AIGOV

Implementing ethical, trustworthy and fair Artificial Intelligence Systems in Public Sector

D1.1 State of Play Analysis

Version-Status:	V1.0 Final
Date:	30/09/2022
Dissemination level:	PU

The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the **"2nd Call for H.F.R.I. Research Projects to support Faculty Members & Researchers"** (Project Number: 2412).

Deliverable factsheet

Project Number:	2412
Project Acronym:	AIGOV
Project Title:	Implementing ethical, trustworthy and fair Artificial Intelligence Systems in Public Sector
Principal Investigator:	Konstantinos Tarabanis
Scientific Area:	SA9. Management & Economics of Innovations
Scientific Field:	9.4 ICT enabled Innovation, Digitisation and Industrial Renewal
Host Institution:	University of Macedonia
Collaborating Organization:	Region of Central Macedonia, Greece

Deliverable title:	State of Play Analysis
Deliverable number:	D1.1
Official submission date:	30/09/2022
Actual submission date:	30/09/2022
Author(s):	Areti Karamanou, Dimitrios Zeginis, Maria Zotou, Antonia Psaltou, Evangelos Kalampokis, Konstantinos Tarabanis

Abstract:	This document is the first deliverable, entitled D1.1 “State of Play Analysis”, of the first work package of the AIGOV project. D1.1 is the direct outcome of Task 1.1 that includes a thorough state-of-the-art analysis of scientific literature, technical reports, and national strategies regarding the implementation of AI technologies in public sector. To fulfill the requirements of this deliverable, a systematic literature review was carried out using the snowballing literature review method.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

The following table presents the acronyms used in the deliverable in alphabetical order.

Abbreviation	*Description*
AI	Artificial Intelligence
A/IS	Autonomous and Intelligence Systems
IEEE	Institute of Electrical and Electronics Engineers
ML	Machine Learning
NLP	Natural Language Processing
RAI	Responsible Artificial Intelligence
WP	Work Package

Executive Summary

This document is the first deliverable, entitled D1.1 “State of Play Analysis”, of the first work package of the AIGOV project. The objective of the AIGOV project is to facilitate the implementation of fair, ethical, trustworthy, and robust, both from a technical and social perspective, Artificial Intelligence (AI) systems in Public Administration (PA) with a focus on users and those who may be affected.

D1.1 is the direct outcome of Task 1.1 that includes a thorough state-of-the-art analysis of scientific literature, technical reports, and national strategies regarding the implementation of AI technologies in public sector. The aim of this activity is to identify successful cases of applying AI in public sector around the globe in order to create added public value or to solve existing problems. The cases along with the targeted problems will be categorised according to the most important dimensions that will be identified. Moreover, the most important challenges and implications described in the literature will be analysed and categorised along with available tools and methods to address them. These include societal, economic, ethical, cultural, organisational, and legal implications and challenges. The result of this task will constitute the solid theoretical basis for the other WPs.

To fulfill the requirements of this deliverable, a systematic literature review was carried out using the snowballing literature review method. The objective was to locate scientific literature, technical reports, and national strategies that deal with the adoption of AI technologies in the public sector. First, Google Scholar was searched using the search strings “Artificial Intelligence” AND “public sector”. This resulted in a plethora of studies. Next, a set of inclusion criteria were applied:

- Include only papers written in English
- Include only papers that are available in full-text
- Include only papers published the last five years (i.e., 2017-2022).

The criteria were applied and the papers were further eliminated after (i) conducting a title screening, (ii) reading their abstracts, and (iii) reading their full text.

1 Introduction

The aim of this section is to introduce the background of the work pursued with Task1.1 “Sota Analysis & Background Knowledge” of the AIGOV project. The scope and the objective that the current document has set out to achieve are presented in sub-section 1.1. The intended audience for this document is described in sub-section 1.2 while sub-section 1.3 outlines the structure of the rest of the document.

1.1 Scope

The present document is the deliverable “State of Play Analysis” (henceforth, referred to as D1.1) of the AIGOV project. The main objective of D1.1 is to document the results of Task1.1 “Sota Analysis & Background Knowledge” of WP1.

1.2 Audience

The intended audience for this document includes who are interested in cases, challenges, strategies etc. of deploying Artificial Intelligence in the public sector.

1.3 Structure

The structure of the document is as follows:

- Section 2 provides the methodology used for this deliverable.
- Section 3 presents Artificial Intelligence including its definition and potential, its stakeholders, history, future, impact, benefits, challenges, and ethics.
- Section 4 presents the state of play of Artificial Intelligence in the Public Sector including a number of cases found in literature.
- Finally, Section 5 draws conclusions.

2 Methodology

In this section we present the methodology that we follow in order to achieve the objectives of the first task of WP1.

To fulfill the requirements of this deliverable, a systematic literature review was carried out using the snowballing literature review method recommended by Wohlin [51]. This approach is particularly useful for expanding existing literature reviews by including new perspectives. The main advantage of snowballing is that it focuses on cited or referenced papers, which reduces noise when compared to the database approach. Additionally, it is common for new studies to cite previous relevant studies or systematic literature reviews in a specific area [51]. The snowballing process involves identifying a tentative set of papers and then conducting forward and backward snowballing in iterations. Wohlin suggested using Google Scholar to identify the starting set of papers and avoid publisher bias [26].

The objective was to locate scientific literature, technical reports, and national strategies that deal with the adoption of AI technologies in the public sector. First, Google Scholar was searched using the search strings “Artificial Intelligence” AND “public sector”. This resulted in a plethora of studies. Next, a set of inclusion criteria were applied:

- Include only papers written in English
- Include only papers that are available in full-text
- Include only papers published the last five years (i.e., 2017-2022).

The criteria were applied, yielding a total of 3,720 results. After conducting a title screening, 3,173 results were eliminated. The remaining papers' abstracts were read, resulting in 189 papers being retained. Subsequently, the full papers were read, and independent judgments were made regarding their inclusion or exclusion. Following the application of all the inclusion criteria and thorough reading of each paper, a total of 53 candidate papers were selected. These papers were thoroughly analyzed in order to find successful cases, stakeholders, impacts and benefits, challenges, ethics, benefits, and tools for applying AI in the public sector.

3 Artificial Intelligence

3.1 Definition and Potential

Although Artificial Intelligence (AI) lack of an official definition, many definitions have been given by various global organizations over time. For example, one of the oldest definitions is the one that was given in 1956 by John McCarthy, who is considered to be the father of AI. John McCarthy define AI as *"the science and engineering of making intelligent machines"*. Since then more other definitions have also been proposed with many of them reflecting McCarthy's approach. A typical example is OECD's more detailed definition for AI systems [38]:

"An AI system is a machine based system that can, for a given set of human defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. It does so by utilising machine and/or human based inputs to:

1. *perceive and/or analyse real and/or virtual environments;*
2. *abstract such perceptions/analyses into models manually or automatically; and*
3. *use model interpretations to formulate options for outcomes.*

AI systems are designed to operate with varying levels of autonomy."

All around the world, 42 governments have endorsed this definition of AI, which views it as capable of mimicking human behavior and matching but not exceeding human intelligence [5]. In this context, literature occasionally refers to hypothetical general AI systems that could significantly outperform human abilities as "Artificial Super Intelligence," or "superintelligence" [4].

Additionally, the World Economic Forum defines AI systems in terms of their own system-based perspective as "systems that act by sensing, interpreting data, learning, reasoning, and choosing the best course of action" [53].

Another approach envisages AI from the perspective of designing systems: "the discipline of creating algorithms that can learn and reason" [38]. A similar definition of AI is provided by the High-Level Expert Group on Artificial Intelligence of the European Commission, which describes it as "systems that display intelligent behavior by analysing their environment and taking actions - with some degree of autonomy - to achieve specific goals" [11].

The British government also offers a definition that is broader in scope. According to this definition, AI is a "research field spanning philosophy, logic, statistics, computer science, mathematics, neuroscience, linguistics, cognitive psychology, and economics" that makes use of "digital technology

to create systems capable of performing tasks commonly thought to require intelligence"[1]. Additionally, according to the Luxembourg regulatory authority, AI solutions should "focus on a small number of intelligent tasks and be used to assist humans in making decisions" [8].

As another option, the Institute of Electrical and Electronics Engineers (IEEE) Standards Association concentrates more on the Autonomous and Intelligence Systems (A/IS) than AI per se[2].

Although there is no consensus regarding the definition of AI, two perspectives are used to set the expectations on the level of AI's intelligence [5]:

- General AI. This first perspective is also known as "strong AI" or the "Artificial General Intelligence" (AGI) perspective.
- Narrow AI. This perspective is the more granular view of AI and is also known as "weak AI", "applied AI", or "Artificial Narrow Intelligence" (ANI).

All AI that has been used or that is being used now is Narrow, meaning that all AI algorithms and systems are accomplishing their tasks in a human-like way rather than being capable of performing or outperforming any task (intellectual or cognitive) a human can perform. Narrow AI makes use of the fact that while humans are still more adept at handling ambiguous situations or those requiring intuition, creativity, emotion, judgment, and empathy, computers are better at processing large amounts of data quickly and consistently as well as carrying out tasks based on logical and explicit rules [5].

Although AI has been already defined in 1956, it only recently started becoming so popular. The key factors that contributed to this recent rise of AI include:

- Data explosion [4] [55]. The explosion of data that are being produced the recent years within businesses, organisations, governments, and societies and that are coming from various sources and being of various types (e.g., IoT data, multimedia including text, audio, and video) makes AI possible and necessary as well. Especially in the public sector, many government operations relate to the generation and maintenance of citizens' registries (e.g., births, marriages). Governments also maintain tremendous amounts of other data including geospatial and weather data from satellites, property records, and health and safety records, among many others. In recent years, governments have increasingly pursued the publication of government data in machine-readable formats through open government data (OGD)

[1] https://www.gov.uk/government/publications/understanding-artificial-intelligence/a-guide-to-using-artificial-intelligence-in-the-public-sector
[2] https://standards.ieee.org/industry-connections/ec/autonomous-systems/

policies, and associated portals for datasets and Application Programming Interfaces (APIs). This contributes to the availability of data for AI systems to leverage. [4]

- Maturity [4]. A significant body of knowledge has been accumulated with many different projects launched over the last few decades. Old algorithms and models have been refined and new ones have emerged. Programming languages and frameworks have been developed and refined and many new applications created as more people become familiar with AI. For example, the idea of artificial neurons has been around since the 1940s, but the development of Deep Learning AI only took off during the last decade.
- Democratisation of computers and programming [4]. While technology has improved, it has also become available to a growing number of people. New users today are also more connected and better equipped to learn and exchange information about AI. Collaborative platforms and tools supported by vibrant communities are making programming and coding possible not only for experts and companies, but also for individuals from all backgrounds. For example, GitHub and Kaggle allow people to collaborate on digital solutions. This also allows bottom-up ideas and solutions to emerge in ways that were not often possible in the past. Freely available online courses and tutorials, including those provided by the public sector, also contribute to this democratisation.
- Better technology and increased processing power (especially in the cloud) [4] [55]. In the cloud, massive amounts of computing power are available on demand. The cost to process an AI application today is a small fraction of what it would have cost a decade ago, if it were even possible. Computers today are cheaper, have more computing power and take up much less physical space. This increase in processing power allows devices to run larger and more complex programmes, and process more data faster. Data storage costs have also decreased dramatically.
- New algorithm types including neural networks and deep learning that have come to maturity in the last decade or so [55]. These new algorithm types have been a primary driver behind such AI use cases as image, video and voice recognition, speech recognition and natural language processing, and natural language generation.
- Other, socio-political factors [55] like the growth of technology capabilities in China, the relative absence of data privacy standards in the United States, the massive acceptance of digital business offerings in many countries around the world, and spending by government bodies.

AI can be applied to various industries using a plethora of technology types. Figure 1 presents a heatmap for the volume of use cases per industry and technology type. The colors used indicate the number of related use cases; the darker the color the largest the number of cases. According to the Figure, most cases are related to:

- Using knowledge management in professional cases and the public and social sector
- Using Vision in the sectors of consumer goods and services, and public and social sector
- Using Speech recognition in the consumer goods and services sector
- Using Natural Language Processing (NLP) in the consumer goods and services sector
- Using analysis, optimization and prediction in the sectors: basic materials; financial services; technology and social sector; telco; transportation; utilities
- Using robotics and sensors in the sectors: basic materials; energy; industries; transportation; utilities

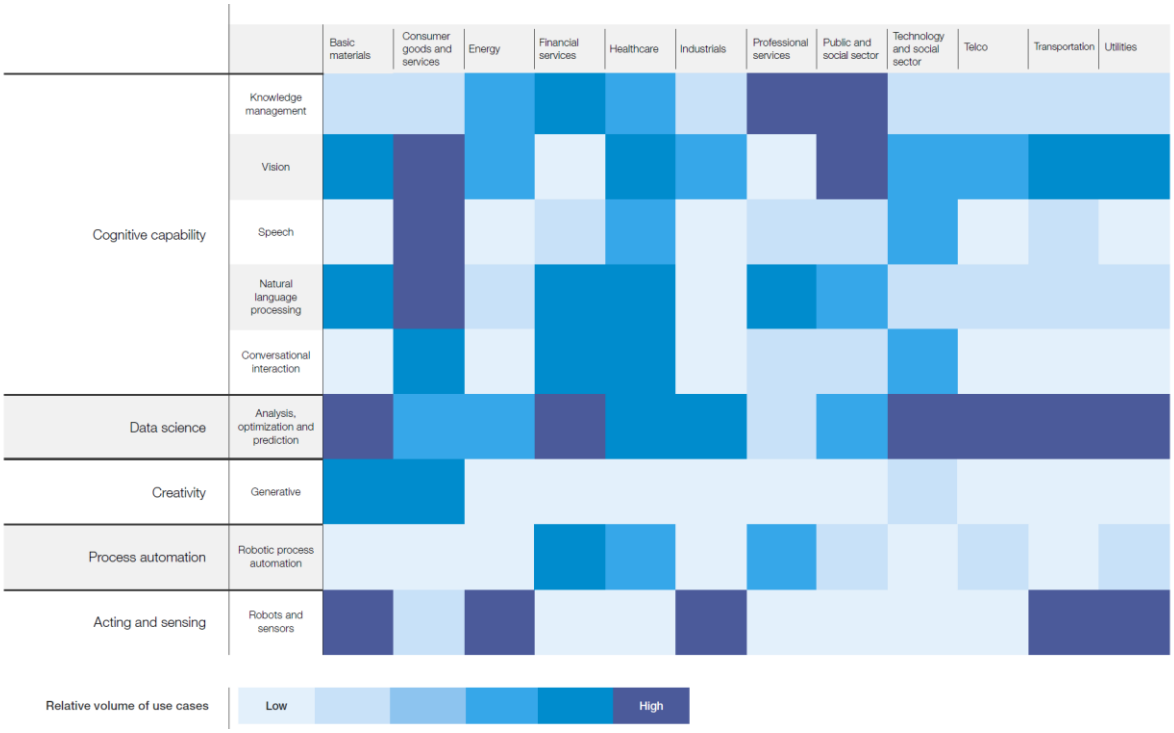

Figure 1 Common AI use cases by industry and technology type [55]

In addition, according to a recent McKinsey's report [29], the most commonly adopted AI case uses service-operations optimization followed by AI-based enhancement of products, contact-center automation, product-feature optimization, predictive service and intervention, customer-service analytics, creation of new AI-based products, customer segmentation, risk modelling and analytics, and fraud and debt analytics.

However, AI is not a general purpose solution which can solve every problem [37]. Current applications of AI focus on performing narrowly defined tasks. AI generally cannot: be imaginative; perform well

without a large quantity of relevant, high quality data; infer additional context if the information is not present in the data.

3.2 Machine Learning

The most widely used form of AI is Machine learning (ML), a field at the intersection of computer science, mathematics and statistics. ML mainly focuses on developing algorithms based on large volumes of data and of various more or less complex structures of data, in order to make predictions, recommendations or decisions. It has contributed to various innovations including navigating autonomous vehicles, predicting infrastructure failures, or recommending products online.

The main types of ML are:

1. supervised learning. In supervised learning the AI model is trained based on past, labeled observations for a particular output. A typical example of supervised learning problem is the prediction of the prices of houses in a region based on labeled data that describe, for example, structural aspects of the houses (e.g., number of rooms etc.). This type of learning is generally useful in clearly defined problems where the content and structure of the data are sufficiently described. Supervised learning problems include regression and classification. The first is used to predict the value of a variable, while the latter to predict the group to which a new data point will be classified. The house prediction mentioned above is, for example, a regression problem, while predicting if a customer will churn is a classification problem.
2. unsupervised learning. In this type of ML the AI model is trained based on unlabeled and unclassified data in order to gain new insights on the data. A typical unsupervised learning problem is market segmentation that aims to cluster customers (e.g., of a superstore) based on their previously purchased products and consuming behavior (e.g., how frequent they order a product). This type of learning is generally used for unguided pattern discovery and is based on identifying common characteristics between different data points. A common approach to do so is using clustering.
3. reinforcement learning, which allows an AI model to learn as it performs a task in an environment. For example, a reinforcement learning model could be trained to predict the next move of an agent in a chess game. Specifically, reinforcement learning is based on getting feedback from the environment about an agent's previous interactions. This helps the agent learn through the trial and error method, since error is "punished" while success is "rewarded: by the environment. The agent then automatically adjusts its behavior over time producing more refined actions. Reinforcement learning has recently grown popularity because of the advances in computing capabilities.
4. semi-supervised learning, which combines supervised and unsupervised learning. Specifically, semi-supervised learning trains the AI model using both labeled and unlabeled data.

The final area of Machine Learning is deep learning. Deep learning is based on neural networks that are algorithms designed to mimic the way the human brain processes information. Deep learning refers to deep neural network, which is a specific configuration where neurons are organized in multiple successive layers [26]. The main distinction with the classical ML algorithms lies in the design of deep learning algorithms, which is inspired by the biology of human brains. Indeed, deep learning is often discussed in conjunction with Artificial Neural Networks (ANN). The "depth" of an Artificial Neural Network relates to its number of hidden layers. Deep learning algorithms use ANNs which have two or more hidden layers.

