COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Strasbourg, 21 May 2021

CAHAI-PDG(2021)06
Provisional

AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI)

POLICY DEVELOPMENT GROUP (CAHAI-PDG)

Artificial Intelligence in the Public Sector[1]

www.coe.int/cahai

[1] This draft document is going to be reviewed by the CAHAI-PDG and should by no means be considered as final.

Table of Contents

1. Introduction 3
2. AI and the public sector: an overview 3
3. Use cases of AI applications deployed by CoE Member States 7
3.1 Examples from the judicial domain 7
3.2 Examples from the tax & finance domain 8
3.3 Examples from the healthcare domain 9
3.4 Examples from the law enforcement domain 9
3.5 Examples from the social security domain 11
3.6 Examples of AI applications used in other domains 12
4. Benefits and risks of the use AI in the public sector 13
5. Policy Guidance for public actors seeking to adopt AI 15
5.1 Design Phase 15
5.2 Procurement Phase 17
5.3 Development Phase 19
5.4 Deployment Phase 22
5.5 Continous Education 25
6. Conclusions 26

1. Introduction

In December 2020, the CAHAI adopted a Feasibility Study which sets out the opportunities and challenges raised by Artificial Intelligence (AI) with regard to the protection and enablement of human rights, democracy and the rule of law. In addition to providing a comprehensive mapping of the issues at stake in the context of AI, the Feasibility Study also laid the groundwork for the CAHAI's endeavours in 2021.

Besides identifying the possible elements for draft legal instrument(s) on AI, the CAHAI also found that it would be beneficial to develop further policy guidance on certain themes that were touched upon in the Feasibility Study. One such theme concerns the use of AI in the public sector, given the specific benefits and risks that arise in this context. Increasingly, public entities, or delegated private subjects, are making use of AI applications in their daily operations, both at the local and at the national level. Moreover, AI applications are penetrating virtually all public activity domains, from tax authorities and social welfare administrations to law enforcement and the judiciary. These applications can lead to important efficiencies that can benefit citizens in both direct and indirect ways, but they can also generate certain adverse effects on human rights, democracy and the rule of law.

This document therefore seeks to map the impact raised by the use of AI in the public sector and aims to fomulate policy guidance for Member States to ensure that their obligation to respect human rights, and to ensure the integrity of the democratic process as well as the rule of law, can be fulfilled.

It should be noted that the scope of this document will be limited to the design, development and deployment of AI by public sector entities and private sector entities when they exercise public power. The use of AI by private entities that can nevertheless have an impact on the public sphere does not fall in the scope of this document. Hence, while the use of AI-powered microtargeting practices on social media could for instance have an impact on electoral integrity – and thereby also affect the public sector – it will not be discussed in what follows.

