the above-presented classification of previous studies into thematic clusters appears to be conceptually helpful because it provides a basis for deriving an integrated overview of AI applications and associated challenges with regard to the public sector, thus contributing to filling in the corresponding research gap.

However, considering the above-mentioned lack of a common AI definition, it is of greater importance to first clarify the meaning of AI, before identifying and describing AI applications and related challenges in the public sector. Therefore, the following section seeks to develop a common definition of AI, which serves as a starting point for examining AI applications and challenges, as well as for creating an unprecedented integrated overview.

AI applications in the public sector

Although AI as a term has been explored over decades, there is still no universally accepted definition available yet (Grosz et al., 2016). This leads to the fundamental problem of understanding AI in its entirety (Legg & Hutter, 2006). To obtain a basic understanding of AI, it seems useful to first define "intelligence" as an explicit term, before applying intelligence to machines and describing the compound term "artificial intelligence." Considering a variety of previous definitions, Legg and Hutter (2007); Legg & Hutter (2006) develop an

integrated definition of intelligence, describing it as the ability to interact, learn, adopt, and resort to information from experiences, as well as to deal with uncertainty. In conjunction with this,, "artificial" means a replica produced by humans (Patrick & Fattu, 1986). Based on this fundamental understanding, it seems expedient to develop a more sophisticated definition of AI in the context of public management for the further analysis. In order to create a comprehensive and clear understanding of the term, we identified six definitions of AI from relevant articles, which are presented in Table 1.

All above-mentioned definitions address the characteristics of machine-based systems and human-like intelligent behavior. AI attempts to replicate human problem-solving practices to achieve solutions that are more efficient. A special feature is the replication of human thinking and learning as well as problem-solving capabilities intended to enhance performance. Based on these definitions, it is possible to derive an integrative definition of AI that is used throughout this article. Accordingly, AI refers to the capability of a computer system to show human-like intelligent behavior characterized by certain core competencies, including perception, understanding, action, and learning. In line with this, our understanding of an AI application refers to the integration of AI technology into a computer application field with human–computer interaction and data interaction. Against the background of recent developments and further innovations, AI can be broken down into three main categories: artificial narrow intelligence (ANI), artificial general intelligence (AGI), and artificial super intelligence (ASI) (Adams et al., 2012; Nilsson, 2003; Thierer et al., 2017). ANI is usually able to solve one specific problem and needs to be programed by humans (Rosa, Feyereisl, & Team, 2016). In contrast, AGI is able to learn on its own and to transmit its experiences and skills to other tasks without any human aid (Adams et al., 2012). Thus, a comparison of human abilities with AI is inevitable. The already mentioned innovative AI application AlphaGo Zero, for instance, is in the transition stage between ANI and AGI, as it is already able to self-learn, but still cannot transfer its knowledge base to other tasks.

Table 1. AI definitions.

Author(s)	Definition
McCarthy et al. (2006, p. 12)	"The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. "
Rich et al. (2009, p. 3)	"[. . .] the study of how to make computers do things which, at the moment, people do better."
Russel and Norvig, (2010, p. 2)	AI may be organized into four categories: Systems that think like humans. Systems that act like humans. Systems that think rationally. Systems that act rationally.
Adams et al. (2012, p. 28)	"[. . .] a system that could learn, replicate, and possibly exceed human-level performance in the full breadth of cognitive and intellectual abilities."
Rosa et al. (2016, p. 6)	"[. . .] programs that are able to learn, adapt, be creative and solve problems."
Thierer et al. (2017, p. 8)	"The exhibition of intelligence by a machine. An AI system is capable of undertaking high-level operations; AI can perform near, at, or beyond the abilities of a human. This concept is further divided into weak and strong AI."

ASI in turn represents the development of software far more advanced than the human mind. However, to date such a software does not exist, though scientists believe that ASI will automatically evolve from AGI (Kreinczes, 2016). The potentially associated loss of human control, as well as the transfer of human activities and knowledge to machines gives rise to concerns that relate to government policies and ethical issues (Boyd & Wilson, 2017; Johnson & Verdicchio, 2017). Against the background of these doubts and concerns in conjunction with the prospective development from ANI to AGI, it is obvious that the government's role should be to act as a technology supervisor in general and as a technology addressee for specific AI applications.

However, before discussing these challenges associated with AI in more detail, we first consider the specific potential and usability of AI applications in the public sector. Based on the above-described literature search, we could identify a set of 16 publications that address AI applications related to the public sector. As already mentioned, AI can reduce administrative burdens and encourages resource allocation (Eggers, Fishman, & Kishnani, 2017). Despite the reduced administrative burdens, there is still little knowledge about the types and overall potential of AI applications for governments, as well as a gap between citizens' expectations and government abilities regarding AI within the scope of arising challenges (Mehr, 2017). In the following, we therefore illustrate AI applications with detailed descriptions and examples for potential use in the public sector. To provide a presentation of contemporary applications in use, we restrict our listing of AI tools to examples from the last 8 years. Against this background, we have identified 10 AI application areas in a public sector context, which are presented in Table 2.

Within the scope our analysis, we found that public organizations and government offices across the globe are already testing AI applications but are not able to keep up with the pace of private business (Mehr, 2017). Since prospective governmental use cases of AI are already common knowledge in the private sector, the respective know-how could be transferred to

Table 2. Potential AI applications for the public sector.

AI Application	AI Value Creation and Functional Proposition	Public Sector Use Cases
AI-Based Knowledge Management (KM) Software	• Generation and systematization of knowledge – gather, sort, transform, record and share knowledge • Expert systems can support the codification of the knowledge of KM • Use of neural networks enables to analyze, distribute and share knowledge with others	• Clinical documentation powered by AI (Lin et al. 2018) • …
AI Process Automation Systems	• Automation of standard tasks; perform formal logical tasks with unpredictable conditions in consistent quality • Complex human action processes (formal logical or dangerous tasks) can be transferred to automation systems, which can support humans in performing tasks • May include rule-based assessment, workflow processing, schema-based suggestions, data mining, case-based reasoning, intelligent sensor technology • Robotic process automation has emerged as a sub-area through further technology innovations. This leverages the ability of software robots or AI-driven workers to mimic human interaction with user interfaces of software systems	• Faster and higher quality request processing for immigration application forms (Chun 2007) • Automated image diagnoses (Collier et al. 2017) • Human-computer interaction for repetitive tasks like data entry etc. (Jefferies 2016) • …
Virtual Agents	• Computer-based system that interacts with the user by means of speech analytics, computer vision, written data input but may also include real-time universal translation and natural language processing systems and affective computing • Software that can perform tasks for humans • Sub-areas are chatbots and avatars	• Task allocation according to the respective area of responsibility of a specific agency (smart HR services) (Zheng et al. 2018) • Virtual nursing assistant (Collier et al. 2017) • A chatbot for helping refugees that seek asylum to fill out and search documents (Mehr 2017) • …
Predictive Analytics & Data Visualization	• These analytics are based on quantitative and statistical analysis of data. • Processing of big data for reporting, prescriptive analysis and predictive analysis • Machine learning as a technical sub-area based on algorithms that can learn from data	• Control and performance monitoring in public areas for police departments to determine terror threats and crime hotspots for preventive action (Power 2016) • Determine high crime-risk situations to secure public transport (Kouziokas 2017) • Forecast model to predict water levels (Kouziokas et al. 2017) • …
Identity Analytics	• Software combined with big data, advanced analytics and identity access management to control the access to IT systems and automate risk-based identity checks • May include deep learning and machine learning, affective computing and artificial immune systems	• Facial recognition software to verify or identify criminals in public areas (Power 2016) • AI fraud detection to secure governmental data (Hemken and Gray 2016) • …
Cognitive Robotics & Autonomous Systems	• Systems with higher-level cognitive functions that involve knowledge representation and are able to learn and respond • Sometimes in connection with affective computing to determine and adapt human behavior as well as respond to respective emotions	• Electric-powered autonomous vehicles for public transport (Christchurch International Airport Limited 2016, Jefferies 2016) • Robot-assisted surgery (Collier et al. 2017) • …
Recommendation Systems	• An information filtering system • Software-based systems that screen personalized information to predict preferences of individuals	• E-service for government offices to provide personalized information for employees (Cortés-Cediel et al. 2017) • …
Intelligent Digital Assistants (IDA)	• Software based on speech analytics • Providing an intuitive interface between a user and a system/device to search for information or complete simple tasks	• Connecting federal programs to IDA's to make public service information available for customers (Herman 2017) • IDA-Amelia to help residents locate information and complete applications forms using speech analytics and affective computing (Jefferies 2016) • …
Speech Analytics	• Software for intelligent recognition and processing of language • Understand or respond to natural language • Translate from spoken to written language or from one to another natural language • May include real-time universal translation and natural language processing systems (Pannu 2015)	• Real-time universal translation (Microsoft 2018) to translate speech and text in face-to-face communications in public service settings • Administrative workflow assistance with voice to text transcription (Collier et al. 2017) • …

(*Continued*)

Table 2. (Continued).

AI Application	AI Value Creation and Functional Proposition	Public Sector Use Cases
Cognitive Security Analytics & Threat Intelligence	• Additional application for cognitive technologies to analyze security information through natural language processing and machine learning • Interpret and organize information and provide reasoning	• Applications like Watson for cybersecurity (Dheap 2017) to support human security analysis in the public sector • …

governmental tasks. In particular, there are already AI application pilots in the public sector that focus on improving public service delivery and support for citizens (Herman, 2017a, 2017b).