3.3 Stakeholders

There are two sets of stakeholders for building and using AI [55]:

- Stakeholders that are required to deliver and manage AI projects, and
- Stakeholders whose trust is required to ensure the successful adoption of AI.

In the first instance, AI project and operational teams need to be built. As with many digital projects, it is important to balance technology and business skills with domain knowledge. Broadly speaking, these roles include:

- Data engineers: responsible for creating the nuts and bolts of data infrastructure and pipelines
- Business intelligence, insight and analytics professionals: responsible for ad hoc reporting and dashboards
- Data scientists: responsible for building predictive algorithms and ensuring the recommendations of the data team are statistically robust
- ML engineers: responsible for scaling data science models and putting them into production
- Data product managers: responsible for coordinating all the above resources as well as aligning with business stakeholders and engineering to deliver value
- Data governance specialists: responsible for managing and documenting data assets and ensuring they are ethical and compliant.
- Depending on the desirable outcome, it is critical to have experienced product managers in the team to ensure the successful delivery of solutions that satisfies the end users' needs.

3.4 History

A brief history of AI including significant milestones is illustrated in Figure 2. AI was first introduced in early 1950's when AI theory went to practice with the first AI programs play checkers and prove theorems being used. Then, in middle 1960's was the first rise of AI. In this decade the Natural

Language Processing (NLP) method was first used to solve algebra word problems. NLP was further improved during the next decade (1970's). In this decade the first scientific discoveries by AI in chemistry were also first made. However, this decade is also known as the "first AI winter" because of the reduced interest and funding for the field.

Although in the period 1980-1987 interest started to grow around the application of AI expert systems, the following 7 years (1987-1993) are known as the "second AI winter" that actually began with the sudden collapse of the market for specialized AI hardware in 1987 [30]. Nevertheless, this period was followed by the so called "AI Spring" until the 2000s. AI springs are periods when AI is in full bloom. The specific time period includes significant milestones like the first publicly available speech recognition software (namely "Dragon Systems") produced in 1997 as well as the appearance of the first AI-driven consumer entertainment products (namely "Furby") in 1998.

Then, the 2000s are known as the years when AI goes interplanetary with autonomous rovers exploring Mars in 2004. In the same decade, new concepts emerged including deep learning (2006). In addition, significant achievements like the first Google's autonomous car and the first usage of speech recognition in mobile phones by Google and Apple are happening.

The final and most recent decade regards the years of AI's explosive growth. In the early 2010's image recognition and classification using AI is achieved. In addition, in 2017, AI beats human players in a series of strategy games including Go, Defense of the Ancients (Dota) 2, and No Limit Hold' em Poker. Finally, the interval 2016-2019 is a significant milestone since the US Food and Drug Administration (FDA) approved for the first time the usage of AI in medicine.

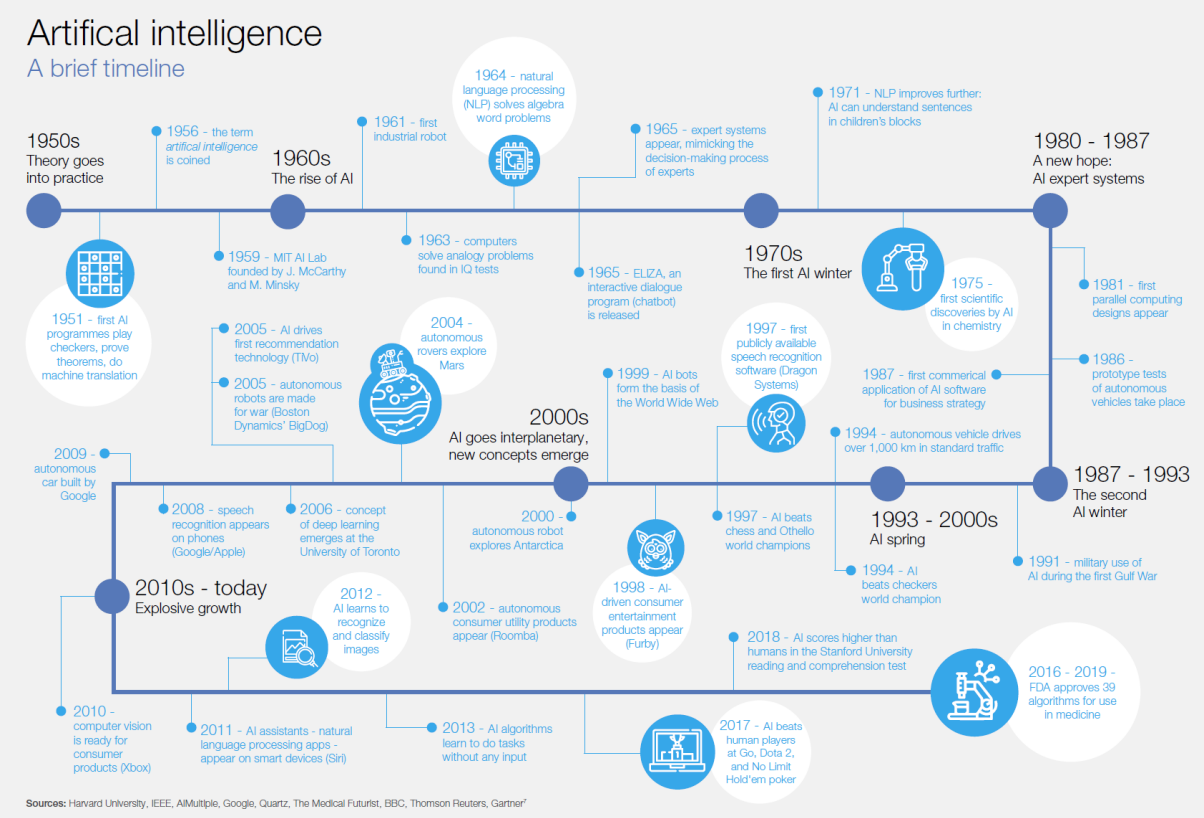

Figure 2 A brief history of Artificial Intelligence [55]

3.5 The future of AI

The 2022 Gartner Hype Cycle™ for Artificial Intelligence (Figure 3) features "must-know" innovations expected to drive extensive benefits to any organization.

In general, the Gartner Hype Cycle[3] is for Artificial Intelligence a graphic representation of the maturity and adoption of AI technologies and applications, and how they are potentially relevant to solving real business problems and exploiting new opportunities. The cycle actually provides a perspective on how an AI technology or application will develop over time and can be used as a reliable source of knowledge to manage its deployment in the context of particular business objectives.

Each Hype Cycle drills down into the five key phases of a technology's life cycle. These include:

[3] https://www.gartner.com/en/research/methodologies/gartner-hype-cycle

- **Innovation Trigger**: Things begin with a possible technological advance. Media attention and early proof-of-concept reports lead to a lot of exposure. There are frequently no useable goods and questionable commercial feasibility.
- **Peak of Inflated Expectations**: Early exposure results in a number of success stories, frequently followed by a large number of failures. While some businesses act, many don't.
- **Trough of Disillusionment**: As experiments and implementations fall short of expectations, interest wanes. Technology manufacturers either succeed or fail. Investments only continue if the remaining providers enhance their goods in a way that appeals to early adopters.
- **Slope of Enlightenment**: More examples of how technology might help the business start to emerge and gain wider acceptance. The release of second- and third-generation goods by technological companies. More businesses support pilots, while conservative businesses maintain their caution.
- **Plateau of Productivity:** Takeoff in mainstream acceptance is beginning. More precise criteria for determining provider viability are now available. The technology is clearly benefiting from its broad commercial application and relevance.

The AI innovations on the Hype Cycle reflect complementary and sometimes conflicting priorities across four main categories:

- *Data-centric AI*, which puts more emphasis on improving and enlarging the data needed to train the algorithms. Traditional data management is disrupted by data-centric AI, but organizations that invest in AI at scale will develop to preserve timeless classic data-management ideas and extend them to AI in two ways: by adding capabilities needed for convenient AI development by an AI-focused audience that is unfamiliar with data management; and by using AI to enhance and supplement timeless classics of data governance, persistence, integration, and data quality. Synthetic data, knowledge graphs, data labeling, and annotation are examples of data-centric AI innovations.
- *Model-centric AI*. Despite the transition to a data-centric strategy, AI models still require maintenance to make sure the results continue to guide our activities in the right direction. Physics-informed AI, composite AI, causal AI, generative AI, foundation models, and deep learning are examples of innovations in this field.
- *Applications-centric AI*. Innovations of applications-centric AI include AI engineering, decision intelligence, edge AI, operational AI systems, ModelOps, AI cloud services, smart robots, natural language processing (NLP), autonomous vehicles, intelligent applications and computer vision.

- *Human-centric AI*. This includes AI trust, risk and security management (TRiSM), responsible AI, digital ethics, and AI maker and teaching kits.

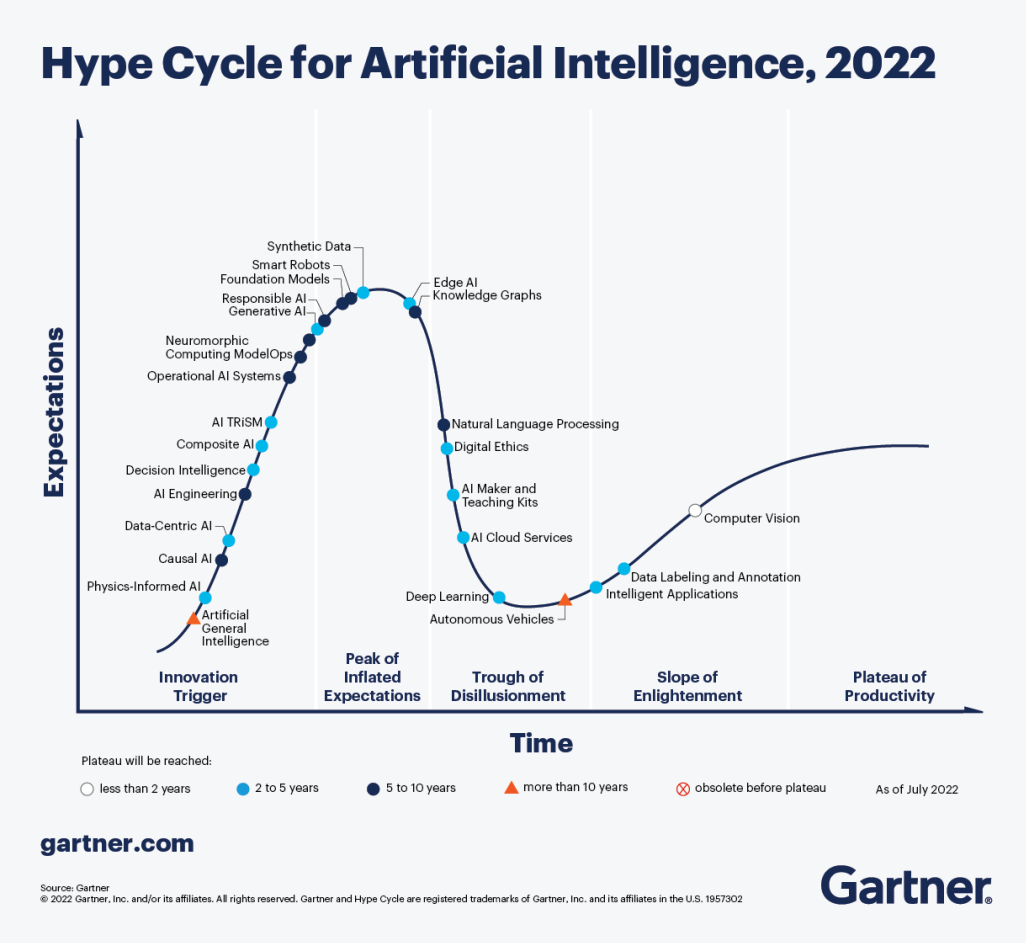

Figure 3 Gartner's Hype Cycle for Artificial Intelligence, 2022 (source: https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle)

3.6 Impact and Benefits of AI

The potential of AI is huge and many researches have tried to estimate it. For example, according to a recent calculation made by the analysts of Accenture, AI has the potential to double annual economic growth rates and also boost labor productivity by up to 40% in developed countries by 2035 [1]. Another research estimates that AI can potentially contribute $13 trillion to the global economy by 2030 [3]. In addition, a McKinsey's recent survey [29] reports at least 5% of earnings before interest and taxes that's attributable to AI. Respondents to the same survey report utilizing AI to considerably reduce costs across all functions, with the highest year-over-year changes in the shares reporting cost takeout from using AI in product and service development, marketing and sales, and strategy and

corporate finance. Finally, by one estimate, AI's contribution to the United Kingdom could be as large as 5% of GDP by 2030 [42].

The main benefits of AI can be summarized in the following [55]:

- **Increased Workforce Productivity**. The productivity of the workforce is increased when then the same number of employees can do more work, or fewer employees can do the same amount of work. Although this could potentially mean that increasing productivity could cause job loss, apart from the case of physical and software robotics, AI is not generally replacing workers. Nevertheless, AI frees up employees giving them more time to perform more complex tasks. Additionally, productivity enhanced by AI may mitigate the negative effects of an aging population and a diminishing labor force participation rate.
- **Better decision making**. Artificial intelligence significantly improves decision-making. It streamlines, expedites, and makes the process more data-driven. With the help of AI, you can quickly make simple decisions, solve difficult issues, start strategic changes, identify risks, and evaluate the success of your entire company or organization.
- **More targeted marketing offers and advertising**. ML is already used in digital marketing to target certain publishers and specific customers with digital adverts.
- **Efficient sales processes**. AI can be used to automate a large part of the sales process based on "propensity to buy" models, i.e., AI model that predict how likely a customer is to purchase a particular product.
- **Optimized supply chains**. The complete supply chain is clearly outlined by AI, which also asserts that it can forecast demand, supply, and movements more accurately and quickly than any human data analyst. This helps firms, despite how chaotic global events may get, make wise judgments and increase resilience.
- **Better managed human resources**. AI provides allows improving the employee experience by automating repetitive, low value tasks and freeing up time to focus on the more strategic, creative work that HR teams need and want to get done. Instead of spending time overseeing every step of the new employee onboarding process, those steps can be intelligently automated leaving teams to invest more time in more important tasks such as mentoring and gathering feedback.
- **More qualitative customer service**. Chatbots and AI intelligent agents have the ability to provide superior customer support that is always available.
- **Better communication with employees**. AI technologies are also being used within organizations to interface with employees.

- **Manufacturing support**. AI is also of benefit to manufacturing companies such as in vehicles, telecommunications and computing devices, industrial machinery, home appliances, and others. For example, vehicle manufacturers are rapidly trying to increase the autonomy of their vehicles. Robots are also used to transform various aspects of the manufacturing process itself. AI may also improve the efficiency and dependability of industrial machinery, therefore minimizing or completely eliminating unscheduled downtime.
- **Healthcare support**. The potential of AI in the support of healthcare is huge, although it has not yet been fully implemented in clinical settings. Example use cases include:
    - Patient diagnosis and treatment based on ML applications (e.g., precision medicine, radiological image recognition).
    - In research laboratories, AI applications have matched or surpassed human skills for illness diagnosis and treatment advice for a wide range of conditions, including cancer, diabetes, and sepsis.
    - AI can also be used to increase patient engagement in care, using customized "nudges" to change behaviour.
    - Speed up the development of and lower the cost of pharmaceutical medications. Although the initial excitement around AI in drug development has partially subsided, several start-ups in the pharmaceutical sector have R&D procedures that extensively rely on ML on genomic and proteomic data.
    - Guide "structural biology" to design new drugs and vaccines, and has already been used to successfully predict the structures of protein folding.
    - Use ML to target the physicians and patients who are likely to achieve better outcomes with their drugs.
- **Sustainability**. AI could make energy consumption more efficient.
- **Improve public services**. AI has the potential to make public services more efficient and effective improving the daily life of citizens. For example, it could provide recommendations for better decisions in many aspects of life, including how to invest, when and how to retire, deciding where to live, what job to take, what school to attend, and so forth.

It is obvious that AI has many positive impacts. Although some of them are still potential, some sectors are have achieved far more benefits than other from adopting AI technology [55].

3.7 Challenges

Although it is obvious that the potential of AI is huge in various domains and through various applications, a number of potential risks can also occur from the extended usage of AI systems.

Potential risks can be classified in two categories [55]; (i) application-level, and (ii) business and national-level risks. The first category includes three groups of risks, namely

- Performance risks;
- Security risks;
- Control risks;

The latter category also includes three groups of risks, namely

- Enterprise risks;
- Economic risks;
- Societal risks;

Table 1 presents 15 application-level risks of applying AI. For each risk, a description, its type and proposed mitigation measures are presented. The majority (11) of identified application-level risks are performance risks, followed by 3 security risks, and 1 control risk.