2. AI and the public sector: an overview

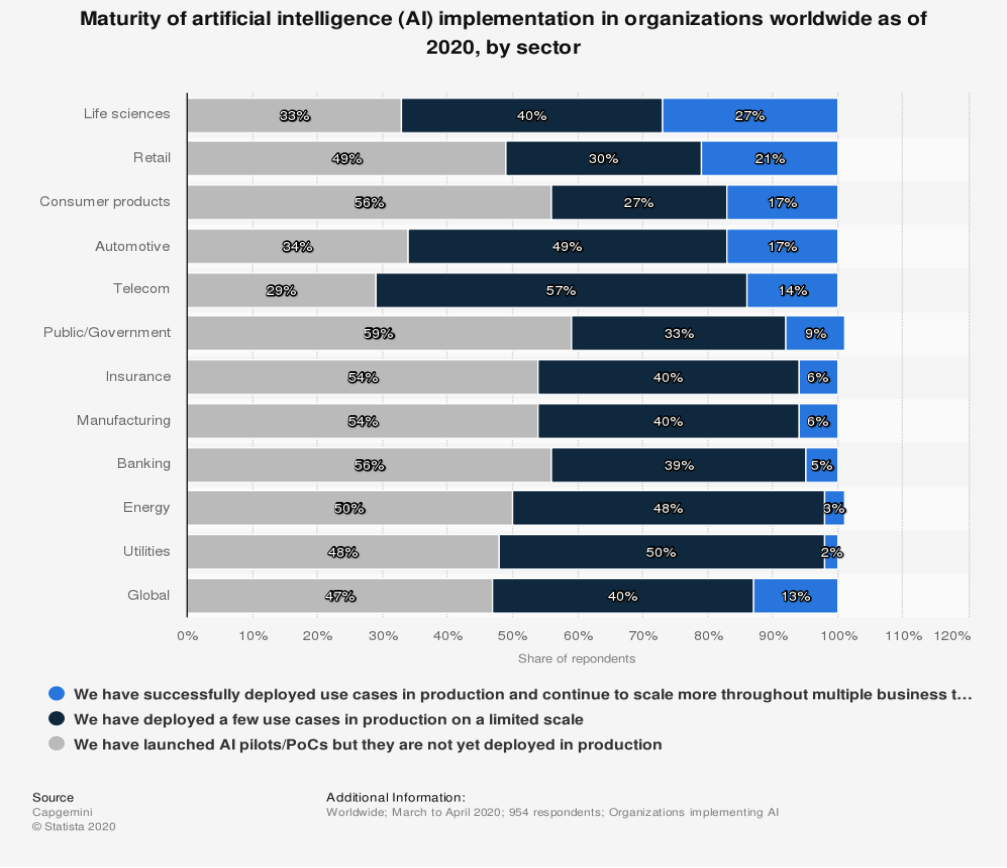
Countries around the world and international organizations such as the European Union, UNESCO, or OECD have understood the tremendous economic potential of AI, which is considered a strategic technology. Most of the Council of Europe member States already adopted AI strategies or otherwise declared the adoption of AI as one of their priorities amidst broader digitalisation strategies. The adoption of AI in the public sector is often mentioned as part of such strategy.

Indeed, most national strategies address the use of AI in the public sector, notably to deliver better public services for the benefit of citizens and enhance efficiency through automating routine government processes, and coordination in the public administration. What is more, some Member States see the public sector as being a leader in pushing for the development and use of AI. Some member States also see potential for AI to help guide governmental decision-making (e.g. in the areas of public safety, public health or policy evaluation).

According to the OECD's initial mapping on AI, out of the 50 countries (including the EU) that introduced national AI strategies, 36 have specific strategies for AI in the public sector. Most of these strategies follow similar themes, which cover economic development, trust and ethics, security and enhancing the talent pipeline. While there is no uniform AI legislation in relation to the use of AI in the public sector, many governments have launched a variety of national projects in recent years to utilize AI. It can be observed that AI is often used to improve efficiency and decision making, foster positive relationships with citizens and business, or solve specific problems in critical fields such as health, transportation and security[2].

AI applications can help provide cost efficiency and savings through intelligent automation, as well as labor and capital augmentation where AI can complement the skills of existing workforces[3]. Ideally, the deploument of AI systems in the public sector can also provide civil servants with a shift from mundane tasks to high-value work, while improving the speed and quality of public services.

Statistics of Statista show the adoption of artificial intelligence in the public sector in comparison with other organizations, by sector (Please see Fig.2).



[2] 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.

[3] Bernd W. Wirtz, Jan C. Weyerer & Carolin Geyer (2019) Artificial Intelligence and the Public Sector—Applications and Challenges, International Journal of Public Administration, 42:7, 596-615, DOI: 10.1080/01900692.2018.1498103

Fig.1. Source (• Maturity of AI implementation by sector worldwide 2020 | Statista)

However, AI-enabled tools are increasingly being tested or used by the public sector[4]. Much of the debate centres around the use of algorithmic or automated decision making (ADM) systems. 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.