In this connection, the listed AI application areas "AI Process Automation," "Predictive Analytics," "Identity Analytics," "Virtual Agents," and "Cognitive Robotics" are of high importance and hold great benefits for the public sector. These AI applications, for instance, may increase efficiency and lead to cost savings by automating processes, assisting resource allocation and reducing waiting time and administrative burdens (Eggers et al., 2017; Mehr, 2017).

Against this background, opportunities arise for public officials as they may now focus on specified tasks and activities, while machines perform routine procedures. At the same time, however, the threat of human replacement persists, thus raising ethical concerns in connection with the use of AI technology in a public context. For instance, the question of liability and responsibility in case of faulty or wrong decisions made by AI technology is not clarified yet (Boyd & Wilson, 2017). Accordingly, AI is a two-edged sword: On the one hand, governments can use the potential of AI applications to improve public affairs and to increase the efficiency of internal processes. On the other, the threats indicate that AI needs policies and regulation based on principles and core societal values in order to bring benefits to everyone (Boyd & Wilson, 2017). The following section attends to these considerations, addressing the challenges associated with AI applications.

Challenges of AI related to the public sector

As can be seen, there are various opportunities to apply and implement AI technology in the public sector with great potential to increase its efficiency. However, the related challenges should not be set aside, as they may inhibit the implementation and use of AI applications. Within the scope of our literature research, we could identify 17 studies that address AI challenges in a public sector context. This research indicates that AI has sparked a debate in science and society comprising the social and ethical impact of AI (e.g. Bartlett, 2017; Johnson & Verdicchio, 2017; Mehr, 2017; Müller, 2014; Quraishi et al., 2017), the implementation of AI technology in business conduct in the public sector (Amodei et al., 2016; Scherer, 2016; Thierer et al., 2017), as well as legal conditions including privacy regulations and responsibility (e.g. Boyd & Wilson, 2017; Power, 2016; Scherer, 2016).

Based on these streams of debate, we propose four major dimensions of AI challenges, including AI technology implementation, AI law and regulation, AI ethics, and AI society. These dimensions serve as a basis for modeling the challenges associated with the introduction of AI applications in the public sector. Although prior studies consider AI challenges individually and the issues are not completely new to science, they fail to provide an integrated overview of AI challenges. Taking a closer look, we also identified subchallenges that are assigned to the four main dimensions elaborated in the following.

AI technology implementation

In general, implementation deals with the aspects of which an initiative consists when it is provided in a particular context (Durlak & DuPre 2008). In this connection, implementing AI in the public sector requires a thoughtful and strategic course of action to take advantage of the great opportunities promised by AI and ultimately create value therefrom (Mehr, 2017).

However, although government organizations worldwide have launched initiatives for implementing and applying AI in the public sector, the implementation of AI technology poses a demanding challenge for the public sector (Cath, Wachter, Mittelstadt, Taddeo, & Floridi, 2018). According to Thierer et al. (2017), the bulk of implemented AI solutions "will likely remain weak and highly specialized" within the next decades. This apparent struggle of implementation is associated with certain issues that inhibit the progress of AI in the public sector. Based on previous research, we could identify four issues associated with the major challenge of implementing AI technology, including AI safety,

system/data quality and integration, financial feasibility, as well as specialization and expertise.

AI safety has been mentioned in previous research as an important risk factor or challenge of AI and refers to assuring the secure performance and impact of AI (Boyd & Wilson, 2017). This not merely comprises issues of information security but rather security issues in general. According to this, complex and safety-critical situations resulting from circumstances such that AI may learn negative behavior from its environment or misunderstands its surrounding are also included in this connection (Conn, 2017). In this context, Bostrom and Yudkowsky (2014) highlight the importance and necessity that AI technology be resilient against adverse manipulation by humans.

Google, one of the leading companies in AI research, has identified various security issues that have already occurred in practice. In the case of AI applications based on reinforcement learning, it has to be ensured that the AI system learns without executing catastrophic actions. In addition, it is necessary to avoid negative side effects such as disturbing the working environment, while conducting the tasks for which the entity is actually constructed (Amodei et al., 2016). For example, a robot assisting a surgery should be able to learn without endangering the patient by testing cuts or surgery methods. As a result, the implementation and advancement of AI is linked to preventing accidents and ensuring a safe functioning of AI applications to protect humankind.

System/data quality and integration is of high importance because the AI system is only as smart as the provided data from which it learns, and data is regarded as "[t]he fundamental driver of current AI systems" (Thierer et al., 2017, p. 10). In particular, data of low quality or untrusted data represent a major challenge for organizations (EY, 2018). Accordingly, the collection, aggregation, storage, and usage of unbiased and relevant data is necessary for successfully implementing AI within the public sector, as inaccurate or poor data may lead to failures (Mehr, 2017). In this connection, establishing a sophisticated and high-quality AI system that is capable of integrating data respectively and managing the interdependencies among data, technologies, and processes is essential but at the same time represents a great challenge in implementing AI solutions (Gerbert, Hecker, Steinhäuser, & Ruwolt, 2017).

Financial feasibility plays also a critical role in implementing AI technology and insufficient budget is one of the biggest challenges organizations face when initiating AI programs (EY, 2018; PwC, 2017). Accordingly, before developing and launching an AI application, the total cost associated and the expected revenues need to be considered in advance in order to assess whether an AI solution is sustainably viable. In this context, there are two main cost drivers that make financial feasibility a major challenge in the context of implementation. In particular, the investment for creating a sophisticated technological infrastructure to store and collect data is huge (Roberts, 2017). In addition, there is a high demand for a limited number of AI experts, which is associated with increasing cost of education and salaries (Bughin et al., 2017).

Specialization and expertise is another important aspect of implementing AI technology in the public sector. The rapid growth of AI is accompanied by the need for specialists and experts with relevant skills to support and promote AI development (Holdren & Smith, 2016). Accordingly, the global demand for AI experts has increased exponentially in recent years (Gagné, 2018). However, as already indicated, there is a lack of AI specialists and experts, hampering AI implementation and thus representing a great challenge in the context of AI development and implementation (EY, 2018). In this connection, government plays a crucial role and needs to place special emphasis on developing and advancing a well-educated and diverse workforce in order to build and establish a sustainable competence and knowledge base with regard to AI (Holdren & Smith, 2016).

AI law and regulations to control and governance

AI law and regulations refer to the general governance of AI and thus concern the overall ability to manage and control AI technology and its social and economic impact. Due to the broad scope of application, the governance of AI is associated with a variety of legal and regulatory issues, pertaining to data, algorithms, infrastructures, and humans (Gasser, 2017).

Governing AI in a responsible and beneficial manner thus represents a complex task and major challenge, which has so far rather been neglected, particularly when it comes to long-term issues of AI (Bostrom, Dafoe, & Flynn, 2016). Insights from leading research institutions, such as the Future Society at Harvard Kennedy School or the Future of Humanity Institute at the University of Oxford indicate the critical role and necessity of AI-related policymaking (Bostrom et al., 2016; Miailhe, 2017), highlighting the need for "some form of global governance board" (Boyd & Wilson, 2017, p. 40). Clarifying the legal status of AI is essential to remove uncertainty and determine legal responsibility and liability when AI applications cause harm (IEEE, 2017). Against this background, our literature analysis reveals three important aspects in connection

with AI law and regulations, including governance of autonomous intelligence systems, responsibility and accountability, and privacy/safety.

Governance of autonomous intelligence systems refers to the challenge of comprehending and controlling the decisions and actions of AI systems and algorithms that are often referred to as black boxes (e.g. Bleicher, 2017; Knight, 2017b). In this connection, Castelvecchi (2016, p. 21) highlights the opaqueness of AI systems and notes that "[i]nstead of storing what they have learned in a neat block of digital memory, they diffuse the information in a way that is exceedingly difficult to decipher," which in turn makes them also difficult to govern or control.

Governance is of transnational nature in this context, thus referring to both governmental action and norms or measures including other relevant stakeholders such as AI technology companies or NGOs (Bostrom et al., 2016). Accordingly, governments and all other relevant actors have to ensure certain requirements in terms of explicability, transparency, fairness, and accountability with regard to AI systems and algorithms that govern humans' lives, taking on the difficult task of introducing mechanisms of governance that minimize risks and potential pitfalls (Gasser & Almeida, 2017; Rahwan, 2018). In doing so, governments across the globe should agree on global principles and regulations for AI systems that also incorporate the prevailing standards of democracy and human rights (Boyd & Wilson, 2017). Designing and establishing a global and flexible AI governance system that not only lives up to the diverse aspects of AI but also accounts for cultural varieties and different national legal systems is complex and thus represents a major challenge in the context of AI (Gasser & Almeida, 2017).

Responsibility and accountability is closely connected to the aspect of governance and refers to defining the legal status of who is in charge and responsible for decisions made by AI. For instance, the question of who is responsible and liable arises when an autonomous vehicle for public transport harms a pedestrian. Is the hardware or the software designer, the supplier or the operator, the authorities, or even the AI application itself responsible and liable for the consequences of any decisions made by the AI application (IEEE, 2017)?

Since AI systems learn while they are in operation and act autonomously, their developers or operators may not be capable of controlling or predicting their subsequent behavior (Johnson, 2015). Thus, AI systems may defy direct human control, which leads to the so-called responsibility gap, according to which humans cannot be held accountable for the behavior of AI systems due to their lack of control and influence (Matthias, 2004).