Challenge	**Type**	***Proposed mitigation measures***
Project shows **signs of bias** or discrimination (e.g., algorithmic bias, sampling bias). Equity and fairness. [29] [37] [55] [4]. In classification or prediction tasks comes with the risk of bias that can put certain groups at a disadvantage (e.g., be sexist/racing when scoring job applicants, inaccurately identify people of color in facial recognition systems). Some degree of bias in AI-based decisions may still be better than decisions made by humans [55].	Performance	• Make sure the model is fair, explainable, and there is a process for monitoring unexpected or biased outputs [37]. • Data professionals actively check for skewed, biased data in data ingestion and at several stages of model development [29]. • Perform small-scale experiments and simulations before implementing algorithms • Regularly evaluate data sets used for training • Involve human reviewers of algorithms and outcomes • Consider the data at hand and determine whether they are influenced by biased factors, and

		whether anything can be done to mitigate perpetuating bias [4].
Lack of data security [37]**.** Security protocols are not in place to make sure you maintain confidentiality and uphold data integrity.	Security	Build a data catalogue to define the security security protocols required [37].
Poor data quality [37] [55]. Data is of poor quality or cannot be accessed.	Performance	• Map the datasets at an early stage both within and outside the organisation. It's then useful to assess the data against criteria for a combination of accuracy, completeness, uniqueness, relevancy, sufficiency, timeliness, representativeness, validity or consistency. [37]
Model cannot be **integrated** [37]	Performance	• Include engineers early in the building of the AI model to make sure any code developed is production-ready [37]
Lack of accountability [37] [55]. For example, Who is responsible in the case of a traffic accident with a driverless car? Who should be blamed for approving parole to a criminal who eventually commits another crime? Who bears responsibility for a large financial loss in algorithmic trading?	Control	• Establish a clear responsibility record to define who has accountability for the different areas of the AI model [37]. • Managers and political leaders need to proactively focus on the accountability for AI-based decisions before potential harm occurs. They also need to carefully consider the stakeholders of AI systems and outcomes (AI developers and designers, business users, customers, institutions) and clarify responsibility and legal liability upfront [55].
Cybersecurity [4][8][29][55]. Along with the significant progress in developing AI systems, new types of attacks to these systems has also	Security	N/A

advanced. These include adversarial attacks, data poisoning (i.e., the manipulation of data used for training, resulting in the AI system learning the wrong insights), model inversion, deepfakes, etc..		
Lack of Interpretability or Explainability [4] [29] [55]. Especially in the case of deep learning, AI truly acts as a black box: the reasons why the algorithm makes specific decisions are not understood. In some cases, explainability may be of lesser concern, as the results themselves are more important than the process by which they were produced (e.g. correctly predicting whether a patient has a disease). However, in some cases (e.g., in public organisations), explainability is key as decisions made based on AI must fully be understood and explainable for reasons of accountability and transparency [4].	Performance	• Link the results produced by the AI model with explanations.
Personal/individual privacy [29] [55]. AI requires increasingly large amounts of data, often involving consumers. This may conflict with people's rights to privacy and autonomy.	Security	• Increase the representation of protected characteristics and/or attributes in our training and testing data as needed [29]. • Legal and risk professionals work with data-science teams to help them understand definitions of bias and protected classes [29]. • Ensure that data practices comply with the relevant regulatory frameworks on the use of personal data and avoid possible privacy violations [55].

		• Clear descriptions of what data is being used, permissions for using personal data, anonymization approaches and transparent algorithms are means to reduce threats to personal data privacy [55]. • Overall, openness about data and how it is managed and used is necessary to ensure customers' trust [55].
Lack of transparency [55]. It is often argued that the decision outcomes of some ML algorithms, such as deep learning, cannot be easily explained due to their complexity and the vast amount of feature layers involved in their production. This could lead to unexplainable personnel evaluations, student test results or sentencing decisions, all of which have already occurred somewhere.	Performance	• Respond to regulators' calls for explainability by either avoiding the use of "black box" AI applications for important decisions or pairing black boxes with so called "explainable AI" algorithms that can provide explanations for the outputs of these algorithms [55]. • Being open about the data that is used and explaining how the model works in non technical terms can also be necessary to ensure customers' trust and to avoid potential dysfunctions triggered by the lack of transparency [55].
Potential brittleness or instability of model [55].	Performance	N/A
Model errors [55].	Performance	N/A
Continuous self-learning and change [55]. AI systems can continuously evolve based on data provided to them, for example during usage. Selflearning can lead to improved accuracy and quality of the decisions	Performance	N/A

the AI system makes, but can also lead to behaviours that were not expected.		
Potential changes in the context [55]. The context within AI systems operate may change. When this happens, the system's performance can deteriorate, possibly in unexpected ways.	Performance	N/A
Multiple and possibly conflicting objectives [55]. Trade-offs typically need to be made when developing AI systems. For example, increased accuracy of the predictions of a system may come at the cost of explainability and transparency; increased safety may come at the expense of potential loss of privacy; improving fairness may require loss in efficiency.	Performance	N/A
Imperfect accuracy [55]. There is no guarantee for the accuracy of the AI system's predictions, recommendations or decisions. In addition, AI systems making decisions can be deployed at scale, possibly affecting millions of stakeholders. Any inaccuracies can translate into a large number of mistakes.	Performance	N/A

Table 1 Application-level risks of AI

In addition, Table 2 presents the Business and National-level risks of AI accompanied by their proposed mitigation measures. In total, 7 Business and National-level risks are identified. Most of them are societal risks (3), followed by enterprise risks (2), one performance and one economic risk.

Challenge	*Type*	*Proposed mitigation measures*
Compliant with legislation, guidance or the government organisation's public narrative [29] [37] [55]	Enterprise	Consult guidance on preparing your data for AI [37]

Organizational reputation [29] [55].	Enterprise	N/A
Physical safety [29]	Societal	
National security [29]	Societal	N/A
Political stability [29] [55].	Societal	N/A
Potential brittleness or instability of model [55].	Performance	N/A
Workforce/labor displacement [29] [55]. Up to half of all jobs could be automated. However, factors like the time frame of automation, the costs, the scarcity and costs of the human workers who might do the activity, may influence the selection of which job to automate. As a result, the predicted job loss has fall from near 50% to below 10% [55].	Economic	• Augment worker and avoid job loss is certainly possible [55]. • Automate tasks, not entire jobs. Relatively few jobs have so many structured and quantitative tasks that they can be fully automated. Job loss because of technology has been found as a relatively slow process. • Use AI to create many new jobs and tasks.

Table 2 Business and National-level risks of AI

3.8 AI Ethics

A number of initiatives including the European Ethics Guidelines for Trustworthy AI [14], the Ethical Principles for Artificial Intelligence [49] have identified a number of ethical principles that should be considered should be considered when designing and deploying AI in various domains. These include the following 15 ethical principles:

1. **Responsible** [49]. Department of Defense personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities.
2. **Equitable** [49]. The Department will take deliberate steps to minimize unintended bias in AI capabilities.

3. **Traceable** [49]. The Department's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedure and documentation.
4. **Reliable** [49]. The Department's AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across their entire life-cycles.
5. **Governable** [49]. The Department will design and engineer AI capabilities to fulfill their intended functions while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior.
6. **Fair** [14] [37] [55]. The development, deployment and use of AI systems must be fair. Substantively, this implies a commitment to 1) ensuring equal and just distribution of both benefits and costs, and 2) ensuring that individuals and groups are free from unfair bias, discrimination and stigmatisation. If unfair biases can be avoided, AI systems could even increase societal fairness. Equal opportunity to access education, goods, services and technology should also be fostered. The use of AI should never lead to people being deceived or unjustifiably impaired in their freedom of choice. Fairness implies that AI practitioners should respect the principle of proportionality between means and ends, and consider carefully how to balance competing interests and objectives. Procedurally, fairness entails the ability to appeal decisions made by AI systems and the humans operating them.
7. **Accountable** [37] [55]. Consider who is responsible for each element of the model's output and how the designers and implementers of AI systems will be held accountable
8. **Respecting Privacy** [37]. Complying with appropriate data policies, for example, the General Data Protection Regulations (GDPR) and the Data Protection Act 2018
9. **Explainable & Transparent** [14] [37] [55]. So the affected stakeholders can know how the AI model reached its decision. Processes need to be transparent, the capabilities and purpose of AI systems must be openly communicated, and the resulting decisions must be explainable to those affected, to the extent possible. Otherwise, a decision cannot be contested. However, an explanation as to why and how a model has generated a particular decision is not always possible. These cases are referred to as "black box" algorithms and require special attention. In those circumstances, other explicability measures (e.g. traceability, auditability and transparent communication on system capabilities) may be required, provided that the system as a whole respects fundamental rights. The degree to which explicability is needed is dependent on the context and the consequences if that output is inaccurate.

10. **Costs effective** [37]. Consider how much it will cost to build, run and maintain an AI infrastructure, train and educate staff and if the work to install AI may outweigh any potential savings.
11. **Compliant with the human rights law** [14] [55]. It is necessary to tie the ethics of AI to specific human rights to limit regulatory ambiguity, but also for the development of human-centric AI for the common good. Ethical principles, laws and human rights are often inextricable, working in congruence across their spheres of application, and the way in which they apply to AI is no different. Applying this in the right context clarifies how technologies may give rise to different ethical concerns and considerations in the development and application of AI. AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, they should augment, complement and empower human skills. The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over AI systems. It should support humans in the working environment and aim for the creation of meaningful work.
12. **Encourage broad participation** [55].
13. **Build capacity across the community** [55].
14. **Compliant to laws** [55]. Laws and regulation represent the public standards that are morally binding on all citizens of a certain jurisdiction and enforced by institutional mechanisms, including punishments such as fines and imprisonment.
15. **Prevention of harm** [14]. AI systems should not cause or exacerbate harm or otherwise adversely affect human beings. This entails the protection of human dignity and mental and physical integrity. AI systems and their environments must be safe and secure. They must be technically robust and not open to malicious use. Vulnerable persons should receive greater attention and be involved in AI development, deployment and use. Particular attention must be paid to adverse impacts due to asymmetries of power or information, such as between governments and citizens.

3.9 Guidelines and Recommendations for Responsible Artificial Intelligence

Besides providing an explanation, it is important how the algorithm is used and for what purpose. With regard to how the algorithm is used, we can identify the following four ‘application scenarios’ [87]:

1. Descriptive – 'What's happening?'

2. Diagnostic – 'Why is it happening?'

3. Predictive – 'What will happen?'

4. Prescriptive – 'What needs to be done?'

In April 2019, the EC published Ethics Guidelines for Trustworthy AI171 to provide guidance on how to design and implement AI systems in an ethical and trustworthy way. The Guidelines were created by the EC's High-Level Expert Group on Artificial Intelligence (AI HLEG), which consists of 52 AI experts from academia, civil society and industry. A core task of the AI HLEG has been to propose AI ethics guidelines that consider issues such as fairness, safety, transparency, the future of work, democracy, privacy and personal data protection, dignity and non-discrimination, among others. The Guidelines maintain that trustworthy AI has three components that work in harmony:

- Lawful. The AI should comply with all applicable laws and regulations
- Ethical. The AI should adhere to ethical principles and values.
- Robust. The AI should avoid unintentional harm from both a technical and social perspective.

Responsible AI (RAI) Guidelines [47] aim to provide a clear, efficient process of inquiry for personnel involved in AI system development (e.g., program managers, commercial vendors, and government partners) to accomplish the following goals: (i) ensure that the DoD's Ethical Principles for AI are integrated into the planning, development, and deployment phases of the AI system lifecycle; (ii) effectively examine, test, and validate that all programs and prototypes align with the DoD's Ethical Principles for AI; (iii) leverage a process that is reliable, replicable, and scalable across a wide variety of programs.

DIU's RAI Guidelines are presented in the form of detailed worksheets that instruct and guide AI vendors, DoD stakeholders, and DIU program managers on how to properly scope AI problem statements. They also provide detailed guidance on the considerations that each of these stakeholders should keep in mind as they proceed through each phase of AI system development. The DIU RAI Guidelines are framed within three major phases of the technical lifecycle of an AI system: planning, development, and deployment. Planning refers to the process of conceptualizing and designing an AI system to solve a given problem; development refers to the iterative process of writing and evaluating the computer code that comprises that system; and deployment refers to the process of using that system to solve the problem in practice.

In addition, OECD's Recommendation adopted in 2019, identifies five complementary, value-based principles for the responsible stewardship of trustworthy AI [40]:

- AI should benefit people and the planet by driving inclusive growth, sustainable development and wellbeing.

- AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.
- There should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them.
- AI systems must function in a robust, secure and safe way throughout their life cycles, and potential risks should be continually assessed and managed.
- Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.

The same Recommendation also identifies five recommendations for policy makers pertaining to national policies and international co-operation for trustworthy AI.

- Investing in AI R&D
- Fostering a digital ecosystem for AI icon
- Fostering a digital ecosystem for AI
- Providing an enabling policy environment for AI
- Building human capacity and preparing for labour market transition
- International co-operation for trustworthy AI

4 Artificial Intelligence in the Public Sector – State of Play

4.1 Current state

Many governments worldwide are increasingly embracing and prioritizing Artificial Intelligence (AI) as a strategic resource to increase their competitiveness and growth [51]. Although today examples of AI adoption are emerging across the globe and at least 50 governments have developed or are in the process of developing an AI strategy, the pace of AI adoption is uneven, and some countries are not even ready for adopting AI. The lowest percentages of AI adoption regard less developed countries. It is indicative that the 2019 AI Readiness Index developed by Oxford Insights doesn't include any country from Africa or Latin America in the list of the top 20 countries [32]. In addition, one of the regions that are lower in the list is Asia-Pacific, except for four governments.

Nevertheless, according to a survey published by Accenture[4], 83% of public sector leadership are both able and willing to adopt intelligent technologies. Another survey conducted by IBM [23] shows that 87% of government executives agreed that "cognitive computing plays a disruptive role in their organizations, and that they intend to invest in cognitive capabilities". In addition, according to a recent study commissioned by Microsoft and conducted by Ernst & Young LLP [32] based on data from more than 200 respondents, as well as on data from interviews with more than 60 leaders in the Public Sector from 12 Western European countries, although 65% of surveyed public organizations view AI as a digital priority and 67% have adopted one or more AI use cases, only 5% have put it into wide use and 4% have been able to scale AI and achieve a high outcome, resulting in organizational transformation. With respect to the three domains of the public sector (namely Health, Public Administration, Transport) included in the study, Health has the highest AI adoption rate with 71% of Health domain respondents having implemented one or more of the identified AI use cases, while 70% of Transportation respondents have implemented an AI solution and achieved the highest impact from the solutions. The increasing interest of the public sector in adopting AI is also reflected by the growing percentage (84%) of public agencies that believe that their spending for AI will increase by 6% in the coming fiscal years [8].

Furthermore, according to another survey [8] that evaluated the preparedness of public organizations for AI, while there are a substantial number of mature government entities blazing a trail in AI (28%), almost the majority (48%) are in the early stages (e.g., using AI to automate simple tasks, creating efficiency gains and allowing the workforce to focus on more valuable work [32]), developing AI primarily through pilots and exploration. This, however, may be impeding further growth. Public

[4] https://www.accenture.com/no-en/insight-ps-emerging-technologies

organizations expect AI to move into the organizational core in the future, contributing to the improvement of society.

4.2 Objectives of AI in the Public Sector

Applying AI in the public sector has various objectives. These include [32]:

- Increase Efficiency [7]
- Employee satisfaction [7] [2].
- Reduce Risk
- Stakeholder Experience
- Greater Sustainability
- Expand Accessibility
- Assure Quality [7]
- Promote Equality [7]
- Qualify Decisions

The role of governments in achieving the above objectives varies. Spesificaly, governments may act as the following roles [47]:

- A **financier** or **direct investor**. Governments may help subsidize the development and adoption of innovative technologies. Some are actively seeking funding through call for projects or pilot tenders. Such programs encompass both public-sector projects and private-sector R&D projects having implications for the broader economy.
- A **smart buyer** and **co-developer**. Governments can act as smart buyer of existing solutions through innovative procurement practices, or as a co-developer through public-private partnerships (PPP) and other forms of collaboration to build new or tailored solutions. Governments can drive innovation from the demand side by steering the development of new solutions directly towards its needs.
- A **regulator** or **rules maker**. Accelerated innovation cycles of emerging digital technologies call for rethinking the types of policy and regulatory instruments used and their implementation. As both an enabler and a user of emerging digital technologies, governments are facing the challenge of determining how – and how much – to regulate them to maximise their innovative potential while minimising the risks for end users. An aspect of this role involves evaluating compliance with the regulations and rules, and taking action, as appropriate, when they are broken.

- A **convener** and **standards-setter**. Governments often have the ability to bring together stakeholders from many parts of the AI ecosystem (e.g. citizens and residents, businesses, organisations and academics) to help realise their objectives and to understand multiple aspects of relevant issues. Governments can also help develop and put in place standards and informal norms for technology in collaboration with these stakeholders.
- A **data steward**. Governments own, or hold on behalf of their people, vast troves of data. Such data can fuel AI-based technologies, especially when well managed.
- A **user** and **service provider**. Governments provide services and tools empowered by, or made possible through, AI technologies, both for the public and for back-office functions. They therefore play a role in using and adopting technologies themselves.

4.3 Benefits of AI for the Public Sector

AI is expected to deliver an additional US $939bn in value, which means 25% increase in the productivity and impact returned from taxpayer dollars, across the public sectors of 16 major developed economies [1]. Apart from the economic impact, AI has the potential to benefit the Public Sector in the following ways:

- Optimize processes [5] [25] [32] [46]. Increase productivity and efficiency of workflows and delivery of service. In general, AI has the potential to improving the internal efficiency of public administration.
- Transform services and make better decisions [5] [25] [32] [26] [46]. Personalize and improve service quality and develop new services and solutions. Help design better policies and improve decision making in the public sector. Provide more accurate information.
- Enable employees [4] [5] [25] [32] [26] [43]. Empower and augment employees to deliver better results with reduced effort. For example, automate repetitive and time-consuming tasks which frees up valuable time of frontline staff. This will shift civil servants' efforts from mundane tasks to high-value work. It is impressive that the average civil servant spends up to 30% of their time on documenting information and other basic administrative tasks [4]. By automating or otherwise avoiding even a fraction of these tasks, governments could save a tremendous amount of money, as well as re-orient civil servants' work around more valuable pursuits, resulting in more engaging jobs.
- Engage stakeholders [5] [25] [32] [26] [46]. Improve experiences for citizens, businesses, partners and others. Enhance citizen participation in the activities of the public sector and improve citizen/government interaction. Provide personalised public services tailored to individual circumstances. Improve communication and engagement with citizen and residents.

When it comes to the public sector workforce, there is good reason to think that AI will do more to change jobs, than replace jobs, particularly in the next two decades [50]. Part of this is via this augmentation of existing jobs, rather than substitution. Another way jobs can change is via redeployment, where freeing the workforce from mundane, 'robotic' tasks allows people to pursue other more valuable roles (many of which have yet to be created). Some of these may involve more human-to-human interaction, others may involve more problem-solving or more creativity. Then there are the many new workers needed to ensure the effective and responsible use of AI. Research has already identified three broad categories – "trainers, explainers and sustainers" – which are already emerging. These are likely just the first of hundreds of job-types that don't exist yet, much like those created since the internet emerged in the 1990s.

Finally, AI can also serve as an enabler of more fundamental change, giving public service providers a chance to re-examine their remit and the value they provide for government and society [2].