Source: Round Table on Artificial intelligence and the Future of Democracy, Council of Europe, 2019

The legal bases for using ADM may vary in Member States. According to German law[5], for instance, automated decision-making can be used only when there is no margin of discretion and when the decision to be made is yes or no. In all cases, it should be possible to opt out, to re-evaluate the process and to explain how the decision was taken. In the Netherlands and Belgium, the situation is as follows: the principle of legality demands a basis in the law for decision making (with legal consequences/when human rights are in play), irrrespective of whether information systems are used or not. Of course, existing Council of Europe standards on human rights, democracy and rule of law already apply to the use of AI in Council of Europe member states. For instance, the processing of personal data is governed by Convention 108, and hence complements legislation that applies more specifically to the public sector context.

At the same time, it is also recognised that the adoption of AI applications brings forth certain risks. Given the important role and mandate of the public sector in society, the mitigation and prevention of these risks are a crucial task for governments who seek to deploy AI.

Most Member States acknowledge 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, including

4 See for instance the report of Algorithm Watch, Automating Society, 2020, page 6.
5 https://www.gesetze-im-internet.de/englisch_bdsg/englisch_bdsg.html#p0310.

aspects like 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. In this regard, it is also important to stress that Member States are direct addressees of human rights obligations, which they must also fulfil whenever they make use of AI. Moreover, they have the task to safeguard the facilities and infrastructures that enable the democratic system and the rule of law.

Some examples coming from member and non-member States on policies and guidance for the use of AI in the public sector

Alan Turing Institute, Understanding artificial intelligence ethics and safety – This guide is an end-to-end guidance on how to apply principles of AI ethics and safety to the design and implementation of algorithmic systems in the public sector.

Canadian directive on automated decision-making – The Canadian government has developed a risk-based approach to AI adoption in the public sector which divides the AI systems in different levels. The four factors used to determine the risk-level are impact on: the rights of individuals or communities, the health or well-being of individuals or communities, the economic interests of individuals, entities, or communities and the ongoing sustainability of an ecosystem. Based on the risk-level, the guide provides insights on how to best approach AI procurement from a proportionality view and to what extent each requirement should be applied.

In the Feasibility Study, several adverse effects of AI in this context were raised. In particular, the study has clearly demonstrated how AI systems can have a negative impact on a wide range of protected civil, political and social rights protected by the European Convention on Human Rights and by the European Social Charter, as well as on rule of law and democracy. These range from the right to fair trial to the right to respect for the private and family life, from the right to freedom of expression and association to the right to equality and non-discrimination as well as the right to social security.

At the same time, it also point out that "*Where used responsibly, they can also enhance the rule of law and democracy, by improving the efficiency of administrative procedures and helping public authorities being more responsive to the public's needs while freeing up time to tackle other complex and important issues. AI systems can also help public actors better identify the needs and concerns of the public, as well as to inform analyses and decisions, contributing to the development of more effective policies*."

The goal is thus to find an appropriate and balanced approach to deploy AI systems in the public sector that can improve efficiency and help public actors to identify the needs and concerns of the public, while at the same time avoiding adverse effects of AI systems. This requires not only an appropriate legal framework that provids the contours in which AI systems can legally operate, but also the right skills. Member States for instance recognise the fact that, if they want to make AI in the public sector a success story, they will also need to invest in capacity building of public officials. Without having the right skills, public services will not be able to reap the benefits of

AI, nor will they know which critical questions to ask in order to ensure that the use of AI complies with standards of human rights, democracy and the rule of law.

3. Use cases of AI applications deployed by CoE Member States

Various Council or Europe member states already make use of AI systems in their public servies. In this section, some concrete examples are provided of AI applications that are currently piloted or deployed.

3.1 Examples from the judicial domain

Hungary: Speech recognition and transcription project

The project explores the use of a speech recognition and transcription software in courts in order to facilitate and expedite the drafting of court decisions and minutes. It would result in reducing manual effort.