On the opposite side, De George (2003) argues that humans are always responsible for the consequences associated with technology. Although there have been various approaches addressing the challenge of responsibility and accountability in the context of AI and seeking to overcome the responsibility gap, there is so far no consensus on how to deal with this major challenge (e.g. Anderson, 2011; Asaro, 2012; Nagenborg, Capurro, Weber, & Pingel, 2008; Santoro, Marino, & Tamburrini, 2008). Johnson (2015, p. 714) indicates that overcoming this challenge is a matter of human decision as well as political and social consensus building, highlighting "that whether or not there will ever be a responsibility gap depends on human choices not technological complexity."

Privacy/safety refers to the challenge of preserving humans' privacy and protecting data and AI-related network resources from security threats in the context of AI. This particularly means that data from individuals are collected and processed with the consent of the respective individual and in compliance with respective laws. According to Rössler (2005, p. 9), there are three ways in which privacy violations can occur: "as illicit interference in one's actions, as illicit surveillance, as illicit intrusions in rooms or dwellings," all of which that may arise in connection with AI (Calo, 2010). For instance, AI systems such as robot applications are vulnerable to cyber attacks, which are particularly dangerous because the attackers gain access to humans' physical living environment and thus their most sensitive area of life in terms of privacy (Calo, 2012). In addition, privacy concerns may also arise, for instance, in connection with AI-based government surveillance (Gasser & Almeida, 2017). Against this background, a recent survey demonstrates that the vast majority of citizens "are worried about the threat that AI poses to their privacy" (Soo, 2018, p. 1). As can be seen, AI may have a great impact on peoples' privacy, making privacy protection a major challenge in the context of AI (De Montjoye, Farzanehfar, Hendrickx, & Rocher, 2017). This challenge particularly refers to technological and legal issues. On the one hand, AI systems need to contain sophisticated cybersecurity precautions to ensure data safety and privacy (Holdren & Smith, 2016). On the other, legislation and policymaking need to be adapted to new developments and changing conditions resulting from AI (Krausová, 2017). This also includes, for instance, the handling of discrepancies among jurisdictions with regard to privacy standards etc. (Gasser, 2017).

AI ethics

An intensely discussed challenge in connection with AI applications in the public setting refers to the ethics of AI. Major ethical aspects of AI refer to the field of robot ethics or machine ethics. On the one hand, this includes considerations whether the development and use of certain AI applications and their consequences are ethical and morally justifiable, for instance, most prominently with regard to lethal autonomous weapons (Patrick., Abney, & Bekey, 2012; Russell, 2015). On the other hand, this pertains to the matter of how to embed ethical principles into AI systems to ensure that they act morally (Anderson & Anderson, 2011).

Ethics of AI do not just follow the codified law but also recognize social norms and standards referring to reasonable obligations and those including virtues of loyalty and honesty. This also requires a continuous study of moral beliefs and behavior to ensure reasonable and well-founded standards (Velasquez, Andre, Shanks, Meyer, & Meyer, 2010). Against this background, our literature analysis reveals that the challenge of AI ethics covers a broad spectrum of aspects, ranging from AI rulemaking for human behavior, to the compatibility of machine versus human value judgment, to moral dilemmas and AI discrimination.

AI rulemaking for human behavior refers to the consequences for the population resulting from AI-based decision-making and represents one of the greatest ethical challenges associated with AI. AI systems are usually designed to reproduce or imitate human behavior and make decisions for humans with the objective of maximizing effectiveness and efficiency, as well as minimizing errors in order to ultimately make the right or best decision for them. Accordingly, AI systems aim "to think and act rationally [...] [and] replicate the natural decision-making process as compared to humans" (Banerjee, Pradeep Kumar, & Bajpai, 2018).

However, as AI systems are not flawless and misjudgment from a human perspective may occur, AI-based decision-making may pose a threat to human beings (Krausová, 2017). To avoid misjudgment, AI systems are reliant upon unbiased data of high quality whose provision in its own right already represents a major challenge, as mentioned earlier (EY, 2018).

Another important challenge arises from the circumstance that humans, unlike AI systems, are not merely acting instrumentally rationally but are also characterized by other traits that influence or determine their decisions and behaviors, such as consciousness or emotion (Banerjee et al., 2018). These traits or concepts are important requirements for achieving what Solum (1992, p. 1255) calls "constitutional personhood", which indicates the equality of AI systems with humans (Krausová, 2017). Since AI systems lack these human traits, they may not be entitled to the rights associated with constitutional personhood (Solum, 1992).

Against this background, the question and challenge arises whether it is ethical and morally justifiable for AI systems to make the rules for human behavior. In addition, it is open to debate whether it is desirable to put into action AI systems that decide detached from human traits such as emotion or consciousness, especially in situations where humaneness should play an important role from a human perspective and there is legal latitude or scope of discretion. For instance, it is problematic when a decision simply changes because a machine decides and not a human being. In this context, self-learning AI systems might decide no longer on a normative ethical basis and thereby become unpredictable for human beings. For instance, the above-mentioned AI process automation system for immigration application forms might reject a query because of biased data, while a human case worker facing the same data might accept the request.

Compatibility of machine versus human value judgment is based on the assumption that human value judgment differs from prospective machine value judgment and is thus closely connected to the above-mentioned deliberations with regard to AI rulemaking for human behavior. Given that human decision-making in contrast to that of AI systems is not only influenced by rational principles, but also by factors such as emotion and consciousness (Banerjee et al., 2018), it stands to reason that their value judgments may differ in certain situations.

In this context, previous research has also drawn attention to the challenge of creating ethical consistency among AI systems and humans, which may emerge in connection with AI-based decision-making (Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016; Turilli, 2007). Turilli (2007), for instance, implies the importance of binding AI systems to the same ethical principles as individuals in order to achieve ethical consistency for an organization's overall behavior or performance.

However, despite the ambitions in the context of machine ethics to embed ethical principles into AI systems to ensure that they act morally (Anderson & Anderson, 2011), "ethical principles as used by human decision-makers may prove difficult to define and rendered computable" (Mittelstadt et al., 2016, p. 11). In this context, the self-learning algorithm and the resulting machine autonomy play a crucial role. AI systems

might develop their own system of values and create their own frame of reference, which may be incompatible with the human value system. In the worst case, the machine judgment will oppose human value judgment and might lead to independent AI technology causing harm by turning against human beings.

Moral dilemmas refer to situations in which AI systems need to choose between conflicting alternatives and "no moral choice is without undesirable moral consequences" (Ditto & Liu, 2012, p. 55). Such dilemmas become particularly relevant when it comes to AI applications, such as healthcare robots or autonomous driving, which may help or harm human beings and where the AI systems has to make an ethical decision between two negative alternatives (Deng, 2015). For instance, if an autonomous vehicle is about to be involved in an accident, its AI system may be faced with a decision that has only two bad options. Should it protect the driver of the vehicle at all cost if several other humans from another vehicle involved could otherwise be rescued? This example makes clear that some applications, particularly those involving life-or-death decisions need to command an ethical reasoning skill before being applied fully autonomously in the field (Conitzer, Sinnott-Armstrong, Borg, Deng, & Kramer, 2017).

As mentioned earlier, the field of machine ethics deals with the challenge of integrating ethical principles into AI systems and teaching them ethical behavior (Anderson & Anderson, 2011). However, the great challenge remains "to figure out what's relevant for artificial intelligence to reason successfully in ethical situations" (Deng, 2015, p. 25).

AI discrimination relates to the challenge of preventing inequality and unfairness caused by AI applications. Previous research indicates that AI systems and their profiling may lead to discrimination and thus violate ethical principles such as equality and fairness (Thierer et al., 2017). Since humans program the algorithms of AI systems or serve as their source of data input, AI systems also may pick up their values and biases (Citron & Pasquale, 2014). For instance, AI systems may absorb gender or racial prejudices hidden in human language patterns and amplify human stereotypes (Bass & Huet, 2017; Devlin, 2017). In response to these drawbacks and potentially even illegal discrimination, respective approaches are increasingly emerging in the context of data mining and machine learning in order to detect and prevent discrimination (e.g. Barocas, 2014; Bolukbasi, Chang, Zou, Saligrama, & Kalai, 2016; Hajian & Domingo-Ferrer, 2013). However, despite these first attempts at counteracting discrimination in connection with AI, the challenge of eliminating discrimination from AI systems persists and thus the necessity as well as demand for further detection and prevention mechanisms remains (Mittelstadt et al., 2016).

AI society

AI society refers to the transition of social life and human interaction driven by AI and the social challenges associated with these changes. AI has reached the everyday world, influencing our daily routines and many core areas of society, such as healthcare, transportation, and finance (Cath et al., 2018). At the same time, there are also increasing concerns about AI's future development and diffuse anxieties emerging in society due to its potentially adverse effects on mankind and society (Johnson & Verdicchio, 2017).

For instance, prominent AI researchers have launched an initiative in the form of open letters published by the Future of Life Institute, which have received broad support by famous AI opinion leaders and experts, like Stephan Hawking or Elon Musk. In these initiatives, they address the threat AI may pose, expressing their concerns about AI and the harm it may do to humanity and society, as well as advocating a development of AI that is beneficial to the latter (Future of Life Institute, 2015a; 2015b).

Fears and objections toward the application of AI are also present among citizens. The European Commission's survey on autonomous systems in 2015, for example, shows that a large proportion of citizens have reservations with regard to AI in various areas of application. For instance, the majority of respondents feel uncomfortable with AI systems providing services to elderly people or using AI systems for medical surgery (European Commission, 2015).

Against this background, our literature analysis identified three important aspects in the literature referring to the challenge of AI to society. These include workforce substitution and transformation, social acceptance/trust in AI, and transformation of human-to-machine (H2M) and machine-to-machine (M2M) interaction.