4.4 AI Framework

To reap the benefits and objectives of AI, government organizations must be successful in integrating functions with capabilities. Figure 4 presents a comprehensive AI framework to unlock benefits for the public sector. The framework reveals (i) the six functionalities of AI in the public sector, i.e., the six different ways namely, empower, adapt, augment, prevent, prescribe, and automate, through which AI can be applied to the public sector to enhance their services; and (ii) the 5 organizational and technical capabilities, namely data, talent, ethics, and culture that are required in order to succeed with AI in the public sector.

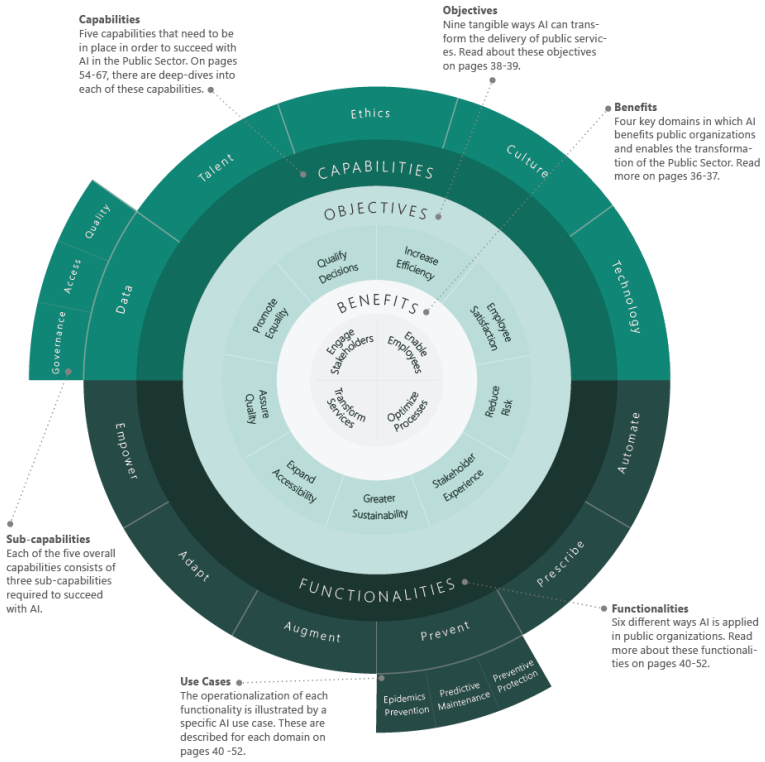

Figure 4 A comprehensive AI framework to unlock benefits for the Public Sector [32]

In addition, according to another report [8], in order to build the organizational capabilities needed for AI at scale, organizations need to adapt their:

- Strategy. Because AI is a transformative technology, alignment on direction and level of ambition is crucial.
- People. Agencies may face challenges around accessing and recruiting necessary technical skills, as well as helping existing employees develop and deploy AI skills.
- Process. AI can be a powerful new tool, but simply embedding it within existing business processes designed for older tools will limit its benefits.
- Data. AI is only as good as the data upon which it is built, and its appetite for data is voracious.
- Ethics. While any technology's deployment should be ethical, AI brings issues such as transparency, privacy, and bias into particular focus.
- Technology and platforms. A variety of models for pursuing AI exist that vary in terms of platforms and ownership of technology (e.g., internal or in partnership), but, in all cases, AI

requires a coherent approach that considers future requirements as AI scales within the organization and its usage evolves.

Successful design, development, and deployment of AI require an appreciation of the nuances of six elements: big data, AI systems, analytical capacity, innovation climate, governance and ethical frameworks, and strategic visioning [25].

In addition, according to the same study (Microsoft and conducted by Ernst & Young LLP [32]), public organizations leading in AI view the following important elements:

1. Make sure that AI is sponsored by senior leadership. This ensures strategic focus, alignment with the organization's mission, and investment readiness.
2. Develop a formalized approach to ensure that AI is managed in a structured way. This can include guidelines, processes and procedures that address the why, when and how to use AI.
3. Cultivate an AI development mindset in the organization. Encourage and incentivize upskilling for hard skills like data science, engineering and domain expertise, and soft skills like innovation, change management and collaboration. The public workforce of tomorrow will need both sets of skills.

4.5 The future of AI in the Public Sector

Figure 5 presents the recent Gartner's Hype Cycle for Digital Government Services, 2022. According to the Hype Cycle, Responsible AI reaches the peak in the cycle, meaning that it is becoming a high priority to governments.

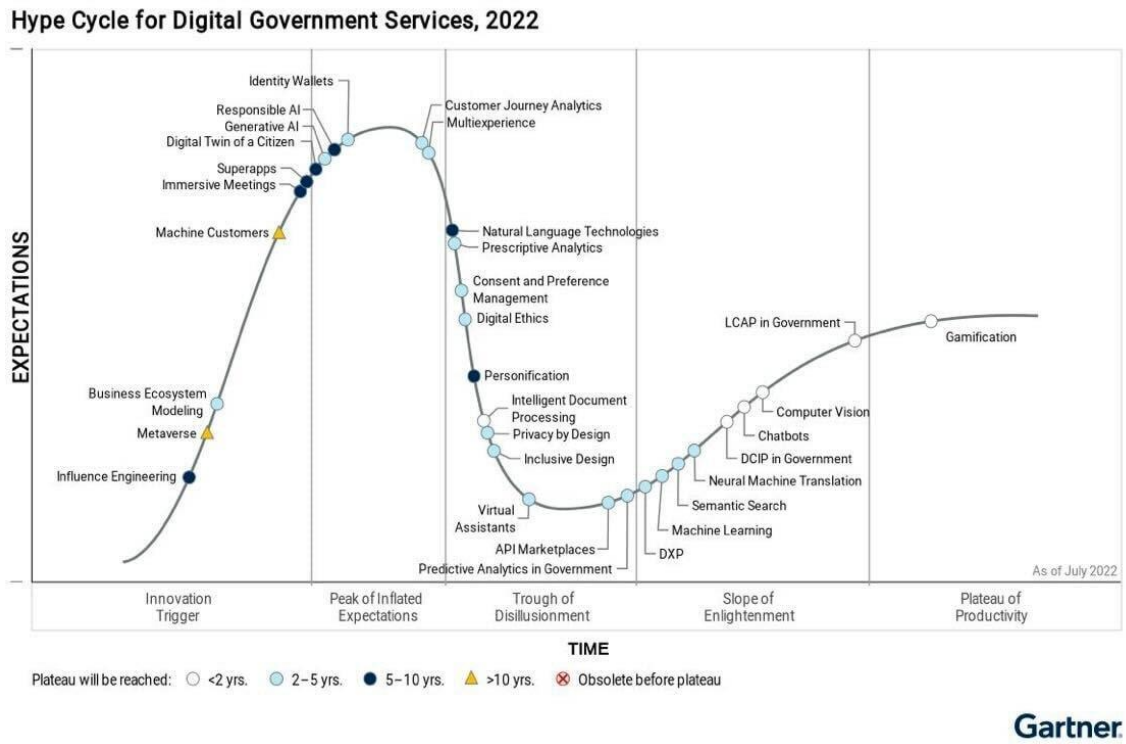

Figure 5 Gartner's Hype Cycle for Digital Government Services, 2022

4.6 AI methods for building applications for the Public Sector

There are various types of AI methods that could be used for building applications for the public sector. These methods can be classified in four categories, namely AI foundation, Machine Vision, Language and speech, and other.

- AI Foundation. The core AI methods used in the public sector. These include:
    - *Machine Learning (ML)*, which enables learning from data, either supervised or unsupervised. For example, ML can be used in the public sector to predicting traffic congestion based on sensor data,
    - *Neural Networks and Deep Learning*, that emulate the human brain, hence, enabling AI models to learn from data like humans. For example, deep learning can be used in the public sector to fast and easy detect cyber threats.
- Machine Vision. Includes methods that aim to derive meaningful information from digital images or videos. These include:
    - *Computer Vision*. It provides systems the ability to "see" and behave similar to how humans see and behave. It is based on gaining high-level understanding from digital

images or videos. For example, *Image Recognition* recognizes and identifies objects, places, people, writing, actions, and patterns in digital images. In the public sector, one of the most widespread applications of computer vision is the detection of license plate numbers in live footage.

- *Simulation*. Uses AI models and imitate the operation of real-world events in order to predict an outcome. Simulation is employed in many cases of the public sector like civil defense modeling, maintenance scheduling, population modeling, and policy simulation and outcome analysis.
- *Biometrics*. Includes the analysis of human physical and emotional characteristics, e.g., sentiment, finger, retina, body heat, sound of voice. Voice identification, for example, can be useful for public services as it can be used to save time to governmental call centers.

- Language and Speech
  - *Natural Language Processing (NLP)*. It focuses on the interaction between computers and human language. It is "the machine's ability to identify, process, understand and/or generate information in written and spoken human communications" [46]. In the public sector, NLP is used for digital translations, identifying handwriting differences and similarities, transcribing, and others.
  - *Text Mining or text analytics*. Uses NLP to analyse texts and transform them to structured data that can be read and analysed by other AI or computer systems. Text mining can be used in the public sector, for example, to classify citizens' complaints to administrative agencies.
  - *Machine Translation*. It is based on NLP and refers to the automatic translation of vast volumes of text from one natural language to another.
- Other
  - *Robotics.*
  - *Intelligent Automation*. The integration of robotics with multiple components from various emerging technologies. It is used to automate low-level tasks.
  - *Virtual Agents.* Computer-generated virtual personas that can be used to interact with citizens, businesses and other users. Virtual agents are used in the public sector as assistants and are deployed to primarily answer citizens' questions (e.g., using chatbots).

4.7 Regulations and Strategies for using AI in the Public Sector

Regulation refers to the diverse set of instruments by which governments set requirements on enterprises and citizens. Regulation include all laws, formal and informal orders, subordinate rules, administrative formalities and rules issued by nongovernmental or self-regulatory bodies to whom governments have delegated regulatory powers [39]

One may argue that the public sector is now lagging behind the commercial sector in terms of AI adoption. Governments, on the other hand, are attempting to catch up and close the gap [7]. A number of related legislative initiatives have been already issued. For example, the European Commission recently presented its first AI legislative framework. The European Parliament has already adopted a resolution [21] with recommendations to the European Commission. Prior to that, EU legislators passed the 2017 Resolution [20] and the "Report on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things, and Robotics" [15], which will be published with the European Commission's "White Paper on Artificial Intelligence" [16] in 2020. In addition, the European Commission's proposal for regulation [17] develops a set of guidelines for AI system market entrance. It forbids some AI activities, creates standards and duties for high-risk AI systems, and mandates transparency and market monitoring norms. After the adoption of the Commission's proposal by the EU Parliament and member states, the new legal framework will be directly applicable throughout the European Union [26].

Given the initial successes with deploying AI and the massive investments being made in AI by national governments, individual countries have also developed national-level AI strategic plans [20]. For the majority of these governments modernizing the public sector through the deployment of AI is a key national priority. Since 2019, 16 European countries have set national agendas/strategies for adopting AI in the public sector, 6 (Ireland, Latvia, Italy, Lithuania, Portugal, UK) have run pilot activities and trials, three (Austria, Sweden, UK) published guidelines, two (Finland, Germany) had a widespread adoption and only one (UK) ran an early project [7]. Three years later, according to OECD's [4] initial mapping on AI, 50 countries (including EU) introduced national AI strategies, with 36 of them having specific strategies for AI in the public sector. The majority of these strategies are based on shared principles like economic growth, ethics and trust, security, and strengthening the talent pipeline. Although there isn't a single piece of legislation that governs the use of AI in the public sector, numerous governments have launched a number of national AI programs in recent years.

For example, Australia Securities & Investments Commission's recent 2022-2026 Corporate plan [4] states that *"In the next 12 months, we will focus on increasing our use of automation and expanding our use of advanced analytics, including artificial intelligence and machine learning"*. The organization intends to *"use data analytics, artificial intelligence and machine-learning technologies to more quickly and accurately identify harms in our environment"*.

Moreover, New Zealand's Ministry of Business, Innovation, and Employment authored a business plan to support the use of AI. The plan is entitled "Building a Digital Nation and the Strategic Science Investment Fund 2017–24 Business Plan" [34] and its main vision is to "accelerate the safe adoption of AI technologies". The plan aims to form a partnership between the New Zealand government, Callaghan Innovation, business, and the nascent AI forum to do research on finding potential and minimizing hazards associated with the use of AI.

Furthermore, the G20 adopted the "G20 AI Principles" [23], which are drawn directly from the OECD's *Recommendation of the Council on Artificial Intelligence* [40], while in 2018, all EU member countries signed the "Declaration of Cooperation on Artificial Intelligence" [11] and committed to collaborating to increase European AI capabilities and adoption, solve socioeconomic issues and ethics, and create a suitable legal and ethical framework. They also agreed to make AI available to and helpful to public administrations, to share best practices in procurement and using AI in government, and to promote open data standards. The subsequent EU Coordinated Plan on Artificial Intelligence [12] builds upon the declaration and seeks to *"maximise the impact of investments at EU and national levels, encourage synergies and cooperation across the EU."* Ten governments (Denmark, Estonia, Finland, the Faroe Islands, Iceland, Latvia, Lithuania, Norway, Sweden and the Åland Islands) also signed the "Declaration on Artificial Intelligence in the Nordic-Baltic Region"[5], pledging to, among others, improve skills development and access to data, and to develop ethical guidelines.

Table 3 presents regulation initiatives as strategies, action plans and declarations worldwide for adopting AI in the public sector. More than 50 countries worldwide have already created at least one initiative for adopting AI in the public sector.

Region	*Title*	*Type*	*Year*
Global	G20 AI Principles	Statement	2019
Global	Recommendation of the Council on Artificial Intelligence (OECD)	Recommendation	2019
EU	AI Legislative Package (AI Act)	Legislation	2021
EU	Declaration of Cooperation on Artificial Intelligence	Declaration	2018
EU	Coordinated Plan on Artificial Intelligence	Plan	2018

[5] https://www.norden.org/sv/node/5059

Nordic-Baltic Region	Declaration on Artificial Intelligence in the Nordic-Baltic Region	Declaration	2018
Australia	Securities & Investments Commission's recent 2022-2026 Corporate plan	Plan	2022
Austria	Artificial Intelligence Mission Austria 2030	Strategy	2021
Argentina	AI National Plan	Strategy	2019
Belgium	Action Plan AI (Flemish)	Plan	2019
Belgium	AI4Belgium	Strategy	2019
Brazil	Brazilian AI Strategy - EBIA	Strategy	2021
Bulgaria	Concept for the Development of Artificial Intelligence in Bulgaria until 2030	Concept	2020
Canada	Pan-Canadian AI Strategy	Strategy	2022
Chile	AI National Policy	Policy	2019
China	National New Generation Plan	Plan	2017
Colombia	AI National Strategy	Strategy	2019
Croatia	National Plan for the Development of AI (draft)	Plan	2020
Cyprus	National AI Strategy: Key Action for Promoting the Integration and Development of AI in Cyprus	Strategy	201+
Czech Republic	National AI Strategy of the Czech Republic	Strategy	2019
Denmark	National Strategy for AI	Strategy	2019
Estonia	National AI Strategy	Strategy	2019
Finland	AI 4.0	Strategy	2020
France	National Strategy on AI	Strategy	2018
Germany	National AI Strategy	Strategy	2018
Greece	Digital Transformation Bible	Policy	2020
Hungary	Hungary's AI Strategy	Strategy	2020
Iceland	AI Strategy (forthcoming)	Strategy	2021
Ireland	AI - Here for Good. A National Artificial Intelligence Strategy for Ireland	Strategy	2021

India	National Strategy on Artificial Intelligence	Strategy	2018
Italy	AI Strategic Programme	Strategy	2021
Japan	AI Strategy	Strategy	2019
Korea	National Strategy for AI	Strategy	2020
Latvia	Latvia's National AI Strategy	Strategy	2020
Lithuania	Lithuanian Artificial Intelligence Strategy	Strategy	2019
Luxemburg	AI: A Strategic Vision for Luxemburg	Strategy	2019
Malta	Malta: The Ultimate AI Launchpad	Strategy	2019
Mauritius	Mauritius Artificial Intelligence Strategy	Strategy	2018
Mexico	Mexican National AI Agenda	Agenda	2018
Morocco	Strategy for Developing an AI Ecosystem	Strategy	2021
Netherlands	Strategic Action Plan for AI	Action Plan	2019
New Zealand	Building a Digital Nation and the Strategic Science Investment Fund 2017–24 Business Plan	Plan	2017
Norway	National Strategy for AI	Strategy	2020
Peru	National AI Strategy	Strategy	2021
Poland	Policy for AI Development in Poland	Policy	2020
Portugal	AI Portugal 2030 - National Strategy for AI	Strategy	2019
Romania	Romanian's AI Strategy	Strategy	2019
Saudi Arabia	National Data and AI Strategy	Strategy	2020
Serbia	Strategy for the Development of AI in the Republic of Serbia for the Period 2020-2025	Strategy	2020
Singapore	National AI Strategy	Strategy	2019
Slovenia	Slovenia's National Programme on AI	Strategy	2021
Spain	National AI Strategy	Strategy	2020
Sweden	National Approach to AI	Strategy	2018
Switzerland	Digital Switzerland Strategy	Strategy	2020

Thailand	Thailand's National AI Strategy and Action Plan (draft)	Strategy	2022
Turkey	National AI Strategy	Strategy	2021
Ukraine	National Strategy for the Development of AI in Ukraine	Strategy	2021
United Arabic Emirates	UAE National Strategy for AI	Strategy	2021
United Kingdom	National AI Strategy	Strategy	2021
United Kingdom	AI Barometer	Report	2019
USA	FUTURE of AI Act of 2017	Proposed law	2017
USA	AI in Government Act	Bill	2018
USA	National Security Commission Artificial Intelligence Act of 2018	Proposed law	2018
Uruguay	AI Strategy for the Digital Government	Strategy	2019
Viet Nam	National Strategy of R&D and Application of AI	Strategy	2021