This project is exploring the use of speech recognition and transcription software in the courts and aims to facilitate compliance with deadlines associated with the obligation to put decisions and minutes into writing. It would also result in more efficient use of work time by reducing time spent on transcription. In 2018, 726 speech recognition and transcription licenses were purchased for the courts.

Switzerland: use of AI in the Penal System

The Swiss Execution of Penal Sentences and Justice is based on a system of levels. According to this system, inmates are generally granted increasing amounts of freedom as the duration of their imprisonment continues. Of course, the risk of escape and recidivism are decisive factors when it comes to granting these greater freedoms. In recent years, and in response to convicted felons committing several tragic acts of violence and sex offenses, the ROS (Risk-Oriented Sanctioning) was introduced. Today, ROS is used in all German-speaking Cantons of Switzerland.

The primary objective of ROS is to prevent recidivism. During a triage, cases are classified according to their need for recidivism risk assessment. Based on this classification, a differentiated individual case analysis is carried out, which is later developed into an individual execution plan for the sanction of the corresponding offender. This triage is performed by an ADM-tool called the Fall-Screening-Tool (Case Screening Tool, FaST). FaST automatically divides all cases into the different classes according to the presumed risks of escape and recidivism.

This classification is determined by using criminal records and is based on general statistical risk factors including age, violent offenses committed before the age of 18, youth attorney entries, number of previous convictions, offense category, sentences, polymorphic delinquency, offense-free time after release, and domestic violence. This classification is carried out fully automatically by the ADM-application. However, it is important to note that this is not a risk analysis, but a way of screening out the cases with increased assessment needs[6].

[6] Treuhardt/Kröger 2018 p. 24-32

Nevertheless, the triage classification has an effect on how those responsible at a particular institution make decisions and which assessments are to be made. No ADM applications are apparent in the other stages of ROS. FaST is, therefore, only used during the triage stage.

3.2 Examples from the tax & finance domain

Ireland: Revenue AI voicebot helping citizens handle tax clearance

In early 2018, the Irish Revenue Commissioners initiated a pilot project to examine if AI-based Natural Language Processing (NLP) technologies could be used to deliver an improved customer service, reduce costs and increase efficiencies. This resulted in the implementation of a Virtual Digital Agent (VDA) or voicebot designed to focus on a subset of calls from the Irish taxpayer relating to tax clearance[7].

A suite of integrated technologies was used to convert customer speech to text, understand the text using NLP so that a response could be formulated, and then convert this response back to speech so the customer could hear the answer. The Irish Revenue demonstrated that voicebot technology can offer a fully automated service, providing an efficient, effective experience for customers.

Ireland: Machine Learning to predict appeals after Revaluation at Ireland's Valuation Office (pilot)

The Valuation Office is an independent Government Office, under the aegis of the Irish Department of Housing, Local Government and Heritage. The core business of the organisation is to provide stakeholders with accurate, up-to-date valuations of commercial and industrial properties. These valuations are integral to the business rating system in Ireland and form the basis for a very significant element of local government revenue each year.

"Applying Machine Learning to predict appeals after Revaluation at the Valuation Office (IRL)" was pilot research as part of AutoVal - Automated Commercial Property Valuation using Machine Learning[8]. The purpose of the research module outlined was to predict whether a commercial property will go to Valuation Tribunal after Revaluation by applying machine learning algorithms for binary classification. The benefits of such

[7] VDA is a Conversational voice bot, enabled via IBM Watson's cloud-based NLP (Natural Language Processing) engine. It uses additional cloud services for speech-to-text and text synthesis (Google & AWS/Polly), adopts a smart-suggestions conversation-flow, driven by the existing customer status & recent previous correspondences and manages queries relating to Tax Clearance. It provides a capability to apply for Tax Clearance over the phone, via the voicebot and transfers to a Live Customer Service Agent as necessary. Finally, it provides a UI interface for Revenue staff, containing summary dashboards, call transcripts & links to the audio recordings.