Workforce substitution and transformation refers to the impact of AI on the labor market and represents one of the most pervasive challenges to society in the context of AI (Bataller & Harris, 2016; Boyd & Wilson, 2017). There is increasing concern that further advancement and implementation of AI will lead to unemployment as work activities and jobs become increasingly subject to automation (Mehr, 2017; Thierer et al., 2017). According to a recent study of PwC, more than one-third of workers were concerned

about a potential job loss as a consequence of automation (Brown et al., 2018).

Various study findings support these misgivings, although the expert estimations on this issue vary widely. While a study of PwC figures that about one-third of jobs are endangered in major industrial nations (Hawksworth, Kupelian, Berriman, & Mckellar, 2017), studies of the Oxford University (Frey & Osborne, 2017) and the McKinsey Global Institute (Manyika et al., 2017) reach quite similar conclusions, stating that about half of the occupations or workforce activities could be automated and are thus at risk. Other estimates draw an even more pessimistic picture of 80–90% automation-induced job elimination within the next 15 years (Lever, 2017).

Although the approaches leading to these figures have been put into perspective and have received criticism (Atkinson, 2017), it becomes apparent that AI greatly fosters the substitution of human workforce and society is facing technological unemployment on a large scale. Besides the specific challenge of workforce substitution, the complementation of many jobs by AI will change their requirement profiles to the effect that workers may experience a shift in roles toward a more supervisory function and will require AI-specific skills (Bataller & Harris, 2016; Mehr, 2017). The development of AI also leads to the emergence of completely new job profiles, such as data scientists or machine learning engineers, requiring experts with AI-specific skills, which are currently few and far between on the market (EY, 2018). The private sector and the public sector must prepare for and address these major challenges in order to protect the economic systems and society from adverse effects that may arise from this transition.

Social acceptance/trust in AI represents another major social challenge in the context of AI to be met that is crucial for AI to be successful. While, the social debate about AI and impacts on society has become more intense and has rather moved toward a positive end in recent years, certain concerns such as losing control over AI, ethical worries, and the negative impact on workforce have increased (Fast & Horvitz, 2017). Findings of a recent study support this development, suggesting that citizens are quite favorable toward AI, but only as long as it does not directly affect them or their health (Nitto, Taniyama, & Inagaki, 2017).

Social acceptance and trust in AI is closely connected with and dependent on the resolution of other challenges such as AI safety, privacy, discrimination, and workforce substitution that may negatively affect citizens in a direct way. A mismatch between citizen or user expectations and the reality of AI systems may negatively influence their acceptance (Hameed, Ten, Thomsen, & Xiaodong, 2016).

If citizens feel discriminated by AI or that it threatens their safety, privacy, or employment, thus contradicting their expectations with regard to AI, they will very unlikely accept or trust it. Moreover, the combination of unemployment and increases in profit through AI may reinforce social inequality and widen the gap between the rich and the poor, which in turn leads to social dissatisfaction instead of social acceptance (Boyd & Wilson, 2017; Wisskirchen et al., 2017).

Against this background, the challenge arises to reduce social concerns, to prepare the next generation for a digital age by initiating and promoting a social dialog and to create confidence in AI systems. In this connection, Hameed et al. (2016) highlight the importance of understanding and taking into consideration the factors that drive acceptance of AI systems in society. However, it should generally be kept in mind that the acceptance of technology per se is dependent on context (Petit, 2018).

Transformation of H2M/M2M interaction is connected to the understanding of communication between humans and machines or among the latter in the context of AI and refers to the challenges associated with these interactions. Due to the advancements in speech, gesture and pattern recognition, AI can substantially change our interactions (Cath et al., 2018) and future H2M interaction is said to be characterized by systems that can anticipate movements of humans and can be steered by voice, gesture and environment communication. (Ice, 2015). However, since most AI systems are designed and developed by technical experts who are not familiar with cognitive aspects, the H2M interaction is not ideal and users need "to adapt to these machines, often resulting in miscommunication and undue anxiety" (Lee & Sathikh, 2013).

A great challenge in this connection refers to making H2M and M2M interaction more intuitive, which requires stronger consideration of cognitive components (Ducatel, Bogdanowicz, Scapolo, Leitjen, & Burgelman, 2005). In this connection, the user also needs to understand the way in which AI is actually influencing the interaction (Holmquist, 2017).

Furthermore, designing an AI system that harmonizes with the user and realizing the great benefits that AI offers, demands many more ways of communication with the AI system (Holmquist, 2017; Mankins, 2015). Besides verbal communication, nonverbal communication is an important attribute of human behavior and interaction. Nonverbal communication represents communication aspects "that are not part of the words themselves, including facial expressions, body posture and gestures" (Dodds,

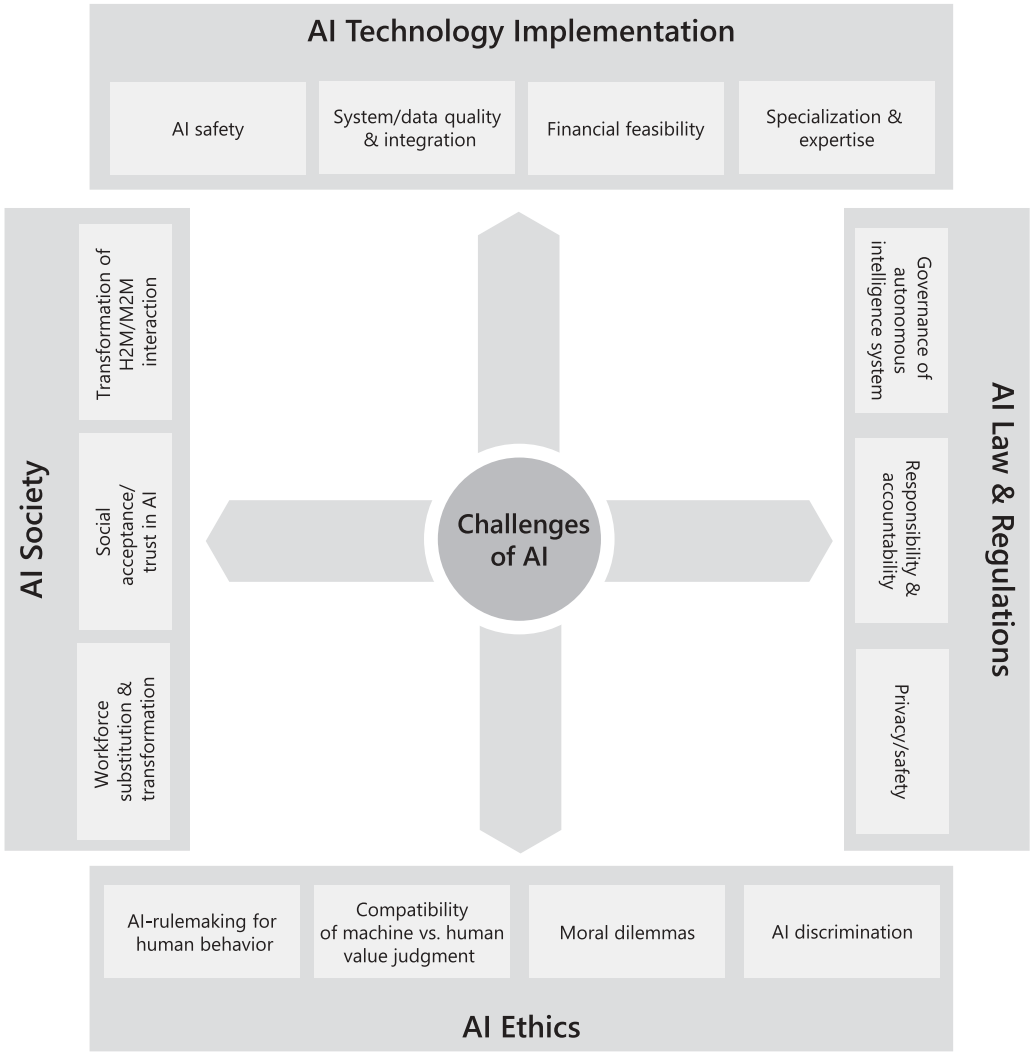

Figure 1. Four-AI-challenges model.

Mohler, & Bülthoff, 2011, p. 1). The ability to understand emotional expressions may be crucial and poses a severe challenge to rational AI systems (Banerjee et al., 2018), resulting in a different H2M interaction behavior.

This circumstance may also contribute to the homogeneity problem that H2M interaction currently faces, according to which AI systems such as robots are only able to interact with humans in a uniform way and cannot approach humans in a variable way in every detail to accomplish a plethora of interactions (Tanaka & Kobayashi, 2015). In order to supersede human-to-human interaction, AI will need a genuine understanding of how the world works and the intentions of the interaction partner (Holmquist, 2017).

Furthermore, another great challenge pertains to the rapidly developing and increasing M2M interactions that require appropriate human control (Minchev, 2016). Given that self-algorithms become more and more sophisticated and humans do not play an active part in these interactions but rather assume a moderating role, there is a risk of losing access to these interactions, which thus may defy human control and understanding. As with AI governance in general, the challenge will be to ensure that M2M interactions in the context of AI are transparent and accountable to, as well as controllable by humans.

Based on the above-mentioned deliberations, Figure 1 depicts the Four-AI-Challenges Model illustrating the major dimensions and sub-aspects of AI challenges in the public sector. In the following, we discuss the findings with regard to the AI applications as well as their potential for the public sector and their connection to all dimensions of challenges.