Table 3 Legislation Initiatives for adopting AI in the Public Sector

In general, all the proposed AI legislation initiatives for the public sector are focused on a number of key. These include:

- experimentation with AI in government and the identification of specific AI projects currently underway or that will be developed in the near future [4]:
- collaboration across sectors, such as through public-private partnerships and facilitated through innovation hubs and labs [4]:
- fostering of cross-government councils, networks and communities to promote systems approaches [4]:
- automation of routine government processes to enhance efficiency [4]:
- use of AI to help guide governmental decision-making (e.g. in policy evaluation, emergency management and public safety) [4]:
- strategic management, leverage and opening up of government data to develop tailored and anticipatory services, as well as to fuel AI in the private sector [4]:

- provision of guidance on the transparent and ethical use of public sector AI [4]:
- enhancement of civil service capacity through training, recruitment, tools and funding. [4]:
- Member States recognise the fact that they need to invest in capacity building of civil servants and public sector officials [7]. Some national strategies explicitly address "up-skilling" as an issue. Access to more and better data is also often mentioned as a key element in order to improve the quality of public services. The national strategies contain different approaches to data governance. Some national strategies explicitly mention open data and sharing data transversally as well as with the private sector. Most Member States stress the need to embed AI design, development and deployment firmly within an ethical framework. Values and principles frequently mentioned in this context are human centred, trustworthy and responsible AI, transparency and human oversight. While most Member States mention an ethical framework, some also specifically mention the need to regulate AI and see the public sector in the regulatory role. Currently, agencies in the public sector use automated decision making mostly in the category of assisted or conditional automation. In few cases, complete processes or services are automated. There are no fully autonomous systems in use in the public sector.
- Public sector functions discussed in these plans cover the gamut from national security to immigration, smart cities, energy, and the environment, and transportation, among others [20].
- With the adoption of the 2030 Agenda for Sustainable Development, nations worldwide committed to a set of universal, integrated and transformational goals and targets, known as the Sustainable Development Goals (SDGs). The 17 goals and 169 targets represent a collective responsibility and a shared vision for the world. Governments are working to make progress to reach them by 2030, and many are exploring the potential of AI to help achieve this objective. In the public sector, the OECD has found that governments are pursuing uptake of AI geared towards preserving the environment, natural capital and climate resilience [47]. These aims support a number of SDGs, such as: Clean Water and Sanitation (SDG 6), Affordable and Clean Energy (SDG 7), Responsible Consumption and Production (SDG 12), Climate Action (SDG 13), Life Below Water (SDG 14), and Life on Land (SDG 15). In its report Harnessing Artificial Intelligence for the Earth [53], the World Economic Forum (WEF) explores the ways that AI can help address environmental challenges. Governments are also broadly adopting, or planning to adopt, AI projects that support citizen-facing welfare services and better lives for individuals (e.g., to help civil servants make decisions about whether citizens and businesses receive financial and other assistance from the government) . These aims cut across the SDGs for No Poverty (SDG 1), Zero Hunger (SDG 2), Good Health and Well-Being (SDG 3), Gender Equality (SDG 6) and Reduce Inequalities (SDG 10).

Despite overall similarities between the individual focus areas of the European national strategies, there are variations in their approaches to developing AI solutions for society. These differences in approach can be divided into three main clusters of countries [32].

1. Nurture economic growth by strengthening the ecosystem. One cluster of countries (Italy, Portugal, Switzerland, Spain) focuses their AI efforts on developing a strong R&D foundation to ensure future progress within AI. They outline specific initiatives to foster a dynamic ecosystem centered around start-ups, private companies, and the Public Sector. Through this dynamic innovative ecosystem, they intend to expand knowledge and AI solutions, thereby becoming attractive countries for talent and developing the right competencies. This is expected to spill over into the Public Sector.
2. Driving Public Sector innovation by re-inventing the delivery of public services. A cluster of comparable countries (Netherlands, Denmark, Finland, Iceland) are focusing their AI efforts on outlining and initiating specific initiatives to foster innovation in the Public Sector through policies and pilot projects. Policy actions ensure ethical and sustainable use of new AI solutions in the Public Sector, and lay the foundation for more efficient public services that benefit all of society. Pilot projects and Proofs of Concept are being launched in the Public Sector in order to experiment with AI, and provide the foundation for implementing new solutions as well as the future development of public services based on AI.
3. Strategic vision ensuring sustainable AI development for all of society. The final cluster of countries (Sweden, Belgium, Austria, Ireland) has developed wide-ranging approaches to AI that provide strategic direction for the development of society as a whole, and through consensus ensures long-lasting support and sustainable adoption. By setting a broad yet clear path for the overall direction of AI, these countries are focusing on unified development of AI, centering around key aspects which can be formulated as specific initiatives and approaches for various areas and sectors.

4.8 Recommendations

AI Watch [46] recently proposed a set of recommendations that are grouped in four intervention areas namely (i) Promote an EU value-oriented, inclusive, human-centric and trustworthy AI in the public sector, (ii) Enhance coordinated governance, convergence of regulations and capacity building, (iii) Build a shared and interactive AI digital ecosystem, (iv) Apply and monitor sustainability though value-oriented AI impact assessment co-created frameworks. The recommendations address policymakers and stakeholders at different operational and governance levels, which may need to be adapted to match countries' specific political, administrative, territorial, economic social, cultural and organisational dimensions.

The recommendations and related actions provided by this handbook highlight specific requirements to be put in place over the coming years, and common issues to be addressed at early stages of development of AI-based solutions by the public sector

Intervention Area 1: Promote an EU value-oriented, inclusive, human-centric and trustworthy AI in the public sector.

- Recommendation 1.1. Harmonise and complement EU regulations to promote human-centric and trustworthy AI-enabled public services for all citizens.
    - Actions
- Recommendation 1.2. Promote the adoption of ethical principles, the development of guidelines and the identification of mitigating measures to minimise the risks of deployment of AI by the public sector.
    - Actions
- Recommendation 1.3. Develop and promote dedicated AI-enabled solutions based on co-creation approaches to increase relevance of services and trust in citizens' and businesses' to stimulate confidence in the use of AI by the public sector.
    - Actions

Intervention Area 2: Enhance coordinated governance, convergence of regulations and capacity building.

- Recommendation 2.1. Create an EU-wide network of governance bodies for a streamlined management of AI in the public sector.
    - Actions
- Recommendation 2.2. Design national and European capacity-building programmes for public sector innovators aiming to develop and/or adopt AI in support of the digital transformation of public services.

- Actions

- Recommendation 2.3. Build upon and promote the use of regulatory sandboxes for public administrations, allowing experimentation of AI-enabled solutions in controlled environments.
    - Actions
- Recommendation 2.4. Optimise funding in support of AI in the public sector to promote the spreading and scaling of reusable solutions.
    - Actions
- Recommendation 2.5. Promote the development of multilingual guidelines, criteria and tools for public procurement of AI solutions in the public sector throughout Europe.
    - Actions

Intervention Area 3: Build a shared and interactive AI digital ecosystem.

- Recommendation 3.1. Support multidisciplinary research and knowledge creation amongst European universities and Research and Development (R&D) institutions around AI for the public sector.
    - Actions
- Recommendation 3.2. Build a common European Data Space for public sector bodies and their operators, drawing from the compilation of relevant AI datasets and related registries throughout Europe.
    - Actions
- Recommendation 3.3. Reinforce and advance existing initiatives on open data and interoperability.
    - Actions
- Recommendation 3.4. Share reusable and interoperable AI components at all operational levels of European public administrations.
    - Actions
- Recommendation 3.5. Create a European marketplace for GovTech solutions in support of public sector digital transformation.
    - Actions

Intervention Area 4: Apply and monitor sustainability through value-oriented AI impact assessment co-created frameworks.

- Recommendation 4.1. Set up an EU observatory on AI, built on a pan-European network of AI national observatories, to gather, share, and collectively manage best practices and experiences learned from different stakeholders in the public sector throughout Europe.
    - Actions
- Recommendation 4.2. Develop and apply umbrella impact assessment frameworks based on key influencing factors to measure the use and impact of AI in the public sector.
    - Actions
- Recommendation 4.3. Promote AI in the public sector in support of sustainability while developing sustainable AI in compliance with environmental principles and leveraging on civic engagement and participation.
    - Actions

OECD also provides five recommendations [39] to policy makers pertaining to national policies and international co-operation for trustworthy AI. Although these recommendations extend beyond public sector innovation and transformation, this primer seeks to help policy makers explore potential avenues for implementation in the public sector [4].

- Investing in AI research and development. This primer provides information and real-world examples that show how governments are investing to support the use of AI in the public sector with a view to transforming and innovating the way they design and implement policies and services. This includes investment in managing data to serve as fuel for AI, as well as opening up data for reuse in order to catalyse innovation.
- Fostering a digital ecosystem for AI. This primer details how governments are building connection points, knowledge-sharing mechanisms and ecosystems within the public sector, as well as with partners in industry and civil society, in order to support AI experimentation in government.
- Shaping an enabling policy environment for AI. This primer and its corresponding webpage “AI Strategies & Public Sector Components” (https://oe.cd/aistrategies) document how countries are developing strategies for AI, and the extent to which they specifically provide for public sector transformation. It also proposes a potential framework for evaluating a given policy environment in order to align activities to meet AI goals. Over time, OPSI plans to integrate this resource with the OECD AI Policy Observatory’s upcoming knowledge repository of national AI strategies to allow users to obtain full, regularly updated information on strategies.
- Building human capacity and preparing for labour market transformation. This primer discusses how governments are building AI capacity among civil servants. It also explores the ways in which governments are securing capacity from external partners and providers, as

well as how they are preparing for the potential for fundamental changes in the future driven by AI, including changes to the labour market. In addition, the corresponding webpage "AI Tools & Resources" (https://oe.cd/airesources) provides a categorised repository of hands-on tools and resources (e.g., online courses) that can help government officials learn more about AI and its potential roles in the public sector.

- Establishing international co-operation for trustworthy AI. This primer provides an overview of ongoing international collaboration opportunities, including those convened by the OECD, as well as information on the products of such collaborations, such as agreed-upon principles for trustworthy AI.

An AI strategy should include the following [5]:

- Baselines: an assessment of the organisation's current strategic situation and challenges that AI might help address.
- Objectives: what the organisation wants to achieve using AI and the principles that will underpin the actions it takes to achieve them.
- Approaches: the concrete actions that will be undertaken to achieve these objectives.

A framework for an AI strategy

Baseline	Objectives	Approaches
Determine current strengths and weaknesses by mapping: - Internal AI capabilities. - Government and external data assets. - Existing government AI and data science projects. *Assess the strategic context*: - Public and workforce attitudes to government and AI . - Current legal and regulatory framework. - Existing government and international commitments, institutions. - Academic and private sector expertise that might be drawn upon. *Identify specific operational problems that AI has the potential to help solve*: - Adopt a multi-disciplinary and diverse approach to decide whether AI is the best solution to a policy problem, and if so, how it should be designed and implemented. - Put in place processes to understand user needs. - Create mechanisms to match resources to priority problems. - Define the specific decision AI will make or support. - Consider who will be impacted by this decision and associated risks if it fails. - Explore how the service will need to be redesigned to leverage the impact of AI.	*Decide what goals the AI should help government achieve*: - Articulate how AI will generate public value and specify missions to which AI can be part of the solution. - Engage stakeholders in goal definition. - Make space for experimentation and learning. *Define and communicate to stakeholders the principles that will shape how AI is used in government*: - Fairness and unbiasedness. - Transparency and accountability. - Privacy and individual autonomy.	*Ensure government access to AI capability and capacity*: - Construct talent pipelines, and develop recruitment and retention plans for internal technical expertise. - Harness external expertise through partnerships and collaboration. - Design effective public sector AI procurement processes. - Build a cadre of civil servants who understand the legal, ethical, technical and managerial issues around AI. - Establish funding schemes and secure the availability of resources in the government's fiscal plans. *Secure ethical access to, and use of, quality data and infrastructure*: - Determine what data are needed to address the problems. - Decide how to obtain input data of sufficient quality and that are sufficiently representative of the target population to make accurate predictions with minimal bias. - Develop a data strategy that complies with data protection law and best practice and is consistent with AI principles. - Ensure important data infrastructure are in place (e.g., hybrid cloud services). *Put in place legal, ethical and technical frameworks to operationalise the principles*: - Monitor compliance with principles during implementation to track progress, and identify and respond to emerging issues. - Put in place safeguards against bias and unfairness. - Clarify the appropriate role for humans in the decision-making process. Develop open and transparent accountability structures. *Prepare for future shifts:* - Ensure openness and flexibility in future plans and contracts. - Follow OPSI Anticipatory Innovation Governance concepts.

Figure 6 A framework for an AI strategy for governments

4.9 Cases

AI can be used in various cases in the public sector and in different domains including healthcare, transportation, security, education, public administration, and military. Countries establish AI efforts in response to their most pressing requirements, but it is also advised that the approach to AI be part of the planning and accounting for future digital initiatives, with a whole-of-government approach to infrastructure, standards, governance, and execution [40].

AI cases can be found in many resources online. For example, OECD's Observatory of Public Sector Innovation (OPSI) encourages public servants to keep up with the latest AI developments by accessing the following resources:

- OPSI's Case Study Platform[6] collects and shares hundreds of government innovations to help disseminate good ideas. Any public sector innovator may submit innovations to the platform.
- OECD AI Policy Observatory (AIPO)[7] that aims to help countries encourage, nurture and monitor the responsible development of trustworthy AI. It includes an open and comprehensive database of AI policy initiatives, information on AI public policy topics (e.g. jobs, skills, data, health, transport), and AI metrics and measurement tools (e.g. OECD methods and measurement, live data points from partners).
- The UN International Telecommunications Union (ITU) has developed a Global AI Repository[8] of projects that promote progress towards the SDGs.
- The OECD Digital Government Toolkit[9] provides resources on good digital government practices by country, including many on managing data as an asset.

The pattern of government adoption typically follows this typology of use cases [51]:

- Citizen engagement [7] [51]. The introduction of AI tools such as chatbots that answer citizen queries. For example: Where is my ballot? Where is the nearest emergency department? How can I apply for social welfare benefits? Additionally, aggregation and pattern determination can be used to collect feedback from millions of citizens on a draft policy published online.
- Compliance and risk management [51]. AI systems are used to cross-reference and reconcile terabytes of data from multiple sources to create alerts for noncompliance. For example, financial intelligence units and central banks use AI to track illicit fund flow and beneficial

[6] https://oecd-opsi.org/case_type/opsi/?_innovation_tags=artificial-intelligence-ai
[7] https://oecd.ai/en/
[8] https://www.itu.int/en/ITU-T/AI/Pages/ai-repository.aspx
[9] https://www.oecd.org/governance/digital-government/toolkit/goodpractices/

ownership as well as terrorism financing to comply with the Financial Action Task Force. Tax authorities can use AI to track tax filers who use duplicate profiles to avoid taxation.

- Fraud detection, prevention, and investigation [7] [51]. Closely related to compliance AI can be used to detect and prevent fraud for example by procurement agencies, anti-corruption units, or audit agencies.
- Business process automation [51]. AI automation tools can scan websites to get currency exchange rates and present information.
- Personalized service delivery [51]. Based on a profile, AI sends automatic alerts such as when to renew a driving license.
- Asset management [51]. AI can be used to tracking asset movements across multitudes of systems, aggregating data from the Internet of Things devices.
- Analytics and decision-making [51]. AI or machine learning helps aggregate and cross-reference data such as household survey data with information on school enrollment, address changes, satellite images of floods, mosquito swamps, and pandemics to produce policy insights and identify areas needing greatest attention for targeted policy actions.
- Cybersecurity [7].
- Social Services [7]. Support and/or drive decisions in fields such as law enforcement, crime prevention, public safety, children welfare, social programs.
- HR [7]. Takes on key HR tasks including hiring, retaining talent, training, benefits and employee satisfaction.

Table 4 presents a large number of cases that apply AI in the public sector. For each case we present the objective of the initiative, the AI technology it uses, the region, the type, a brief description, and the measured impact.

Objective	*AI technology*	*Region*	*Type*	*Description*	*Impact*
AI to improve MOT testing[10]	Clustering	the UK	Analytics & Decision making	The UK's Driver and Vehicle Standards Agency (DVSA) developed an approach that analyses vast amount of testing data, which it then combines	• 50% decrease in examiners' preparation time for

[10] https://www.gov.uk/government/case-studies/how-the-department-for-transport-used-ai-to-improve-mot-testing

				with day-to-day operations to develop a continually evolving risk score for garages and their testers. From this the DVSA is able to direct its enforcement officers' attention to garages or MOT testers who may be either underperforming or committing fraud.	enforcement visits. • Increase in disciplinary action against garages, meaning standards are now being better enforced. As more garages are delivering better MOT standards, there are more cars on the road that comply with roadworthiness and environmental requirements.
Help developing countries better understand their population distribution[11]	Computer Vision	N/A	Analytics & Decision making	The Department for International Development partnered with the University of Southampton, Columbia University and the United Nations Population Fund to apply a random forest machine learning algorithm to satellite image and micro-census data. The algorithm then used this information to predict the population density of an area. The model also used data from micro-censuses to validate its	DFID has deployed the programme in Nigeria, Zambia, Mozambique and Democratic Republic of Congo. DFID has also conducted scoping missions to Tanzania, Ethiopia, and South Sudan. The model - called GRID3 - is being used to: develop a

11 https://www.gov.uk/government/case-studies/how-dfid-used-satellite-images-to-estimate-populations

				outputs and provide valuable training data for the model.	hybrid census model which combines population estimates for small areas of uniform, detailed grids with modelled population estimates support developing countries as they plan their full censuses plan vaccination campaigns and other services support developing countries such as DRC safely gather population estimates for areas in conflict
Make GOV.UK more accessible[12]	Classification, natural language processing (NLP), deep learning	gov.uk		GDS had designed a taxonomy to organise information locally but the process of tagging content was resource-intensive. It needed engagement with publishers in other government departments, which was time-consuming. GDS wanted to find a way to do this activity in a more time-efficient and cost-effective way. GDS had 100,000 untagged pages, which they needed tag to around 210 sub-branches. The GDS data science team	The final model, a deep learning convolutional neural network, was able to provide tags to 96% of existing content and suggest tags to new content with high accuracy. GDS predicted the original task might last years, but with machine learning this reduced to

12 https://www.gov.uk/government/case-studies/how-gds-used-machine-learning-to-make-govuk-more-accessible

				embedded staff in the existing team responsible for tagging content, working with them to build a supervised machine learning model to solve this problem. The GDS team trained the model on the pages they had already tagged to recognise patterns in the page contents. They used NLP to make the text content on the page machine readable. They then used these results alongside the page metadata (such as date published and department) to learn patterns. These patterns could predict where the untagged pages would best fit in the sub-branches.	under 6 months and was a relatively quick and easy solution for GOV.UK and publishers across government.
Land Registry documentation[13]	NLP	Sweden	Citizen engagement	The SLR's case handlers receive many requests about properties and property rights from citizens. To make decisions, case handlers need to review historical information for that property, often dating back to the 1850s. Case handlers spent approximately 48,000 hours a year manually translating and evaluating the handwritten documents. The government charges citizens for the number of hours case handlers spend working on their request. The handwritten documents were of low quality and resolution which made analysis challenging. To fix	The model now allows case handlers to rapidly respond to citizen requests and focus more on urgent decisions.