[8] The ML techniques used post exploratory data analysis where: Logistic regression - a statistical model that in its basic form uses a logistic function to model a binary dependent variable; Random forest classifier - an ensemble model that fits multiple decision trees classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting in comparison with individual decision trees; Gradient boosting classifier - a machine learning technique which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them to get the outcome.

ML research has tangible benefits in terms of 'flagging' of properties likely to appeal and also in the reduction of the number of appeals to the Valuation Tribunal following a Revaluation programme.

Models on data from two Revaluation programmes achieved an accuracy of between 80-90%. Applying the models on the Revaluation 2019 data did not work as well (~35%) and it was concluded that retraining the model using newer data was needed.

3.3 Examples from the healthcare domain

Russia: City of Moscow Centre for-Diagnostics and Telemedicine

The AI pilot in radiology combines two different uses of AI :

- assisting doctors in tumour detection on x-ray images; and
- taking dictation for observation notes as dictated by the doctor completing the patient's medical chart. This is estimated to save 25% of doctor's time and delivers an automatic, complete and accurate record for each patient. The system is based on Voice2Med toolkit by Centre of Speech Technologies (Sankt Petersburg) and can be scaled up and replicated at minimum cost.

Both uses of AI significantly improve the performance, accuracy and efficiency of doctors' work.

Russia: medical supplies in the Tomsk region

The region introduced a special legal regime as a pilot to authorise delivery of medical supplies by autonomous drones to remote regions over hundreds of kilometres. The region measures 600km north to south, and 780km east to west.

Turkey: AI applications to counter the COVID-19 pandemic

AI applications have been implemented in the field of health within the scope of fighting with COVID pandemic. Turkey has set up a 'COVID-19 Technology Platform' to coordinate R&D projects for the discovery of new medicines and vaccines. It currently hosts 17 different projects. Big data search techniques are used to identify drug candidates among thousands of molecules with the help of AI. A specific call was also launched for project proposals for researchers and R&D companies to propose solutions to fight the COVİD-19.

3.4 Examples from the law enforcement domain

Netherlands: The processing of online declarations of internet scams by the police

An AI system is being used in the Netherlands in the field of law enforcement , to offer an advisory function to citizens in making online reports of internet fraud.

The citizen first tells, in a free text field, what happened. The application then asks additional questions using text analysis. Based on the text and additional answers, the model classifies the declarant and gives the declarant advice and perspective on what the most meaningful follow-up action is.

Netherlands: Housing Fraud detection, Municipality of Amsterdam[9]

This AI system is being tested[10] in Amsterdam to detect hausing frauds. Since housing is scarce in Amsterdam, he municipality is committed to prevent illegal use of homes - for example with double registrations, fraud with housing subsidies or excessive holiday rental. Law enforcers are increasingly helped with detection of this type of fraud by data analysis and forecasting models.

Based on a large number of factors, the risk of fraud per address gets calculated in advance. So the likelihood that enforcers catch actual fraud during their unannounced house visits increases.

Switzerland: use of AI for Predictive Policing

In some cantons, in particular in Basel-Landschaft, Aargau, and Zurich, the police use software to help prevent criminal offenses. They rely on the commercial software package “PRECOBS” (Pre-Crime Observation System), which is solely used for the prognosis of domestic burglaries. This relatively common crime has been well researched scientifically, and police authorities usually have a solid database regarding the spatial and temporal distribution of burglaries as well as crime characteristics. Furthermore, these offenses indicate a professional perpetrator and thus show an above-average probability of subsequent offenses. In addition, corresponding prognosis models can be created using relatively few data points. PRECOBS is, therefore, based on the assumption that burglars strike several times within a short period if they have already been successful in a certain location.