Discussion and conclusion

The starting point of this study refers to the increasing relevance of AI in science and practice as well as its

groundbreaking potential for the public sector on a global level—both in positive and negative terms. A number of countries in particular the United States and China have recognized the great value of AI for public usage and has launched various cost-intensive AI initiatives, revealing the broad range of potential application areas (Holdren & Smith, 2016; Knight, 2017a). However, for some reason or another no government has so far comprehensively addressed the whole AI application spectrum. Associated challenges that may hamper successful implementation of AI in the public sector are occasionally neglected in government reports (Boyd & Wilson, 2017). It seems that despite the great benefits and recent efforts regarding AI, its implementation in the public sector is struggling.

Since the public sector represents a relatively new area of AI application and many applications have been employed as innovative pilot projects (Herman, 2017a; Singh, 2017), governments, and public authorities may not be aware of the full range of AI applications opportunities nor of related challenges. This is aggravated by the fact that respective public sector-specific AI research is still in its infancy and so far fails to provide an integrated view of AI applications and challenges for the public sector.

In response to this practical shortcoming and the corresponding gap in research, this study seeks to develop a comprehensive understanding of AI, examining its applications and challenges in the public sector. In doing so, this study contributes to AI research by (1) enhancing the understanding of AI in a public context, (2) identifying AI applications and explaining their value creation and functional proposition, (3) providing public sector-specific AI use cases, (4) presenting a Four-AI-Challenges Model that incorporates major dimensions and sub-aspects of AI challenges, and (5) deriving both concrete implications for AI-related research and specific guidelines for public managers dealing with AI implementation.

The extensive analysis of applications and challenges-related AI literature in this study initially revealed a lack of a unified integrated definition of the AI concept (Scherer, 2016), which is crucial to ensure an efficient and consistent development of AI research. In this connection, the integrated AI definition proposed in this study contributes to the development and establishment of a common conceptual understanding of AI in the public sector, which is an essential starting point and prerequisite for conceiving AI applications and challenges. Moreover, the categorization of the pertinent literature within our literature review contributes to research by structuring the respective field of AI research and providing a conceptually helpful starting point for future research endeavors. The categories identified thereby give already first indications for the potential application areas and challenges of AI, of which many important aspects have been admittedly addressed by previous research, but in an isolated and fragmentary way (e.g. Boyd & Wilson, 2017; Mehr, 2017; Thierer et al., 2017). Accordingly, prior approaches have remained short of providing a systematic and integrated description of AI applications and challenges in the public sector. The study at hand makes this contribution, which is particularly important for a comprehensive understanding thereof.

The holistic consideration of applications and challenges also indicates their interdependencies. AI applications and challenges are closely connected to each other, as these challenges may arise when introducing AI applications to the public sector and should therefore be considered in concert. There are also interdependencies among the challenges, allowing to deduce potential synergy effects when it comes to opposing and overcoming these challenges. For instance, the challenge of achieving social acceptance and trust in AI is dependent on whether other challenges that may negatively affect citizens, such as AI safety, privacy, discrimination, and workforce substitution, are successfully addressed. Moreover, making regulations to control and govern AI applications requires addressing and discussing ethical issues or implementation-related issues in advance. Overall, these interdependencies are a further argument in favor of an integrative approach.

Besides the study's thorough scope, another important contribution refers to the level of detail and the resulting specificity and reference points for implementing and regulating AI in the public sector. The description of AI applications exceeds further approaches (e.g. Mehr, 2017; Thierer et al., 2017) by specifying their value creation and functional proposition as well as highlighting concrete field-tested use cases for the public sector, demonstrating the broad and multifaceted spectrum of AI applications as well as underscoring the importance of an integrated overview to reveal their entire potential for the public sector. The identification of various use cases from previous research and governmental programs and the thereupon based formation of 10 superordinate categories of AI applications is to the best of our knowledge the first comprehensive overview of AI applications for the public sector and thus constitutes a central point of reference in view of the formerly fragmented state of knowledge. Likewise, the elaborate deliberations about the challenges of AI go beyond the mere consideration of challenges by former approaches that fall short of providing detailed descriptions of respective issues or

leverage points specific enough to serve as a practical orientation or decision aid in the context of AI implementation and regulation (e.g. Bataller & Harris, 2016).

Prior approaches in the context of AI ethics, for instance, strongly focus on the ethical challenges of AI, neglecting the aspect of how to solve or regulate these ethical issues (e.g. Gasser & Almeida, 2017). Here, there a few exceptions that may act as role model for AI research, such as a recent publication of the IEEE Global Initiative on Ethics of Autonomous and Intelligent that provide concrete highly demanded recommendations on how to deal with the ethical challenges of AI (IEEE, 2017).

Against this background, our integrative approach identified specific AI challenges to the public sector in the literature and formed four main dimensions to which these challenges were assigned, developing the Four-AI-Challenges Model for the public sector. The relevance and severity of these challenges may be dependent on context. For instance, social acceptance may vary across countries (Nitto et al., 2017) and the challenge of workforce substitution and transition may have a greater impact on society in terms of unemployment and inequality in countries with larger automation potential than others (Arntz, Gregory, & Zierahn, 2016; Chui, Manyika, & Miremadi, 2017).

Speaking of challenges, prior AI literature has also largely focused on its counterpart, the benefits of AI, neglecting the challenges associated with AI. Even government reports are regarded to have an imbalanced view in this respect (Boyd & Wilson, 2017). Balanced presentations of benefits and risks like in the recently published government report of the United States rather constitute an exception, but may serve as a useful benchmark for AI research in terms of its double-sided perspective (Holdren & Smith, 2016). However, the recently kicked off public debate whether AI is good or bad for society in which famous AI experts and opinion leaders, such as Elon Musk or Stephen Hawking, have voiced fears about the threats of AI (Newman, 2017), has raised increasing awareness of the challenges of AI and may result in stronger consideration thereof in future AI literature. Against this background, the study at hand also represents an initial contribution towards redressing the respective balance in research between benefits and challenges of AI.

Furthermore, the challenges of AI play particularly in the public sector a special role, as the protection and education of citizens and their provision with goods and services they cannot provide on their own is a central part of governmental duties (Slaughter, 2017). Due to the rapid changes in AI technology, it is difficult for any government to keep up with the pace of development and create comprehensive regulations, while at the same time meeting the variety of challenges, satisfying ethical concerns, building a broad understanding and acceptance of AI in society, as well as creating a sense of security and so on.

Against this background, the study findings offer several practical implications and carry valuable advice for public management. First, the study imparts a thorough understanding of AI and provides an AI applications checklist for public managers including their value creation and functional proposition, which may thus serve as a benchmark for prospective AI ventures in the public sector. In addition, the Four-AI-Challenges Model may serve as a reference point to avoid pitfalls when introducing AI applications to the public sector. When considering certain AI applications for implementation, public managers should be aware that it is critical to simultaneously take into account the variety of challenges in terms of risks and adverse effects that may endanger its success. Acknowledging their interdependencies may also help public managers to identify and realize synergy effects when coping with these challenges.

Moreover, public managers should bring to mind that a number of challenges is in their direct sphere of influence and can be quite simply addressed. For instance, in response to the lack of specialization and expertise, public organizations could initiate educational programs such as training courses or workshops for their employees to develop and strengthen AI-specific skills, which also may smoothen the changes and mute potentially negative impacts in the context of workforce substitution and transition. Moreover, such educational measures for employees may also be promising in connection with issues of IT or data security and privacy in the context of AI, to ensure a responsible handling of these critical areas that greatly affect and are highly relevant to citizens.

Furthermore, sensitive aspects that raise concerns among citizens, such as workforce substitution or AI safety and privacy, could generally be addressed by measures that create transparency with regard to potential consequences of AI-based changes in the working, social, and personal environment, for instance, in the form of respective information events. This may contribute to allaying potential fears among employees or citizens and to achieving social acceptance and trust in AI. To this end, the implementation of AI applications should take place gradually in the form of smaller, ethically inoffensive pilot projects that are well manageable and readily comprehensible for all actors involved, promising fast success and encouraging social acceptance of AI. On grounds of effectiveness and efficiency,

it makes sense for public managers to tackle these low hanging fruits first. By contrast, governance-related and ethical challenges may be more difficult to address and rather represent a long-term issue.

Nonetheless, it is essential that public organizations establish an AI governance with respective capabilities, structures, and processes that regulate the issues of responsibility and accountability, privacy and safety and so on. Likewise, ethical measures are important requirements and duties public organizations need to fulfill, which public organizations may institutionalize by defining ethical guidelines and monitoring their adherence, as well as establishing an ethics council that deals with ethical matters of AI. Overall, the successful future of AI requires public organizations to rethink existing approaches and structures and adapt them in accordance with the prevailing challenges.

Beyond the above-mentioned implications for research and practice, this study is also subject to some limitations, which may serve as useful starting points for future research. While this study has a clear focus on applications and challenges of AI, the aspects of benefits and regulation are rather set aside and superficially addressed. As all challenges are closely connected to benefits and regulation (Scherer, 2016), future research could broaden the scope of study, examining and comparing all aspects in a common framework to better understand their relationships.

Moreover, investigating the interdependencies among AI challenges in more detail would have gone beyond the scope of this study, but may be of interest for further studies to improve our understanding of the dynamics of these challenges and to provide detailed insights with regard to potential synergy effects when coping with them.

In addition, the conceptual deductive research approach of this study does not allow to make empirical statements or to draw causal conclusions regarding AI applications and challenges. Accordingly, further empirical examinations are necessary, for instance, in terms of surveys with public managers, to prove whether our study findings such as the Four-AI-Challenges Model are empirically tenable.