13 https://www.gov.uk/government/case-studies/natural-language-processing-for-land-registry-documentation-in-sweden

				this, the SLR carried out pre-processing work on the handwritten documents to improve the quality of the input data. The SLR then used handwritten text recognition (HTR) to extract information from the handwritten documents. The SLR then used a neural network to apply word corrections and associations to complete any sentences with words not captured by HTR. Once SLR had extracted the text, they used an AI model to highlight key features in the document, for example location, names and summary.	
Using data from electricity meters to predict energy consumption [14]	Clustering	the UK	Analytics and decision-making	A research institution needed to understand which electric appliances were being used in a house at a certain time to optimise heating and energy consumption. The research institution did not know when particular electric appliances were being used. This meant they were unable to optimise heating and energy consumption resulting in higher prices and energy waste. The research institution used non-intrusive load monitoring to gather unlabelled data from electricity meters to see which appliances were being used and when. They	The model was able to predict future energy needs of a property; could help plan when households might use appliances; enabled smart use of heating - for example turning off heating while the occupier is out and turned on when they are coming home

[14] https://www.gov.uk/government/case-studies/using-data-from-electricity-meters-to-predict-energy-consumption

				used unsupervised machine learning techniques to convert the unlabelled data into patterns. From this, the research institution could cluster the different types of appliances based on their power consumption patterns.	
Use AI to compare prison reports[15]	NLP	UK	Social services	Ministry of Justice (MoJ) needed to compare how various factors including geography, and incidents such as inmate conflict, affected different prisons. MoJ had over 250,000 sentences of unstructured text in over 500 reports detailing their inspections of prisons and other institutions. These reports were from: HM Inspectorate of Prisons; the Independent Monitoring Board; the Prisons and Probation Ombudsman; Ofsted reports into secure training centres. There were too many reports for staff to quickly access relevant information. MoJ trained a neural network on the prison reports to track how people use specific words in prison contexts. The algorithm groups words with similar meanings to form an 'intelligent search' tool. New reports are automatically added to the tool's library so the data remains up-to-date. This means staff can rapidly	The tool helps MoJ: identify patterns of issues and incidents across prisons; identify geographic patterns affecting prisons; inform data-driven decisions about prison inspections and policy.

[15] https://www.gov.uk/government/case-studies/how-the-ministry-of-justice-used-ai-to-compare-prison-reports--2

				uncover information buried in the reports and identify trends.	
An AI-driven helpdesk to handle citizen calls [2]	N/A	Italy	Citizen engagement	The system can deal with greater volume than human operators, meaning citizens get faster services and the MEF is more productive.	Since the AI helpdesk was deployed, customer satisfaction has risen by 85 percent.
Answer queries from the public [2]		Singapore	Citizen engagement		
Process incoming correspondence [2]		UK	Citizen engagement		
Process thousands of public comments on regulatory proposals [2]		US	Citizen engagement		
Transform the way of managing trees across the city [2]	AI filtering and processing	Washington DC	Analytics & decision-making	Use an AI filtering and processing system to: (i) Rapidly, and accurately, count the number of trees, their height and crown width, (ii) Organize and filter the collected data points to create a detailed citywide map, (iii) use the map to drive decision-making and resource allocation.	In this case, as well as rapidly providing the city with new, rich information, the administration also found a way to pinpoint illegal felling activity, amounting to $100,000 worth of unissued fines.
Randomize boat patrol routes [43]		NY, LA	Fraud detection, prevention, and investigation	The Coast Guard uses AI to randomize its boat patrol routes, making their day-to-day security-related activities less predictable for criminals.	

Classify and understand citizen petitions and then route them to the relevant department [26]	Classification	Mexico			
Free up operators' lines and customer help desks[16]	Chatbot	North Carolina	Citizen engagement	AI-based chatbots deployed in a North Carolina government office free up operators' lines and customer help desks. Most service questions from the government office are simple and repetitive (for example, almost 90 percent of requests are about resetting passwords).	Using the chatbots to answer the simpler questions has allowed customer agents to focus on more complex and time-sensitive service inquiries
Efficiently responds to service inquiries.[17]		Surrey, British Columbia	Citizen engagement	The MySurrey app—an app deployed in Surrey, British Columbia, that deployed IBM Watson—efficiently responds to service inquiries. The app can handle 65 percent of inquiries, for which there is already self-help information on city websites. Watson, which uses machine learning to learn over time, reviewed over 3,000 documents related to sixteen city services and responds to 10,000 service inquiries	

[16] Goldsmith, S. (2017). Artificial Intelligence Will Help Create a More Responsive Government. Government Technology. https://www.govtech.com/opinion/Artificial-Intelligence-Will-Help-Create-a-More-Responsive-Government.html .

[17] Pereira, D. (2017, March 27). Watson helps cities help citizens. Medium. https://medium.com/@darylp/watson-assists-cities-with-311-3d7d6898d132 .

Respond to service inquiries from citizens [26]	AI-based Natural Language Generation	Japan	Citizen engagement	Using AI-based Natural Language Generation (NLG), agencies can draft documents such as self-help information and service application files. The technique can also support non-data science staff in more easily understanding and using the data. Japan's Ministry for Economy, Trade, and Industry deployed an innovative solution to help parliament member offices respond to service inquiries from citizens by drafting responses using AI	
Assistance in completing forms for unemployment benefits [36].	Chatbot	Australia	Citizen engagement	An AI-based chatbot has been deployed for providing targeted assistance with unemployment benefits. When used, the form can be auto-filled based on the applicant's profiles.	Reducing the number of questions that need applicant's input from 150 to 10-15.
Partially automate cargo risk assessments [25]	NLP	Canada	Compliance & risk assessmnet	The agency was therefore able to use their AI to partially automate accurate risk assessments. They successfully designed a dashboard and an initial version of a targeting interface that help employees with figuring out which cargo is potentially high-risk. There are initial planning and deployments underway about improving the approach, start implementation and then adapting it to other modes of transportation and customs and the border related activities	
CitizenLab. Crowdsource public decision-	NLP, Classification, clustering	Belgium	Analytics & decisio	Classifying contributions from citizens, identify emerging ideas, highlights major trends, and clusters ideas by theme,	

making[18] [38] [5].			n making	demographic, and geographic location. For example, the government may realize that a neighborhood in a geographic location is prioritizing fixing issues with roads or installing more traffic stops. During the deployment phase, the agency learned that an effective human-machine interaction is essential for success of the project. The agency's employees needed to learn how to interpret the output generated by the system, trust it, and use it in day-to-day workflow. Moreover, the agency realized that the quality of input data (citizens' ideas and feedback) is crucial to reliably understand citizens' needs, showing the importance of providing guidance on submitting their contributions. Ongoing evaluation of the system through auditing and refinement of the platform has been conducted by the agency.	
Governance Risk Assessment System. Optimize the process of detecting fraud in public expenditure [51]		Brazil	Compli ance & risk assess ment	The system optimizes the process of detecting fraud in public expenditure substantially, saving valuable resources – time and money – and increasing the effectiveness of audits and investigations.	

[18] www.citizenlab.co

Zero Trust: Monitor, evaluate, and scrutinize the work and lives of public servants [51]	NLP, anomaly detection	China	Fraud detection, prevention, and investigation	China has faced enormous challenges of controlling corruption and has 50 million employees on the government payroll. Zero Trust can cross-reference more than 150 databases in central and local government systems. The system detects an individual's property transfers, infrastructure, construction, land purchases, and house demolitions. Zero Trust also detects unusual increases in a civil servant's bank savings, new car purchases, and if an official is bidding for government contracts or is doing so under the name of family members or friends. The system then calculates a probability that those actions are corrupt and alerts officials to highly probable cases of corruption.	Zero Trust was rolled out in 30 counties and cities and identified 8,721 government officials suspected of engaging in embezzlement, abuse of power, misuse of government funds, and nepotism.
DataCrowd. Increase the capacity of the agencies to receive, analyze, and respond to citizen feedback [51]	NLP	Nigeria	Citizen engagement	Citizen engagement and feedback are helpful tools to complement formal mechanisms of accountability as they offer compelling insights for monitoring and evaluation of policies, project designs, and implementation. The AI solution has several features including an AI-powered tag cloud; image classifier; image matching; opinion mining and sentiment analyzer	
Northern Border Surveillance System. Help detect illegal trade,	Convolutional neural network, computer vision,	the US	Fraud detection, prevention,	The US Customs and Border Patrol is one of the world's largest law enforcement organizations and is charged with keeping terrorists and their	

including drug smuggling and human trafficking, and weapons entering the US [51]	pattern matching, anomaly detection, prediction.		and investigation	weapons out of the U.S. while facilitating lawful international travel and trade. There are 300 ports of entry into the United States that need to be secured without disrupting trade and transit. An AI system is used to detect and monitor vessels from miles away and alert authorities when it recognizes unusual vessel movements.	
Contact Tracing and Temperature Detecting Camera Apps [51] Smarter and targeted response on COVID-19 quarantine and social distancing policy to save economies from economic disasters.	Artificial neural network, reinforcement learning, prediction.	the US	Social services	The US Customs and Border Patrol is one of the world's largest law enforcement organizations and is charged with keeping terrorists and their weapons out of the U.S. while facilitating lawful international travel and trade. There are 300 ports of entry into the United States that need to be secured without disrupting trade and transit. Contact tracing and screening to target policy response on quarantine for minimum disruption on economic life and contain the spread of COVID-19.	
Hospital and health information app for doctors and front-line health workers [51]	Natural language processing, data mining, chatbot	Singapore	Personalized service delivery	Doctors and front-line health workers need information on the latest health protocols, staff rosters, operational directives, and dosage to effectively manage the COVID-19 pandemic. Bot MD is an AI Chatbot mobile app that acts like 'google' for hospital and clinical information on COVID19 for doctors and frontline health workers.	More than 13,000 doctors in 52 countries are now using the app.

Review and Approval of Judgement System. Promote consistency in the judicial decisions [51]	Natural language processing, Big Data, data mining, and automation	China	Social services	China's Supreme People's Court is promoting the policy of "Similar Judgments in Similar Cases" to promote consistency in judicial decisions. All prior judgments were digitized and stored in a database. Next, the SPC deployed NLP AI capabilities, through the Similar Cases Push System, to match key text relevant to pending cases using the database. The system presents relevant judgments before a judge using a pre-populated judgment template that the judge reviews and edits. Also, an AI pilot program records court proceedings. Some courts in China are now using AI speech recognition products to directly translate the court hearing recordings into texts in real-time and convert these into written court proceedings using Speech-to-Text NLP methods.	The system reduces the time it takes to formulate a written judgment and all legal procedural documents by 70 percent and 90 percent, respectively.
AI legal assistant [51]	Natural language processing, chatbots.	the UK	Personalized service delivery	An AI legal assistant is necessary for improvements in the analysis of legal contracts; support of private legal bureaucracy among citizens; and guided legal advice.	Automated AI legal assistants and lawyers have surpassed human-level accuracy. In one month, post-launch, DoNoPay.com helped people overturn 160,000 of 250,000 parking tickets-a success rate of 64 percent

Reviewing hundreds of complex and voluminous bidding documents, issued by the federal agencies, to ensure compliance with regulations [51]	Natural language processing, automation	the US	Social services	The U.S. government is harnessing the power of AI to strengthen procurement compliance. The U.S. General Services Administration (GSA) has an Office of Government-wide Policy, which developed a new pilot using AI for scanning bidding documents to determine regulatory compliance. The tool is known as the Solicitation Review Tool (SRT). The SRT AI platform uses NLP, text mining, and machine learning (ML) algorithms to scan and review whether federal solicitations posted on fbo.gov are compliant with Section 508 of the Rehabilitation Act. It alerts responsible parties of noncompliance and flags the need for corrective actions. Through the independent review, the predictions have an accuracy of 95 percent.	
Korea's Fair Trade Commission's Bid Rigging Indicator Analysis System	NLP	Korea	HR	The Fair Trade Commission ensures fair competition in procurement practices in the government. Officials converted a manual process that was in place since 2004 to detect bid-rigging cases using AI. Bid rigging refers to collusion between procurement officials and a pre-ordained vendor to award a contract using corrupt practices. Bid rigging can take various forms, including short bid submission windows, split procurements to capture funds below detectable thresholds,	The introduction of the AI system greatly increased speed and effectiveness.

				significant change orders, and substitution of low priced items with high priced items after the award. Korea's Fair Trade Commission (KFTC) is leveraging an AI and analytics platform, the Bid Rigging Indicator Analysis System (BRIAS) to combat corrupt practices.	
Detecting tax evasion and conducting criminal investigations in cases of tax fraud and identity theft.		the US	Fraud detection, prevention, and investigation	The Internal Revenue Service (IRS) created the Office of Compliance Analytics (OCA) to construct analytics programs that could identify potential refund fraud, detect taxpayer identity theft, and handle noncompliance issues efficiently. OCA leverages an advanced analytics program that relies on the use of Big Data and predictive algorithms to reduce tax fraud.	
The AI Economist. Model and predict tax policy design through data-driven simulations [51]	Artificial neural networks	the US	Analytics & decision making	Modeling data-driven tax policies in most developing countries is hampered by a lack of reliable data, forecasting skills, and robust models. These impediments could be overcome through the use of emerging AI tools if concomitant analog complements are in place. The challenge in most settings is devising a tax policy that optimizes equity and productivity. The AI Economist employs AI models based on RL algorithms to model and predict tax policy design through data-driven simulations using a two-level RL framework composed of agents (workers) and tax policy to model and learn the effects of	

				dynamic tax policies in principled economic simulations.	
Virtual Assistant	Chatbot	Portugal	Citizen engagement	With ISS services receiving around 24,000 calls a day, a Virtual Assistant was developed, providing information to employers, the self-employed, and domestic workers. In addition, citizens have the opportunity to find out about isolation and illness social protection measures, social security contributions, and employment support measures.	The expectation is that this AI solution will increase the added value of ISS staff by reducing routine tasks and the time it takes to resolve bureaucratic matters, thereby making the entity more efficient and closer to the citizens it serves.
Medical decision support [32]	Image Recognition	Italy	Analytics & decision making	IRCCS Policlinico San Donato University Hospital is experimenting with AI in diagnostic imaging and electrocardiography. Automatic image reading ensures that radiologists can concentrate on interpreting complex pathologies. For doctors, these AI solutions are becoming a fundamental part of decision support, enabling more personalized treatment paths. These and other AI diagnostic solutions are still in the pilot phase, yet are providing great value for medical staff.	
Predictive Maintenance for transportation infrastructure [32]		Sweden	Analytics & decision making	Trafikverket's ambition is to predict when transportation infrastructure will require maintenance. Trafikverket also hopes to greatly improve its services utilized by people and organizations in Sweden. To	Reduce costs and ensure the avoidance of critical infrastructure failures that can cause major

				achieve their AI goals, a key early step is solving how to handle information in order to lay the foundation for AI development.	shutdowns and costly repairs.
Intelligent assistance for patent registration searches		Portugal	Personalized service delivery	A machine Learning solution analyzes data, allowing efficient patent registration searches. The solution assists Tribunal de Contas judges in their daily tasks by identifying patterns in order to find the most relevant information in an archive of several hundred thousand records.	
Workplace of the future via intelligent assistants [32]	Intelligent Automation, Virtual Assistants and Predictive Systems	Spain	Citizen engagement	Informática del Ayuntamiento de Madrid (IAM) is focusing their AI efforts on augmenting employees. These new AI solutions enhance the ability of employees to structure their work, answer citizens inquiries more effectively, and facilitate collaboration. They will be used to automate repetitive tasks with limited value-add, as well as to enable more flexible ways of supporting citizens. The intention of the project is to act as an engine of change that encourages growth, and to provide digital training for public employees.	Increase employee self-management and promote new ways of working. All Service efficiency gains and improved employee and citizen satisfaction.
Automated clinical coding [32]	NLP	Portugal		The goal is to automate the clinical coding process for procedures and diagnostics, which is part of CHUSJ's objectives to strengthen institutional capacity and promote efficient public administration. Through a	Increase efficiency and provide faster access to knowledge.

				learning engine based on Natural Language Processing, the organization is able to process unstructured text and turn it into structured classified text.	
Infrastructure inspection [32]	Image recognition		Analytics & decision making	Sund & Bælt is using Image Recognition to identify potential damage of their infrastructure. Inspections are performed using remote control drones that can quickly inspect the infrastructure, reducing the need for inspectors to perform surface inspections, and enabling more frequent inspections.	increased the accuracy and efficiency in the infrastructure inspection process. Reductions in the cost of maintenance.
Cancer detection through AI-enabled image processing [4]	deep learning	US	Analytics & decision making	Lung cancer is one of the leading causes of cancer-related deaths, and catching it early is crucial to treating the disease. Typical processes for diagnosing the disease have high rates of false positives and false negatives. Such errors can result in delays that prevent patients from receiving effective treatment. Google and Northwestern Medicine, an academic medical centre in Chicago, collaborated to develop a "deep learning" AI algorithm to review image scans used to diagnose lunch cancer. The algorithm was then able to review scans independently to predict whether a scan indicated cancer. Researchers compared the predictions of the AI system with those of radiologists with significant experience in the field. In all cases, the AI system's predictions were as accurate as those of the	

				radiologists. In some situations, the AI system outperformed the doctors	
Make city living just that much better [4]	image recognition	China	Asset movement	The city of Hangzhou, which has a metropolitan population of about 6 million, has partnered with tech firm Alibaba to launch the "City Brain" project. The initiative uses hundreds of cameras around the city to collect real-time data on road traffic conditions. These machine-readable data are then centralised and fed into to an "AI hub" which makes decisions affecting traffic lights at 128 city intersections.	The system does not simply monitor and adjust traffic based on vehicle volume; it can also make more strategic decisions, such as identifying and clearing paths for ambulances on emergency calls, reducing their travel time by 50%.
Predict the likelihood that public bus drivers would crash within the next three months [4]		Singapore	Analytics & decision making	SMRT Corporation, a public transportation organisation in Singapore, has worked with private company NEC on a pilot using AI to predict the likelihood that public bus drivers would crash within the next three months. If the AI systems indicated a high chance of a crash for a driver, they are required to take a training course. The AI pilot used historical road performance data, and two data scientists observed bus driver behaviour in order to identify potential risk factors.	
Map out and manage traffic flux [4]		Portugal	Asset management	The City Hall of Lisbon has partnered with the National Civil Engineering Laboratory (LNEC) and an academic partner, Instituto Superior Técnico, to put in place AI systems to gather, treat, classify and use urban	

				mobility and situational context data, in order to map out and manage traffic flux can in an integrated way. In addition, Portugal is implementing a project that aims to minimise response time of emergency medical service vehicles. Predictive models are being developed that can anticipate service demand by combining existing historical data and contextsensitive data from several sources (e.g. weather) to allow for more strategic deployments of vehicles.	
Monitor network traffic and detect suspicious behavior [4]		Thailand	Asset manag ement	Thailand is using AI to monitor network traffic and conduct big data analyses to detect suspicious user behaviour – for instance, two unusual logins with the same credentials, but hundreds of kilometres away	
Intelligent Security Tools [4]		the UK	Cybers ecurity	The UK National Cyber Security Centre has issued guidance on Intelligent Security Tools to help users understand considerations when using off-the-shelf AI security tools, and guide those seeking to build in-house AI security tools. It provides useful information on how to establish needs, deal with data, factor in available resources and get the most from AI. It presents a series of questions to help determine whether an AI solution is a good approach for a particular problem and set of needs.	