The software is used to search for certain patterns in the police reports on burglaries, such as how the perpetrators proceed and when and where they strike. Subsequently, PRECOBS creates a forecast for areas where there is an increased risk of burglary in the next 72 hours. Thereupon, the police send targeted patrols to the area. PRECOBS thus generates forecasts on the basis of primarily entered decisions and it does not use machine learning methods. Although there are plans to extend PRECOBS in the future to include other offenses (such as car theft or pickpocketing) and consequently create new functionalities, it should be noted that the use of predictive policing in Switzerland is currently limited to a relatively small and clearly defined area of preventive police work[11].

Russia: National inspectorate for road traffic (GIBDD)

[9] For more information see (in Dutch): Gemeente Amsterdam magazine , Amsterdamse Intelligentie, accessible here: https://assets.amsterdam.nl/publish/pages/922120/magazine_ai.pdf , page 16.

[10] This prototype is still being tested. Development is done in-house and some UVA students. The Machine Learning technique used is decision tree and random forest.

[11] Blur 2017, Leese 2018 p. 57-72

Image recognition on traffic cameras is gradually rolled out nationwide. Cameras automatically detect traffic violations, including, inter alia, public transport lanes violation, aggressive driving behaviour, mobile phone use, or missing seat belts.

Germany: Predictive Policing, Federal level[12]

Since 2017, the Bundeskriminalamt (Federal Crime Agency, BKA) has used the risk-assessment tool RADAR-iTE[13] to sort “militant Salafists” into three threat levels (high, conspicuous, and moderate). The system was developed in cooperation with the Department of Forensic Psychology at the University of Konstanz. In order to assess a person already known to the authorities, the caseworker fills in a standardized questionnaire about the “observable behavior” of the subject, drawing on data the police previously gathered on the person and everything the police is legally authorized to access. Once the results of the corresponding threat levels are provided, the caseworker (or the respective department) decides what action to take.

Russia: Rosfinmonitoring (Russian financial markets monitoring service, RFM)

In 2019 RFM started testing AI in preventing money laundering, terrorism financing and financial crime (AML) activities. In particular, in collaboration with 40 other states the system is used to track down financing for ISIS, and uses models of behaviour to identify money laundering activities.

3.5 Examples from the social security domain

Netherlands: Application 'Work Explorer' of the Employee Insurance Agency (UWV)

The UWV employment company is responsible for supporting unemployed workers in finding paid work. The support consists of services such as a workshop, webinar, online training and often starts with a personal meeting at a UWV branch. Because both the number of advisers and the available services are limited, the employment company must determine how quickly unemployed workers are eligible for an interview and which services can best be deployed. It is for this purpose that they have created this application.

After unemployment benefits have been granted, the unemployed person will see a questionnaire in the online "Workbook". They answer questions about employment history, the personal situation and an assessment of their own chances on the labor market. Combined with data about employment history and education, the Work Explorer application then predicts the chance of finding paid work within a year (expressed as a percentage), supplemented with a diagnosis of personal obstacles and opportunities for unemployed workers to find paid work (a percentage per factor).

[12] AlgorithmWatch, Automating Society Report, 2020, available at: https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf

[13] Bundeskriminalamt, 2017

The applicant will not see the percentages, unless they ask or this is explicitly discussed during the personal interview.

3.6 Examples of AI applications used in other domains

Germany: Crises Management in Foreign Policy, Federal Level[14]

The Federal Foreign Office uses the data analytics tool PRE- VIEW– Prediction, Visualization, Early Warning[15] to identify evolving international crises. The tool analyzes publicly available data related to current political, economic, and societal trends and conflicts in order to identify developing crises. According to the Federal Foreign Office, AI is used to process the data, which is then used to produce infographics and maps to help provide insights into the state of a particular conflict. Furthermore, trend analyses illustrate how political and societal developments may evolve. *PREVIEW* is deployed by the Federal Foreign Office's *Department S*, which oversees international stabilization measures and crises engagement. The output also supports the civil servants in determining which steps to take next.