Finally, given the indicated context dependency of challenges, future research could check them for cross-national differences. Since research on AI and the public sector is still in an early stage, there are various research opportunities available that scholars need to address, in order to extend our inchoate theoretical and empirical knowledge in this field. This study seeks to make an initial conceptual contribution, by providing an integrative analysis of AI applications and challenges in the public sector.

References

Adams, S., Arel, I., Bach, J., Coop, R., Furlan, R., Goertzel, B., Hall, J. S., Samsonovich, A., Scheutz, M., Schlesinger, M., Shapiro, S. C., Sowa, J. (2012). Mapping the landscape of human-level artificial general intelligence. *AI Magazine*, *33* (1), 25–42. doi:10.1609/aimag.v33i1.2322

Aguilera, A., Barrera, R., & Gabriela, M. (2016). Technological unemployment: An approximation to the Latin American case. *AD-minister*, (29), 59–78. doi:10.17230/ad-minister.29.3

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). *Concrete problems in AI safety*. In *CoRR* abs/1606.06565, 1–29. Retrieved June 29, 2018, from https://arxiv.org/pdf/1606.06565.pdf

Anderson, M., & Anderson, S. L. (Eds.). (2011). *Machine ethics*. New York, NY: Cambridge University Press.

Anderson, S. L. (2011). Machine metaethics. In M. Anderson & S. L. Anderson (Eds.), *Machine ethics* (pp. 21–27). New York, NY: Cambridge University Press.

Ansip, A. (2017). *Making the most of robotics and artificial intelligence in Europe*. Edited by European Commission. Retrieved June 29, 2018, from https://ec.europa.eu/commission/commissioners/2014-2019/ansip/blog/making-most-robotics-and-artificial-intelligence-europe_en

Arntz, M., Gregory, T., & Zierahn, U. (2016). The risk of automation for jobs in OECD countries: A comparative analysis. *OECD Social, Employment and Migration Working Papers*, 189 1–34, checked on June 27, 2018

Asaro, P. M. (2012). A body to kick, but still no soul to damn: Legal perspectives on robotics. In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot ethics: The ethical and social implications of robotics* (pp. 169–186). Cambridge, MA: MIT Press.

Atkinson, R. D. (2017). In defense of robots. *National Review*, *LXIX*(7). Retrieved June 20, 2018, from, https://www.nationalreview.com/2017/04/robots-jobs-industrial-future/

Banerjee, S., Pradeep Kumar, S., & Bajpai, J. (2018). A comparative study on decision-making capability between human and artificial intelligence. In B. K. Panigrahi, M. N. Hoda, V. Sharma, & S. Goel (Eds.), *Nature inspired computing: Proceedings of CSI 2015* (pp. 203–210.). Singapore: Springer (Advances in Intelligent Systems and Computing, 652).

Barocas, S. (2014). *Data mining and the discourse on discrimination*. Proceedings of the Data Ethics Workshop, Conference on Knowledge Discovery and Data Mining (KDD) 2014, New York, NY, USA.

Bartlett, S. J. (2017). The case for government by artificial intelligence. *SSRN Journal*, 1–15. doi:10.2139/ssrn.3089920

Bass, D., & Huet, E. (2017). *Researchers combat gender and racial bias in artificial intelligence*. Edited by Bloomberg. Retrieved June 29, 2018, from https://www.bloomberg.com/news/articles/2017-12-04/researchers-combat-gender-and-racial-bias-in-artificial-intelligence

Bataller, C., & Harris, J. (2016). *Turning artificial intelligence into business value*. Today. Edited by Accenture. Retrieved June 29, 2018, from https://www.accenture.com/t20160401T100530__w__/us-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Technology_11/Accenture-Turning-Artificial-Intelligence-into-Business-Value.pdf

Bleicher, A. (2017). *Demystifying the black box that is AI: Humans are increasingly entrusting our security, health and safety to "Black Box" intelligent machines*. Scientific American. Retrieved June 11, 2018, from https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems 29. Neural information processing systems 2016* (pp. 4349–4357). Barcelona, Spain, 05. 12.-10.12.2016. Red Hook, NY: Curran Associates. Retrieved from http://papers.nips.cc/paper/6228-man-is-to-computer-programmer-as-woman-is-to-homemaker-debiasing-word-embeddings.pdf

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence*. Cambridge, UK: Cambridge University Press.

Bostrom, N., Dafoe, A., & Flynn, C. (2016). *Policy desiderata in the development of machine superintelligence*. Future of Humanity Institute, University of Oxford. Retrieved June 8, 2018, from https://www.fhi.ox.ac.uk/wp-content/uploads/Policy-Desiderata-in-the-Development-of-Machine-Superintelligence.pdf

Boyd, M., & Wilson, N. (2017). Rapid developments in artificial intelligence: How might the New Zealand government respond? *Policy Quarterly*, *13*(4), 36–44.

Brown, J., Gosling, T., Sethi, B., Sheppard, B., Stubbings, C., Sviokla, J., Jon, W., Daria, Z., Liz, F. (2018). *Workforce of the future: The competing forces shaping 2030*. Edited by PwC. Retrieved June 20, 2018, from https://www.pwc.com/gx/en/services/people-organisation/workforce-of-the-future/workforce-of-the-future-the-competing-forces-shaping-2030-pwc.pdf

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Nicolaus., H., Monica, T. (2017). *Artificial intelligence: The next digital frontier?* Edited by McKinsey Global Institute. Retrieved June 8, 2018, from https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx

Byford, S. (2016). *Google's AlphaGo AI Beats Lee Se-dol again to Win Go Series 4-1*. Edited by The Verge. Retrieved June 8, 2018, from https://www.theverge.com/2016/3/15/11213518/alphago-deepmind-go-match-5-result

Calo, R. (2010). Peeping HALs: Making sense of artificial intelligence and privacy. *European Journal of Legal Studies*, *2*(3), 168–192.

Calo, R. M. (2012). Robots and privacy. In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot ethics: The ethical and social implications of robotics* (pp. 187–201). Cambridge, MA: MIT Press.

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, *538*(7623), 20–23. doi:10.1038/538020a

Castro, D., & New, J. (2016). *The promise of artificial intelligence*. Edited by Center for Data Innovation. Retrieved June 8, 2018, from http://www2.datainnovation.org/2016-promise-of-ai.pdf

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': The US, EU, and UK approach. *Science and Engineering Ethics*, *24*(2), 505–528. doi:10.1007/s11948-017-9901-7

Chen, H. (2009). AI, E-government, and politics 2.0. *IEEE Intelligent Systems*, *24*(5), 64–86. doi:10.1109/MIS.2009.91

Chou, J.-S., Lin, C.-W., Pham, A.-D., & Shao, J.-Y. (2015). Optimized artificial intelligence models for predicting project award price. *Automation in Construction*, *54*(June), 106–115. doi:10.1016/j.autcon.2015.02.006

Christchurch International Airport (2016). New Zealands's first trial of autonomus vehicle announced. Retrieved July 2, 2018, from http://www.christchurchairport.co.nz/media/827295/new_zealand_s_first_trial_of_autonomous_vehicle_announced.pdf

Chui, M., Manyika, J., & Miremadi, M. (2017). *The countries most (and least) likely to be affected by automation*. Harvard Business Review. Boston, MA. Retrieved June 27, 2018.

Chun, A., & Wai, H. (2007). *Using AI for E-government automatic assessment of immigration application forms*. In Proceedings of the National Conference on Artificial Intelligence, AAAI 2007, Vancouver, BC, Canada. 2, pp. 1684–1691.

Chun, A., & Wai, H. (2008). An AI framework for the automatic assessment of E-government forms. *AI Magazin*, *29*(1), 52–64.

Citron, D. K., & Pasquale, F. A. (2014). The scored society: Due process for automated predictions. *Washington Law Review*, *89*(1), 1–33.

Collier, Matt, Fu, Richard & Yin, Lucy. (2017). Artifical intelligence: Healthcare's new nervous system. Edited by Accenture. Retrieved June 29, 2018, from https://www.accenture.com/t20170418T023052Z__w__/au-en/_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf

Conitzer, V., Sinnott-Armstrong, W., Borg, J. S., Deng, Y., & Kramer, M. (2017). *Moral decision making frameworks for artificial intelligence*. Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI 2017, San Francisco, CA, USA, pp. 4831–4835.

Conn, A. (2017). *Artificial intelligence: The challenge to keep it safe*. San Francisco, CA: Future of Life Institute. Retrieved June 6, 2018, from, https://futureoflife.org/2017/09/21/safety-principle/

Cortés-Cediel, María E.; Cantador, Iván & Gil, Olga (2017). Recommender systems for e-governance in smart cities. In Yang, Jie; Sun, Zhu; Bozzon, Alessandro; Zhang, Jie & Larson, Martha (Eds.), Proceedings of the international workshop on citizens for recommender systems - CitRec '17, Como, Italy, August 27, 2017, New York, NY: ACM press.

De George, R. T. (2003). *The ethics of information technology and business*. Malden, MA: Blackwell Publishing (Foundations of Business Ethics, 3).

De Montjoye, Y.-A., Farzanehfar, A., Hendrickx, J., & Rocher, L. (2017). Solving artificial intelligence's privacy problem. *Field Actions Science Reports. The Journal of Field Actions Special Issue*, *17*, 80–83.