Virtual Assistants	Chatbot	Latvia	Citizen engagement	Latvia's Register of Enterprises has developed UNA, a 24/7 virtual assistant chatbot that provides answers in writing to frequently asked questions posed by current and future Latvian entrepreneurs, including status updates about submitted registration documents. UNA can be accessed through the Register of Enterprises website, as well as Facebook Messenger. It provides an alternative to an in-person visit or telephone call, and enables users to receive answers to questions at any time of the day.	
Virtual Assistants		Portugal	Citizen engagement	Sigma is a 24/7 virtual assistant chatbot that provides answers in writing to frequently asked questions posed by Portuguese citizens. Sigma can be accessed at ePortugal either by unregistered or registered users (in which case it will increasingly tailor its responses through NLP). In the event that Sigma recognises that its response is not adequate, it will ask the user whether they want to speak to a human, and connect them via telephone or email depending on the user's preference. Sigma provides an alternative to an in-person visit or telephone call, and enables users to receive answers to questions at any time of day. As of July 2019, Sigma had registered over 46,250 interactions.	

Land mapping [5]	Image recognition	Australia		In Australia, the Queensland Government Department of Environment and Science has adopted Machine Learning to automatically map and classify land use features in satellite imagery. Identifying different land uses (e.g. agriculture or housing) is crucial for conserving biodiversity, natural disaster monitoring, and biosecurity disease outbreak readiness and response. It can also be useful in providing a near real-time analysis of potential crops impacted during large disasters such as tropical cyclones. and floods. The process is 97% accurate. With traditional manual methods, mapping land use for the whole state takes years, but the same process takes only six weeks with new AI-based technology.	
Welfare decisions[19]		Denmark	Analytics & decision-making	Develop AI Machine Learning algorithms to help civil servants make decisions about whether citizens and businesses receive financial and other assistance from the government (e.g. support for older Danes, assistance for low income families and housing assistance). To make this possible, the government has focused on two specific challenges:	The government believes that this approach can produce more accurate and objective decisions free from human bias. In addition, it can help address the challenge of an aging population, with only a limited number of civil servants available

[19] https://govinsider.asia/innovation/exclusive-denmark-plans-to-use-ai-for-welfare-payments/

				• How to put in place proper legislation to enable automated decisions. • Making underlying data and decisions flows readable and understandable by machines	to process an increasing number of welfare requests.
Slavery from Space[20]	Image recognition	South Asia	Social services	Study high-resolution satellite data to estimate the number of brick kilns in South Asia's "Brick Belt" – an area where slavery is highly prevalent – thereby helping to calculate the extent of modern slavery in the region. Prior to this work, the full scale of brick kilns and, by proxy, slavery, was unknown, hindering action by the appropriate agencies. This innovation provides data to help NGOs and governments fight modern slavery.	Using this approach, the Rights Lab team estimates that a third of slavery may be detectable from space.
Understand local sentiments regarding refugee inflows[21]	speech-to-text analytics, sentiment analysis	Uganda	Social services	Global Pulse is the United Nations' flagship initiative on Big Data and consists of a network of innovation labs. Global Pulse is working to implement AI-driven speech-to-text analytics on local radio content to help understand local sentiments regarding refugee inflows. For instance, by analysing discussions on local radio, Machine Learning algorithms have uncovered valuable insights not previously gathered by other mechanisms.	

[20] https://oecd-opsi.org/innovations/slavery-from-space/

[21] https://www.unglobalpulse.org/project/pilot-studies-using-machine-learning-to-analyse-radio-content-in-uganda-2017/

				They have been able to identify small-scale disasters and their impact on the public, as well as surface areas of vulnerability for refugees.	
			Finland		

Table 4 Cases of AI in the Public Sector

4.10 Challenges of Applying AI in the Public Sector

It is imperative to exercise caution and vigilance when implementing AI technology to prevent potential misuse and unforeseen consequences. In light of this, the public sector must address the various economic, societal, and other impacts of AI and establish ethical and legislative frameworks to ensure safe usage within communities [2].

One of the main challenges of applying AI in the public sector include mitigating bias [2] [32] [46] [51]. The issue of negative bias can be caused by a variety of factors such as incomplete, inaccurate, or corrupted data (statistical bias), leading to predictive outcomes that favor or discriminate against specific groups of people [51]. Negative bias can have severe consequences, such as limiting access to public services such as housing and social benefits or unjust imprisonment. Even when datasets are scaled up, some degree of bias is inherent in AI models due to finite data availability. To address these issues, AI systems must be continually refined and improved as datasets and tools evolve, and weaknesses emerge. However, it can be challenging to preemptively identify sources of bias, which can make AI results appear unbiased even when they are not. Additionally, cognitive bias from AI development teams or data scientists can be a source of inherent bias that must be carefully monitored. AI companies may also manipulate data and algorithms to maximize profits (economic bias), which must be addressed through policy action and public scrutiny. To manage bias effectively, full disclosure of the datasets and algorithms used in AI is crucial.

Despite efforts to manage bias in AI, it can still emerge unconsciously throughout the project life cycle, often as a result of selective data gathering. This requires additional policies to oversee data selection processes. Data scientists may choose to collect data from certain groups based on personal preferences, resulting in biased outcomes. This is a common polling technique that yields favorable results from a select population based on factors such as gender, race, ethnicity, zip code, color, and disability. The best way to mitigate bias is through policies and processes that ensure inclusion, conscientious oversight, transparency, disclosure, and contestability. If models may influence public policy or mission-critical outcomes, the release of data collection criteria and open-source code for implemented frameworks can help mitigate the risk of producing negative results. Moreover, democratization of data and policymaking can improve practical outcomes of AI frameworks and increase trust in AI infrastructure in government.

Governments can further enhance the likelihood of a positive outcome by developing competing AI systems that focus on the same problem statement, as demonstrated by Singapore and Israel. This practice involves employing multiple solutions for one problem statement, reducing the likelihood of unintended outcomes by converging on results in different ways, even when two systems have varying degrees of bias. Additionally, AI systems could be designed to identify bias, thereby using the same tool that caused the bias in the first place to fight it. Introducing human oversight can provide an additional safeguard against machine-invoked bias, as human oversight can help detect skewed results resulting from influences such as training data manipulation, forgery, and intentional bias.

Gender bias can be inherited by AI-based tools from the data they are trained on [7]. When certain groups are under or over-represented in the data used for machine learning, it can unintentionally either ignore or overemphasize their presence, background markers, and life experiences. If historical or contemporary data is used for teaching AI without refinement or critical examination, it can cause or perpetuate sex-based discrimination and gender stereotypes, for instance in gender marketing, job profiling, recruitment tools, or image searching. AI-based tools can also reproduce and sustain sexism and gender stereotypes by possessing gender-stereotypical characteristics, such as virtual assistants given women's names and personalities associated with sexist or stereotypical "women's reactions," or physical robots, such as rescue robots, being given men's shapes. Associating these systems with a specific gender is unnecessary and reinforces gender stereotypes. Gender equality can also affect the technological side of AI solutions, as voice, speech, and face recognition systems have been found to perform worse for women than for men, with face recognition often being worse for women in certain groups.

Additional challenges of AI in the public sector include:

- **Cybersecurity** [32] [51]**.** AI systems are vulnerable to hacking, which can cause harm through the manipulation of data or training algorithms by bad actors. Phishing is one of the most common techniques used to exploit security vulnerabilities in AI systems. Good governance practices can mitigate the risks by enforcing explainability, transparency, and validation in AI systems, as well as adhering to security best practices at the technical level. Governments can employ standard security measures such as access-control lists (ACLs) and API tokens for inter-process communication (IPC) and human-facing endpoints to prevent adversarial attacks on data sources and computing resources. Proactive cybersecurity operations are necessary to prevent common patterns that can kill critical processes and conceptualize the kill chain. Both hackers and defenders benefit from finding vulnerabilities in AI systems. AI is useful for detecting zero-day exploits, which have no patches and are common targets of cyberattacks. Fuzzers are tools used by cybersecurity and AI teams to discover errors and security loopholes. Redundant systems and no single point of failure (SPOF) must be enforced to ensure data backup. Obfuscation and anti-forensics techniques are used to avoid detection, but AI can be useful in detecting obfuscation attacks and creating them. However, destructive attacks are unlikely candidates for AI prevention.

Proactive measures are necessary because cyberattacks can emerge from obscure corners of the internet, not just political cyberwarfare. Although such incidents fall under the jurisdiction of international authorities and attract stern responses from law enforcement officers and legislators, they are effective in advancing the evolution of cybersecurity best practices.

- **Control** [51]. Due to the autonomous nature of many AI systems unintended consequences can arise from machine-centric feedback loops. To avoid rogue AI systems and prevent edge cases in software development from causing harm, it is essential to proactively control, monitor, test, and validate AI systems. This can help to ensure that the outcomes of AI systems align with human values and prevent unintended consequences from occurring, even if they occur infrequently.
- **Privacy** [51] [46]. Ensuring privacy and protecting human identity should be a top priority. While ethical concerns surrounding AI may seem daunting, protecting individuals' identities within large-scale datasets and implementing access control policies are important steps in the right direction. Privacy legislation and regulatory frameworks provide a solid legal foundation for mitigating privacy risks. Additionally, governance frameworks that encourage self-assessment, peer review, and public involvement can help strengthen compliance with these legal frameworks. The specifics of these frameworks should be tailored to the context and existing mechanisms of transparency, citizen engagement, and accountability. Nevertheless, public inclusion is crucial in this process.
- **Transparency and Interpetability** [2] [7] [46] [32]. The issue of "black boxes" in AI is a new challenge, but it is not insurmountable. Technology can be part of the solution, such as training AI systems to provide more transparent reports. Policies must also be established, taking into account the practical need for understanding, which may vary by application and context. Additionally, policies should consider the amount of understanding that can be feasibly achieved by humans within reasonable timeframes for more advanced systems. Addressing this challenge is just one of several challenges that public services must overcome to implement AI, including ensuring the quality and proper use of data, securing funding, acquiring the necessary skills, integrating with other agencies, building trust, managing organizational change, and establishing the necessary digital infrastructure.
- **Scattered laws and regulations** [25] [46]. Public agencies are well-versed in designing and implementing governance frameworks, given the nature of their operations. However, a closer examination of digital transformation frameworks indicates that there is significant room for improvement. Specifically, there is a need to ensure that the digital transformation strategies align with strategic goals.
- **Actively engaging and involving civil society** [32].
- **Better coordination between IT departments** [25]. Currently, IT departments are frequently tasked with either resolving technical issues or providing updates on ongoing digital

transformation initiatives. However, for public agencies to effectively design and develop AI solutions, there must be greater collaboration and coordination among both internal stakeholders such as AI experts, data scientists, and service designers, as well as external stakeholders such as citizens, customer advocacy groups, academic researchers, and third parties.

- **Employment** [7] [46]. The issue of artificial intelligence replacing human jobs is a major concern. To alleviate these concerns, it is important to increase the awareness of AI technologies among civil servants, so they can understand the benefits of AI for their work. There are several ways to raise awareness, such as organizing regular meetings between civil servants, either within their institutions or at specialized innovation hubs. Participating in policy events organized by European institutions or other relevant parties is another effective way to raise awareness. Of course, some categories of jobs in the public administration are destined to disappear.
- **Societal fragmentation** [46]. AI-enabled algorithms have proven tremendously effective at micro-targeting content and at fostering the creation of groups of like-minded actors in the public space, such as social media platforms (Sunstein, 2017). Such groups function as echo-chambers, where citizens are sealed off from the diversity of other political opinions and social representations. This challenge affects the context in which public governance is exercised – that is the sphere of public opinion formation at large and thus, indirectly, the ability of government to both be seen as legitimate by citizens and to formulate policy actions that draw on a perceived common good. The ability of algorithms to provide personalised content by filtering out inputs that do not match pre-existing user preferences (in e.g., news, entertainment, political discourse) is potentially bringing about societal fragmentation, polarisation, and radicalisation, with the creation of digital echo chambers (Medaglia & Zhu, 2017). Governments that fail to mitigate such disaggregating forces enabled by AI systems will potentially lose the capability to be perceived as legitimate and to formulate policy actions that can be met by sufficient public opinion support. For example, the effectiveness of science-based public health initiatives, such as the ones rolled out to combat the COVID-19 pandemic, is weakened when dealing with citizen groups that unknowingly share misinformation in digital echo chambers, where distrust towards institutions is easily exacerbated.
- **Damage to the natural environment** [46]. AI has a substantial energy demand and contributes significantly to carbon emissions. However, the development of accurate metrics for measuring the environmental impact of AI technologies is still in its early stages and existing AI development practices tend to prioritize algorithm accuracy over energy efficiency.
- **Attracting and developing skills, enabling a growth mindset and a dynamic ecosystem** [32]. The ability to attract, develop, and retain AI talent such as data scientists, engineers, and domain experts is a challenge for public organizations due to budgetary constraints and high wage demands. Mitigation strategies include training current employees, creating new roles for AI, and building partnerships with other public organizations, the private sector, and academia.

- **Culture Strategic focus, openness and experimentation, while co-creating with stakeholders** [32]. The goal is to foster an experimental culture and increase understanding of AI, as well as to develop AI in limited settings before scaling AI solutions.
- **Inadequate data volumes and data management** [46]. Most AI solutions are dependent on large amounts of data, except for applications that use small data modeling or non-data driven models. This means that the effectiveness of AI applications is heavily reliant on the quality and quantity of input data. Learning systems require access to large amounts of high-quality data to refine their algorithms and enhance the accuracy and reliability of their outputs. In addition, due to their smaller size, public administrations in smaller countries often face challenges in providing AI-based services as they do not have access to the large volumes of data required to train AI algorithms. Data sharing across organizational boundaries could help eliminating this problem.
- **Underdeveloped data governance** [46]. Government agencies often lack the necessary curated data resources required for AI solutions. The acquisition of diverse data from various sources requires agreements between stakeholders on data collection, storage, access, and formatting. Data governance is a major challenge for AI solutions, with personal data governance and security being cited as the most significant barriers. If not appropriately handled, personal data governance can result in privacy breaches.
- **Conflicting organisational culture** [46]. The successful adoption of AI requires a transformation of organizational processes, routines, norms, and strategic vision. This means a transformation of the organizational culture, which can be hindered by a lack of leadership support, resistance to change, and resistance to knowledge, resource, and data sharing. To fully realize the innovation potential of AI, organizations must be willing to embrace change and create a culture that supports AI adoption.
- **Lack of skills and expertise** [46]. The successful adoption of AI solutions requires both technical and socio-technical skills, but employees and public administrations often lack the necessary expertise. AI professionals are also relatively scarce, and their high demand and preference for the private sector result in higher salaries, contributing to the high cost of adopting AI solutions in the public sector.
- **Lack of trust** [46]. AI faces a double challenge in gaining trust from users, including citizens and businesses. Firstly, as a new technology, it is still in the diffusion phase, and users may have low levels of trust in its potential impacts until they become more familiar with it. Secondly, AI touches on sensitive issues, and the shift of agency from human to machine can be unsettling, making trust a crucial factor in addressing these issues.
- **Insufficiently known impacts** [46]. The societal, organizational, and economic impacts of AI development, adoption, implementation, and use are not yet fully explored, and there is a need for further development of impact assessment, both ex-ante and ex-post. The assessment of

impact is particularly challenging within the public sector, which operates differently from the private sector. The value-laden vision of AI also needs to be taken into consideration when assessing its impact.

4.11 Ethical Principles for the use of AI

To mitigate risks associated with AI, it is necessary to adopt certain ethical principles. However, striking a balance between oversight and agility is crucial to ensure compliance with these principles.

- **Privacy and Protection** [3] [51]**.** AI solutions need to uphold an individual's civil liberties and respect their right to privacy by ensuring that data is under their control. Data usage and distribution should only be permitted with explicit individual consent, and individuals should have the ability to limit how their data is processed and request corrections or deletions.
- **Accountability** [3] [7] [51]**.** To ensure accountability throughout the AI design and implementation process, appropriate mechanisms must be put in place. This can be achieved by conducting impact assessment frameworks to identify accountability at each stage of the process, with an agency or body being responsible for monitoring accountability.
- **Safety and security** [3] [51]**.** Cybersecurity is of utmost importance in AI solutions, and it is crucial that they exhibit predictable behavior. Leaders must prioritize the well-being of both society as a whole and individual citizens in the private sector.
- **Transparency, explainability, and fairness** [3] [7] [25] [32] [51]**.** To ensure transparency and explainability of AI in the public sector, stakeholders and impacted individuals must have access to information on the algorithm, business case, data collection, design, and policy. Open-source data algorithms could enhance transparency, and individuals should receive notifications when interacting with AI or when AI makes decisions on their behalf. Regular reporting requirements on transparency should be in place, and the rights of citizens to information must be respected. Data quality should be high and representative to ensure accuracy and fairness.
- **Human control of technology** [51]**.** I systems must be subject to human oversight and control. Human review should be conducted for automated decisions made by AI systems. Individuals should have the option to opt-out of automated decision-making processes.
- **Professional Responsibility** [51]**.** Multistakeholder collaboration, accuracy, and scientific integrity of the solution should be ensured.
- **Promotion of Human Values** [3] [32] [51]**.** AI should be human- centric, promote human values and benefit society.
- **Generation of Net Benefits** [3]. The AI system must generate benefits for people that are greater than the costs.

- **Liability** [7]. AI liability refers to the legal responsibility of individuals or organizations for the actions and decisions made by artificial intelligence systems. As AI becomes more prevalent in various domains, questions arise as to who should be held accountable for the outcomes of AI systems, particularly in cases where harm or damage occurs. Some argue that liability should be attributed to the developers, manufacturers, or operators of AI systems, while others argue that the AI system itself should be held liable. This raises complex legal, ethical, and technical questions that require careful consideration and new regulations.
- **Trustwothy & Legitimacy** [7]. As AI becomes more prevalent in the public sector, it is crucial to ensure that it is trustworthy and legitimate in the eyes of citizens. This involves designing and implementing AI systems that are fair, transparent, and accountable, and serve the public interest without perpetuating biases or discrimination. Building trust in AI requires establishing clear ethical and legal frameworks, including guidelines for data privacy and security, mechanisms for holding organizations accountable for AI misuse, and regular audits and evaluations. Transparency is also vital, with clear explanations of AI use and decision-making factors, citizen notification when AI is used, and avenues for challenging or appealing AI decisions.
- **Inclusiveness** [32]. AI inclusiveness in the public sector refers to ensuring that AI is accessible and beneficial to all members of society, regardless of their backgrounds or circumstances. Inclusiveness requires addressing potential biases in AI systems that may result in discrimination against certain groups of people. It also involves ensuring that everyone has equal access to the benefits of AI technology, such as improved public services and increased efficiency in government operations.

In addition, lessons learned from AI leaders include [32]:

1. Create organizational frameworks, guidelines and principles for the ethical application of AI.
2. Follow and be engaged in the development of European and national AI strategies to ensure trustworthy development of AI.
3. Human-in-the-loop to ensure that the final decision is made by humans, with a clear structure for checking AI output.

4.12 Tools

The Oxford Insights' Government AI Readiness Index 2017 [45] assesses countries' readiness to implement AI in public service delivery based on nine input metrics, including digital skills, innovation, and existing data capabilities. The index identifies areas of improvement for OECD nations and highlights which countries need further development before rolling out public sector AI solutions. The factors considered include public service reform, economy and skills, and digital infrastructure, with a focus on AI startups and the quality of data.

The Oxford Insights' Government AI Readiness Index 2019 ranks the readiness of 194 countries to implement AI in the delivery of public services. The overall score is based on 11 metrics grouped under governance, infrastructure and data, skills and education, and government and public services. The data is obtained through desk research and the UN eGovernment Development Index. The top-ranking country in 2019 is Singapore, followed by mostly Western European countries in the top 20.

The Canadian Government has created the Algorithmic Impact Assessment[22], a digital questionnaire designed to ensure accountability, transparency, and fairness of AI outcomes in public agencies. This tool is available on the open government portal and as open-source software for anyone to use. The AIA has received international attention and has been adopted by other countries such as Germany and Mexico. The Digital 9180 network is also considering using the instrument and adapting it to their unique contexts. The AIA has been crucial in preventing potential harm to citizens and ensuring that AI outcomes positively impact society.

The Understanding Artificial Intelligence Ethics and Safety guidelines are a comprehensive set of guidelines for the public sector proposed by Turing researchers and launched by the UK's Minister for Implementation in June 2019. They aim to help public agencies identify and address the potential harm to citizens caused by AI systems and propose measures to counteract it. However, there is currently no guiding policy within the US public sector, and agencies are left to either create their own or follow guidelines from other countries.

IBM [25] has suggested a model for measuring the progress of AI initiatives in the public sector, which is called a maturity model. This model comprises five stages in ascending order of complexity, namely ad hoc, experimentation, planning and deployment, scaling and learning, and enterprise-wide transformation. The model recommends that public agencies should advance through each stage sequentially.

Finally, the European Commission has developed a Trustworthy AI Assessment List [14] to ensure the safe, ethical and transparent use of AI in the public sector. The assessment list includes seven categories that need to be addressed when implementing AI systems. These categories are human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability. The guidelines suggest that AI systems should empower humans, be technically robust and safe, respect privacy and data governance, be transparent, fair and non-discriminatory, be sustainable, and include mechanisms to ensure responsibility and accountability. The aim is to ensure that AI systems

22 https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html

benefit everyone, while also minimizing unintentional harm and negative societal and environmental impacts.

5 Conclusions

This deliverable is the the direct outcome of Task 1.1 that includes a thorough state-of-the-art analysis of scientific literature, technical reports, and national strategies regarding the implementation of AI technologies in public sector. The aim of this activity is to identify successful cases of applying AI in public sector around the globe in order to create added public value or to solve existing problems. The cases along with the targeted problems will be categorised according to the most important dimensions that will be identified. Moreover, the most important challenges and implications described in the literature will be analysed and categorised along with available tools and methods to address them. These include societal, economic, ethical, cultural, organisational, and legal implications and challenges. The result of this task will constitute the solid theoretical basis for the other WPs.

AI has tremendous potential in the public sector, and its applications are already making significant impacts in areas such as healthcare, education, public safety, and transportation. AI has the potential to improve efficiency, accuracy, and decision-making while reducing costs and improving outcomes for citizens. However, AI in the public sector also comes with challenges such as the need for transparency, ethical considerations, and the potential for bias. It is essential to ensure that AI systems are designed and implemented with the utmost care, and with the best interests of citizens at heart. As AI continues to evolve and become more prevalent in the public sector, it will be crucial to strike a balance between innovation and accountability to ensure that the benefits of this technology are realized while mitigating potential risks.

References

1 Accenture (2017). *How AI Boosts Industry Profits and Innovation.* *https://www.accenture.com/fr-fr/_acnmedia/36dc7f76eab444cab6a7f44017cc3997.pdf*

2 Accenture (2018*). Artificial Intelligence, genuine impact: Public Services in the Era of Artificial Intelligence.* *https://www.intelligentautomation.network/events-aiforgovernment/downloads/artificial-intelligence-genuine-impact-public-services-in-the-era-of-artificial-intelligence*

3 Australian Government (2019). Australia's Artificial Intelligence Ethics Framework. https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework

4 Australia Securities & Investments Commission (2022). *Corporate Plan 2022–26 Focus 2022–23.* *https://download.asic.gov.au/media/v3vhdqiw/asic-corporate-plan-2022-26-focus-2022-23-published-22-august-2022.pdf*

5 Berryhill, J., et al. (2019). *Hello, World: Artificial intelligence and its use in the public sector*, OECD Working Papers on Public Governance, No. 36, OECD Publishing, Paris, *https://doi.org/10.1787/726fd39d-en.*

6 Bughin, J., Song, J., Manyika, J., Chui, M. and Joshi, R. (2018). *Notes from the AI Frontier: Modeling the Impact of AI on the World Economy*. Report, McKinsey & Company, Washington, DC. https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-aifrontier-modeling-the-impact-of-ai-on-the-world-economy# .

7 CAHAI (Ad Hoc Committee on Artificial Intelligence) (2021). *Artificial Intelligence in Public Sector*. Council of Europe, Policy Development Group. CAHAIPDG(2021)03

8 Commission de Surveillance du Secteur Financier (2018). *Artificial Intelligence. Opportunities, Risks and Recommendations for the Financial Sector.* *https://www.cssf.lu/wp-content/uploads/files/Publications/Rapports_ponctuels/CSSF_White_Paper_Artificial_Intelligence_201218.pdf*

9 CSSF (2018), Artificial Intelligence: Opportunities, Risks and Recommendations for the Financial Sector, Luxembourg, Commission de Surveillance du Secteur Financier, www.abbl.lu/content/uploads/2018/12/CSSF_White_Paper_Artificial_Intelligence_201218.pdf

10 Deloitte AI Institute (2021). *Scaling AI in Government. How to reach the heights of Enterprisewide adoption of AI.* *https://www2.deloitte.com/content/dam/insights/articles/us164831_cgi-state-of-ai-for-gov_/DI_CGI-state-of-ai-for-gov.pdf*

11 European Commission (2018a). *EU Member States sign up to cooperate on Artificial Intelligence.* *https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence*

12 European Commission (2018b). *Coordinated Plan on Artificial Intelligence*. https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence

13 European Commission (2019a), A Definition of AI: Main Capabilities and Disciplines, definition developed for the purpose of the AI HLEG's deliverables, Independent High-Level Group on Artificial Intelligence set up by the European Commission, Brussels, https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligencemain-capabilities-and-scientific-disciplines

14 European Commission (2019b). *Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai*

15 European Commission (2020a). *Commission Report on safety and liability implications of AI, the Internet of Things and Robotics. https://ec.europa.eu/info/publications/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics-0_en*

16 European Commission (2020b). *White Paper on Artificial Intelligence: a European approach to excellence and trust. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en*

17 European Commission (2021). *Proposal for a Regulation laying down harmonised rules on artificial intelligence. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence*

18 European Commission, Joint Research Centre, Manzoni, M., Medaglia, R., Tangi, L. (2022a). *AI Watch, road to the adoption of artificial intelligence by the public sector: a handbook for policymakers, public administrations and relevant stakeholders,* Publications Office of the European Union. https://data.europa.eu/doi/10.2760/288757

19 European Commission (2022b). *Natural Language Processing for Public Services*. https://joinup.ec.europa.eu/collection/catalogue-services/document/study-natural-language-processing-public-services

20 European Parliament (2017). *Civil Law Rules on Robotics. https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html*

21 European Parliament (2020). *Civil liability regime for artificial intelligence. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html#title1*

22 Fatima, S., Desouza, K. C., & Dawson, G. S. (2020). *National strategic artificial intelligence plans: A multi-dimensional analysis.* Economic Analysis and Policy, 67, 178-194. https://doi.org/10.1016/j.eap.2020.07.008

23 G20 (2019). *G20 Ministerial Statement on Trade and Digital Economy. https://www.mofa.go.jp/files/000486596.pdf*

24 IBM (2021). IBM Institute for Business Value—Research, reports, and insights.. https://www.ibm.com/thought-leader¬ship/institute-business-value.

25 IBM Center for The Business of Government (2021). Artificial Intelligence in the Public Sector: A Maturity Model. https://www.businessofgovernment.org/report/artificial-intelligence-public-sector-maturity-model

26 Kitchenham B. and Charters, S. (2007). "Guidelines for performing systematic literature reviews in software engineering," Keele University & University of Durham, EBSE Technical Report EBSE 2007-01, 2007.

27 LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*. Nature, 521(7553), 436–444. doi:10.1038/nature14539

28 Marcia, V., Desouza, K. C., (2021). *The EU Path Towards a Regulation on Artificial Intelligence*. Brookings. https://www.brookings.edu/blog/techtank/2021/04/26/the-eu-path-towards-regulation-on-artificial-intelligence

29 McKinsey Global Institute (2021). *The state of AI.* https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2021

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

31 Mehr, H. (2017). *Artificial Intelligence for Citizen Services and Government*. Harvard Ash Center for Democratic Governance and Innovation, 19. https://ash.harvard.edu/publications/artificial-intelligence-citizen-services-and-government.

32 Microsoft (2020). *Artificial Intelligence in the Public Sector. European Outlook for 2020 and Beyond. How 213 Public Organizations Benefit from AI. https://info.microsoft.com/rs/157-GQE-382/images/EN-CNTNT-eBook-artificial-SRGCM3835.pdf*

33 Miller, H., Stirling, R. (2019). *Government Artificial Intelligence Readiness Index 2019. https://ec.europa.eu/futurium/en/system/files/ged/ai_readiness_index_2019__0.pdf*

34 Ministry of Business, Innovation and Employment (2017). *Strategic Science Investment Fund Investment Plan 2017 – 2024.* https://www.mbie.govt.nz/dmsdocument/1435-strategic-science-investment-fund-investment-plan-pdf

35 Misuraca, G., & van Noordt, C. (2020). *Overview of the use and impact of AI in public services in the EU* (JRC EUR 30255). Publications Office of the European Union. https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/ai-watch-artificial-intelligence-public-services

36 Nili, A., Barros, A., & Tate, M. (2019). The public sector can teach us a lot about digitizing customer service. MIT Sloan Management Review, 60(2), 84-87.

37 UK Office for Artificial Intelligence, Government Digital Service (2020). *A guide to using artificial intelligence in the public sector.* *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/964787/A_guide_to_using_AI_in_the_public_sector__Mobile_version_.pdf*

38 Organisation for Economic Co-operation and Development (OECD) (2018a). *Science, Technology and Innovation Outlook 2018*, OECD Publishing, Paris, www.oecd.org/sti/oecd-science-technology-and-innovation-outlook-25186167.htm

39 Organisation for Economic Co-operation and Development (OECD) (2018b), *OECD Regulatory Policy Outlook 2018*, OECD Publishing, Paris. https://doi.org/10.1787/9789264303072-en .

40 Organisation for Economic Co-operation and Development (OECD) (2019). *Recommendation of the Council on Artificial Intelligence*, OECD Legal Instruments, OECD/LEGAL/0449, 22 May 2019, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (accessed 20 August 2022).

41 Organisation for Economic Co-operation and Development (OECD) (2021). *Unlocking the potential of crowdsourcing for public decision-making with artificial intelligence*. Observatory of Public Sector Innovation. *https://oecd-opsi.org/innovations/unlocking-the-potential-of-crowdsourcing-for-public-decision-making-with-artificial-intelligence/.*

42 PwC (2017). *The economic impact of artificial intelligence on the UK economy.* *https://www.pwc.co.uk/economic-services/assets/ai-uk-report-v2.pdf*

43 Partnership for Public Service/IBM Center for the Business of Government (2019). More than Meets AI: Assessing the Impact of Artificial Intelligence on the Work of Government, Washington, DC, www.businessofgovernment.org/sites/default/files/More%20Than%20Meets%20AI.pdf .

44 Savitz, S., Davenport, A., & Ziegler, M. (2020). The Marine Transportation System, Autonomous Technology, and Implications for the U.S. Coast Guard. RAND Corporation. *https://doi.org/10.7249/PE359.*

45 Stirling, R., Miller, H., & Martinho-Truswell, E. (2017). Government AI Readiness Index. *https://www.oxfordinsights.com/government-ai-readiness-index.*

46 Tangi, L., Van Noordt, C., Combetto, M., Gattwinkel, D. and Pignatelli, F. (2022) *AI Watch. European landscape on the use of Artificial Intelligence by the Public Sector*, EUR 31088 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53058-9, doi:10.2760/39336, JRC129301.

47 Ubaldi, B. et al. (2019), "State of the art in the use of emerging technologies in the public sector", OECD Working Papers on Public Governance, No. 31, OECD Publishing, Paris, https://doi.org/10.1787/932780bc-en

48 U.S. Defense Innovation Unit (DIU) (2020). *Responsible AI Guidelines in Practice. Lessons learned from the DIU portfolio.* *https://assets.ctfassets.net/3nanhbfkr0pc/acoo1Fj5uungnGNPJ3QWy/3a1dafd64f22efcf8f27380aafae9789/2021_RAI_Report-v3.pdf*

49 U.S. Department of Defense (2020). *Ethical Principles for Artificial Intelligence*. *https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/*

50 Wilson, H. J., Daugherty, P., & Bianzino, N. (2017). The jobs that artificial intelligence will create. *MIT Sloan Management Review*, *58*(4), 14.

51 Wohlin, C. (2014) *"Guidelines for snowballing in systematic literature studies and a replication in software engineering",* in Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, ser. EASE '14. New York, NY, USA: ACM, 2014, pp. 38:1–38:10. [Online]. http://doi.acm.org/10.1145/2601248.2601268

52 World Bank (2020). *Artificial Intelligence in the Public Sector: Maximizing Opportunities, Managing Risks. E*quitable Growth, Finance and Institutions Insight; World Bank, Washington, DC. © World Bank. *https://openknowledge.worldbank.org/handle/10986/35317*

53 World Economic Forum (2018*). Harnessing Artificial Intelligence for the Earth.* *https://www3.weforum.org/docs/Harnessing_Artificial_Intelligence_for_the_Earth_report_2018.pdf*

54 World Economic Forum (2019*). Empowering AI Leadership An Oversight Toolkit for Boards of Directors*. https://wef-ai.s3.amazonaws.com/WEF_Empowering-AI-Leadership_Oversight-Toolkit.pdf

55 World Economic Forum (2022). *Empowering AI Leadership: AI C-Suite Toolkit.* *https://www.weforum.org/reports/empowering-ai-leadership-ai-c-suite-toolkit/*