Deng, B. (2015). Machine ethics: The robot's dilemma. *Nature*, *523*(7558), 24–26. doi:10.1038/523024a

Devlin, H. (2017). *AI programs exhibit racial and gender biases, research reveals*. Edited by The Guardian. Retrieved July 6, 2018, from https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals

Dheap, Vijay (2017). IBM QRadar advisor with Watson: Revolutionizing the way security analysts work. Edited by IBM. Retrieved July 2, 2018, from https://securityintelligence.com/ibm-qradar-advisor-with-watson-revolutionizing-the-way-security-analysts-work/

Ditto, P. H., & Liu, B. (2012). Deontological dissonance and the consequentialist crutch. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 51–70). Washington, DC: American Psychological Association.

Dodds, T. J., Mohler, B. J., & Bülthoff, H. H. (2011). Talk to the virtual hands: Self-animated avatars improve communication in head-mounted display virtual environments. *PLOS ONE*, *6* (10), e25759. doi:10.1371/journal.pone.0025759

Ducatel, K., Bogdanowicz, M., Scapolo, F., Leitjen, J., & Burgelman, J.-C. (2005). That's what friends are for - Ambient intelligence (AmI) and the information society in 2010. In A. Zerdick, A. Picot, K. Schrape, J.-C. Burgelman, R. Silverstone, V. Feldmann, C. Wernick & C. Wolff (Eds.), *E-merging media: Communication and the media economy of the future* (pp. 181–200). Berlin, Germany: Springer.

Durlak, J. A, & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3–4), 327–350.

Eggers, W. D., Fishman, T., & Kishnani, P. (2017). *AI-augmented human services: Using cognitive technologies to transform program delivery*. Edited by Deloitte, Christian Wernick, Carolin Wolff. Retrieved July 2, 2018, from https://www2.deloitte.com/content/dam/insights/us/articles/4152_AI-human-services/4152_AI-human-services.pdf

European Commission. (2015). *Special Eurobarometer 427: Autonomous systems*. Report. Retrieved July 2, 2018, from http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_427_en.pdf

EY. (2018). *The growing impact of AI on business*. Edited by MIT Technology Review. Cambridge, MA: Massachusetts Institute of Technology. Retrieved June 7, 2018, from https://www.technologyreview.com/s/611013/the-growing-impact-of-ai-on-business/

Fast, E., & Horvitz, E. (2017). *Long-term trends in the public perception of artificial intelligence*. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), 963–969.

Frey, C. B., & Osborne, M. A. (2017). The future of employment. *How Susceptible are Jobs to Computerisation? in Technological Forecasting and Social Change*, *114*, 254–280. doi:10.1016/j.techfore.2016.08.019

Future of Life Institute. (2015a). *Autonomous weapons: An open letter from AI & robotics researchers*. Retrieved June 20, 2018, from https://futureoflife.org/open-letter-autonomous-weapons/

Future of Life Institute. (2015b). *Research priorities for robust and beneficial artificial intelligence*. Retrieved June 20, 2018, from https://futureoflife.org/ai-open-letter

Gagné, J.-F. (2018). *The global artificial intelligence talent report: 2018*. Edited by DX Journal. Retrieved June 7, 2018, from http://www.digitaljournal.com/digitaltransformation/2018/03/global-artificial-intelligence-talent-report-2018/

Gasser, U. (2017. *AI and the law: Setting the stage*. Berkman Klein Center for Internet & Society at Harvard University. Retrieved June 14, 2018, from https://medium.com/berkman-klein-center/ai-and-the-law-setting-the-stage-48516fda1b11

Gasser, U., & Almeida, V. A. F. (2017). A layered model for AI governance. *IEEE Internet Computing*, *21*(6), 58–62. doi:10.1109/MIC.2017.4180835

Gaurav, B., Queirolo, A., & Santhanam, N. (2018). *Artificial intelligence: The time to act is now*. Edited by McKinsey & Company. Retrieved July 2, 2018, from https://www.mckinsey.com/industries/advanced-electronics/our-insights/artificial-intelligence-the-time-to-act-is-now

Gerbert, P., Hecker, M., Steinhäuser, S., & Ruwolt, P. (2017). *Putting artificial intelligence to work*. Edited by BCG Henderson Institute. Munich, Germany: The Boston Consulting Group. Retrieved January 22, 2018, from https://www.bcg.com/de-de/publications/2017/technology-digital-strategy-putting-artificial-intelligence-work.aspx

Grosz, B. J., Mackworth, A., Altman, R., Horvitz, E., Mitchell, T., Mulligan, D., & Shoham, Y. (2016). *Artificial intelligence and life in 2030: One hundred years study on artificial intelligence*. Edited by Stanford University. Retrieved July 2, 2018, from https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf

Hajian, S., & Domingo-Ferrer, J. (2013). A methodology for direct and indirect discrimination prevention in data mining. *IEEE Transactions on Knowledge and Data Engineering*, *25*(7), 1445–1459. doi:10.1109/TKDE.2012.72

Hameed, I. A., Ten, Z.-H., Thomsen, N. B., & Xiaodong, D. (2016). *User acceptance of social robots*. The Ninth International Conference on Advances in Computer-Human Interactions, 274–279.

Hawksworth, J., Kupelian, B., Berriman, R., & Mckellar, D. (2017). *UK economic outlook*. Edited by PwC. Retrieved June 20, 2018, from https://www.pwc.co.uk/economic-services/ukeo/pwc-uk-economic-outlook-full-report-march-2017-v2.pdf

Hemken, Terry & Gray, Chris (2016). Smart move: Technologies make their mark on public service. Edited by Accenture. Retrieved June 29, 2018, from https://www.accenture.com/t20170224T040950Z__w__/us-en/_acnmedia/PDF-36/Accenture-Public-Service-Emerging-Technologies-Slideshare.pdf

Herman, J. (2017a). *Federal pilot to integrate public services into intelligent personal assistants*. Edited by DigitalGov. U. S. General Services Administration. Retrieved June 5, 2018, from https://digital.gov/2017/04/24/federal-pilot-to-integrate-public-services-into-intelligent-personal-assistants/

Herman, J. (2017b). *Opening public services to artificial intelligence assistants*. Edited by U.S. General Services Administration. Retrieved June 6, 2018, from https://www.gsa.gov/blog/2017/06/06/Opening-Public-Services-to-Artificial-Intelligence-Assistants

Holdren, J., & Smith, M. (2016). *Preparing for the future of artificial intelligence*. Edited by Executive Office of the

President National Science and Technology Council Committee on Technology. Washington, DC. Retrieved July 2, 2018, from https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

Holmquist, L. E. (2017). Intelligence on tap: Artificial intelligence as a new design material. *Interactions*, *24*(4), 28–33. doi:10.1145/3085571

Ice, B. (2015). *Gesture technology, invisible user interface and the evolution of human-to-machine interaction*. Edited by Marketing Magazine. Retrieved June 22, 2018, from https://www.marketingmag.com.au/hubs-c/gesture-technology-invisible-user-interface-evolution-human-machine-interaction/

IEEE. (2017). *Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems*. Edited by The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Retrieved June 14, 2018, from http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf

Jefferies, Duncan. (2016). The automated city: Do we still need humans to run public services? Edited by The Guardian. Retrieved July 2, 2018 from https://www.theguardian.com/cities/2016/sep/20/automated-city-robots-run-public-services-councils

Johnson, D. G. (2015). Technology with no human responsibility? *Journal of Business Ethics*, *127*(4), 707–715. doi:10.1007/s10551-014-2180-1

Johnson, D. G., & Verdicchio, M. (2017). AI anxiety. *Journal of the Association for Information Science and Technology*, *68*(9), 2267–2270. doi:10.1002/asi.23867

Knight, W. (2017a). China's AI awakening: The west shouldn't fear China's artificial-intelligence revolution. It should copy it. *MIT Technology Review*, *120*(6), 66–72.

Knight, W. (2017b). The dark secret at the heart of AI. *MIT Technology Review*, *120*(3), 54–61.

Kouziokas, G., Chatzigeorgiou, A., & Perakis, K. (2017). Artificial intelligence and regression analysis in predicting ground water levels in public administration. *European Water Publications*, *57*, 361–366.

Kouziokas, G. N. (2017). The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment. *Transportation Research Procedia*, *24*, 467–473. doi:10.1016/j.trpro.2017.05.083

Krausová, A. (2017). Intersections between law and artificial intelligence. *International Journal of Computer*, *27*(1), 55–68.

Kreinczes, C. (2016). *Artificial Intelligence Innovation Report*. Edited by Deloitte. Retrieved July 2, 2018, from https://www2.deloitte.com/content/dam/Deloitte/at/Documents/human-capital/artificial-intelligence-innovation-report.pdf

Lee, S. G., & Sathikh, P. (2013). A framework for effective human-to-machine communication for artificial interactive systems. In U. Lindemann, S. Venkataraman, Y. S. Kim, S. W. Lee, P. Badke-Schaub, & K. Sato (Eds.), *DS 75-7: Proceedings of the 19th international conference on engineering design (ICED13), Design for harmonies, Vol.7: Human behaviour in design*. Seoul, Korea: Design Society.

Legg, S., & Hutter, M. (2006). *A formal measure of machine intelligence*. In Proc. 15th Annual Machine Learning Conference of Belgium and The Netherlands (Benelearn 2016), 73–80.

Legg, S., & Hutter, M. (2007). A collection of definitions of intelligence. *Frontiers in Artificial Intelligence and Applications*, *157*, 17–24.

Lever, R. (2017). *Tech world debate on robots and jobs heats up*. Edited by Phys.org. Retrieved June 20, 2018, from https://phys.org/news/2017-03-tech-world-debate-robots-jobs.html

Lin, Steven Y., Shanafelt, Tait D. & Asch, Steven M. (2018). Reimagining clinical documentation with artificial intelligence. *Mayo Clinic Proceedings*, *93*(5), 563–565.

Mankins, P. (2015). *Can we design trust between humans and artificial intelligence?* Edited by Fast Company. Retrieved June 22, 2018, from https://www.fastcodesign.com/3047500/can-we-design-trust-between-humans-and-artificial-intelligence

Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., & Dewhurst, M. (2017). *A future that works: Automation, employment, and productivity*. Edited by McKinsey Global Institute. Retrieved June 20, 2018, from https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works_Full-report.ashx

Massaro, T. M., Norton, H. L., & Kaminski, M. E. (2017). Siri-ously 2.0: what artificial intelligence reveals about the first amendment. *101*(6), 2481–2525.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, *6*(3), 175–183. doi:10.1007/s10676-004-3422-1

McCarthy, John, Minsky, Marvin L, Rochester, Nathaniel & Shannon, Claude E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*, *27*(4), 12–14.

Mehr, H. (2017). *Artificial intelligence for citizen services and government*. Cambridge, MA: Harvard Kennedy School, Ash Center for Democratic Governance And Innovation. Retrieved July 2, 2018, from https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

Metaxiotis, K., Ergazakis, K., Samouilidis, E., & Psarras, J. (2003). Decision support through knowledge management. The role of the artificial intelligence. *Information Management & Computer Security*, *11*(5), 216–221. doi:10.1108/09685220310500126

Miailhe, N. (2017). The policy challenges of automation. *Field Actions Science Reports. The Journal of Field Actions*, *17*, 66–71. Retrieved June 8, 2018, from https://journals.openedition.org/factsreports/4476

Microsoft (2018). Machine translation. Retrieved July 2, 2018, from https://www.microsoft.com/en-us/translator/mt.aspx

Minchev, Z. (2016). Challenges to human factor for advance persistent threats proactive identification in modern social networks. *Information & Security: An International Journal*, *34*, 57–70.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 1–21. doi:10.1177/2053951716679679

Müller, V. (2014). Risks of general artificial intelligence. *Journal of Experimental & Theoretical Artificial Intelligence*, *26*(3), 297–301. doi:10.1080/0952813X.2014.895110

Nagenborg, M., Capurro, R., Weber, J., & Pingel, C. (2008). Ethical regulations on robotics in Europe. *AI & Society*, *22* (3), 349–366. doi:10.1007/s00146-007-0153-y

Newman, D. (2017). *Artificial intelligence: To be feared or embraced*. Edited by Forbes. Retrieved June 29, 2018, from https://www.forbes.com/sites/danielnewman/2017/07/11/artificial-intelligence-to-be-feared-or-embraced/#3881ae166f09

Nilsson, N. J. (2003). *Artificial intelligence: A new synthesis* (5th ed.). San Francisco, CA: Morgan Kaufmann Publishers.

Nitto, H., Taniyama, D., & Inagaki, H. (2017). *Social acceptance and impact of robots and artificial intelligence: Findings of survey in Japan, the U.S. and Germany*. Edited by Nomura Research Institute (NRI Papers, 211). Retrieved June 21, 2018, from https://www.nri.com/~/media/PDF/global/opinion/papers/2017/np2017211.pdf

Patrick, E. A., & Fattu, J. M. (1986). *Artificial intelligence with statistical pattern recognition*. Englewood Cliffs, NJ: Prentice Hall.

Patrick., L., Abney, K., & Bekey, G. A. (Eds.). (2012). *Robot ethics: The ethical and social implications of robotics*. Cambridge, MA: MIT Press.

Petit, N. (2018). Artificial intelligence and automated law enforcement: A review paper. *SSRN Journal*. doi:10.2139/ssrn.3145133

Power, D. J. (2016). "Big Brother" can watch us. *Journal of Decision Systems*, *25*(Sup1), 578–588. doi:10.1080/12460125.2016.1187420

Purdy, M., & Daugherty, P. (2016). *Why artificial intelligence is the future of growth*. Edited by Accenture. Retrieved July 2, 2018, from https://www.accenture.com/t20170927T080049Z__w__/us-en/_acnmedia/PDF-33/Accenture-Why-AI-is-the-Future-of-Growth.PDFla=en

PwC. (2017). *Sherlock in health: How artificial intelligence may improve quality and efficiency, whilst reducing healthcare costs in Europe*. Retrieved June 7, 2018, from https://www.pwc.de/de/gesundheitswesen-und-pharma/studie-sherlock-in-health.pdf

Quraishi, F. F., Wajid, S. A., & Dhiman, P. (2017). Social and ethical impact of artificial intelligence on public: A case study of university students. *International Journal of Scientific Research in Science, Engineering and Technology*, *3*(8), 463–467.

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, *20*(1), 5–14. doi:10.1007/s10676-017-9430-8

Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). *Reshaping business with artificial intelligence: Closing the gap between ambition and action*. Edited by MIT Sloan Management Review. Retrieved July 2, 2018, from https://sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/

Rich, Elaine, Knight, Kevin & Nair, Shivashankar B. (2009). Artificial intelligence. Third edition. New Dehli, India: Tata McGraw-Hill.

Roberts, A. (2017). *Five big challenges to AI adoption and success*. Edited by ClickZ. Retrieved July 2, 2018, from https://www.clickz.com/five-big-challenges-to-ai-adoption-and-success/112795/

Rosa, A., Feyereisl, J., & Team, T. G. (2016). *A framework for searching for general artificial intelligence*. CoRR abs/1611.00685, 1–54. Retrieved from http://arxiv.org/abs/1611.00685

Rössler, B. (2005). *The value of privacy*. Cambridge, UK: Polity Press.

Russel, Stuart & Norvig, Peter (2010). Artificial intelligence: A modern approach. Third edition. Englewood Cliffs, NJ: Prentice Hall.

Russell, S. (2015). Take a stand on AI weapons. *Nature*, *521* (7553), 415–416. doi:10.1038/521415a

Santoro, M., Marino, D., & Tamburrini, G. (2008). Learning robots interacting with humans: From epistemic risk to responsibility. *AI & Society*, *22*(3), 301–314. doi:10.1007/s00146-007-0155-9

Scherer, M. U. (2016). Regulating artificial intelligence systems: Risk, challenges, competencies, and strategies. *Harvard Journal of Law & Technology*, *29*(2), 353–400.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hassabis, D., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D., (2017). Mastering the game of go without human knowledge. *Nature*, *550*(19), 354–359. doi:10.1038/nature24270

Singh, P. (2017). *CIA set to be run by AI, has 137 pilot projects in place*. Edited by Analytics India Magazine. Retrieved June 29, 2018, from https://analyticsindiamag.com/cia-set-run-ai-137-pilot-projects-place/

Slaughter, A.-M. (2017). *3 responsibilities every government has towards its citizens*. Edited by World Economic Forum. Retrieved June 29, 2018, from https://www.weforum.org/agenda/2017/02/government-responsibility-to-citizens-anne-marie-slaughter/

Solum, L. B. (1992). Legal personhood for artificial intelligences. *North Carolina Law Review*, *70*(4), 1231–1288.

Soo, Z. (2018). *The increasing use of artificial intelligence is stoking privacy concerns in China*. Edited by South China Morning Post. Retrieved June 13, 2018, from http://www.scmp.com/business/companies/article/2135713/increasing-use-artificial-intelligence-stoking-privacy-concerns

State Council of the People's Republic of China. (2017). *China issues guideline on artificial intelligence development*. Retrieved July 2, 2018, from http://english.gov.cn/policies/latest_releases/2017/07/20/content_281475742458322.htm

Sun, T., & Medaglia, R. (2017). *Artificial intelligence and public healthcare service innovation: A service ecosystem perspective*. The 40th Information Systems Research Conference in Scandinavia, IRIS 2017. Halden, Norway: Østfold University College.

Tanaka, T., & Kobayashi, K. (2015). Developing a dividual model using a modular neural network for human-robot interaction. *Journal of Robotics, Networking and Artificial Life*, *2*(1), 34–39. doi:10.2991/jrnal.2015.2.1.9

Tang, A., & Wen, A. (2009). An intelligent simulation system for earthquake disaster assessment. *Computers & Geosciences*, *35* (5), 871–879. doi:10.1016/j.cageo.2008.03.003

Thierer, A., O'Sullivan Castillo, A., & Russell, R. (2017). *Artificial intelligence and public policy. Mercatus research*. Edited by Mercatus Center at George Mason University.

Retrieved July 2, 2018, from https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf

Turilli, M. (2007). Ethical protocols design. *Ethics and Information Technology*, *9*(1), 49–62. doi:10.1007/s10676-006-9128-9

Velasquez, M., Andre, C., Shanks, T., Meyer, S. J., & Meyer, M. J. (2010). *What is ethics? Edited by Markkula Center for Applied Ethics, Santa Clara University*. Retrieved July 2, 2018, from https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/what-is-ethics/

Wisskirchen, G., Biacabe, B. T., Bormann, U., Muntz, A., Niehaus, G., Jiménez Soler, G., & Brauchitsch, B. V. (2017). *Artificial intelligence and robotics and their impact on the workplace*. Edited by International Bar Association Global Employment Institute. Retrieved from https://www.ibanet.org/Document/Default.aspx?DocumentUid=c06aa1a3-d355-4866-beda-9a3a8779ba6e

Zheng, Y., Han, Y., Cui, L., Miao, C., Leung, C., & Yang, Q. (2018). *SmartHS: An AI platform for improving government service provision*. The Thirtieth AAAI Conference on Innovative Applications of Artificial Intelligence (IAAI-18), 7704–7711. Retrieved July 2, 2018, from https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16041/16369