Germany: Identity Check of Migrants, Federal level[16]

The Bundesamt für Migration und Flüchtlinge (Federal Office for Migration and Refugees, BAMF) has been using automated text and speech recognition systems to identify refugees17 since 2017. Agency employees can ask asylum seekers to give them access to their cell phone, tablet, or laptop to verify if they are telling the truth about where they come from. The agency has the ability to obtain all the data contained on the devices and run software on it. The software presents the employee with a limited overview of the content, which also includes language analysis of the text retrieved. According to the BAMF, both the software and the hardware was provided by the firm Atos SE18 , however, VICE Magazine found evidence19 that the mobile forensic technology firm MSAB was also involved. Another tool deployed by the BAMF aims to identify disguised dialects in speech 20. When an asylum seeker does not have a valid proof of ID, a two-minute voice recording of the person describing a picture in their mother tongue is analyzed by software, which then calculates a percentage of how close the speech comes to a certain dialect.

Ireland: Using AI to count trees

An initial feasibility study has been conducted with CeaDAR (Ireland's National Centre for Applied Data Analytics and Machine Intelligence) to explore potential to use AI

14 AlgorithmWatch, Automating Society Report, 2020, available at: https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf
15 Auswärtiges Amt, 2019
16 AlgorithmWatch, Automating Society Report, 2020, available at: https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf
17 Thüer, Köver and Fanta, 2018
18 Biselli, 2017
19 Biselli, 2018b
20 Biselli, 2018a

and/or machine learning to improve forecasting of timber volumes and quality from Ireland's forests using multi-modal data such as Lidar, inventory data, historical harvesting data, climate, soils, etc.

European Commission: using AI to identify risk of fraud and irregularities in publicly funded projects

Currently, the EDPS can only provide one example of AI applications used by EU Institutions, Bodies or Agencies, namely the Arachne system[21], whose users are EU member states auditors and the EU Commission auditor. This system, used in the field of public project funding, aims to identify the projects, beneficiaries, contracts and contractors which might be susceptible to risks of fraud, conflict of interest and irregularities. It is based on collection of data from internal and external sources and data mining for calculating risk indicators.

Poland has also provided a variety of examples of the use of AI in all the sectors mentioned above. AI systems being used aim to detection fraud in tax system, inform citizens through a chatbot, make predictions as to the development of COVID-19 pandemic, analyze voice and text of protocols of trials, calculate critical pools those need public aid intervention, compare education predictions of students and analyze social health edge system. The main capabilities of AI systems used are recognition, predicition and provision of advice, supported by different core techniques[22]. The AI systems being used have a basic level of automation, with human supervision.

4. Benefits and risks of the use AI in the public sector

The examples above demonstrate that AI applications, or, more precisely, expert systems and applications based on machine learning, are being introduced in many different fields of public sector intervention in several Council of Europe member states, and at different levels (municipal, regional and state level).

Some systems aimed at helping citizens directly, others assist public servants in the exercise of their functions, and yet others are universally relevant.

The use of AI should in principle help increase citizens' satisfaction with public services; this is the case for instance of AI systems which can answer their frequently asked questions through chatbots or virtual assistants, and provide assistance 24 hours for 7 days a week. In addition, AI systems can contribute to improving public servants' job satisfaction when deployed in a manner that allows them to carry out repetitive administrative tasks more speedily and spend more time on interaction with citizens or other core tasks and allow, ensuring at the same time a more rationale use of public resources.

[21] More information available at European Social Fund (ESF) - Employment, Social Affairs & Inclusion - European Commission (europa.eu)
[22] Machine Learning, Computer vision, Reinforcement learning, supervised learning, expert systems, knowledge representation, neural networks, planning and scheduling, searching and optimization, natural language processing.

In terms of the Council of Europe's driving values, most of the applications indicated above appear to be either neutral or to have a potential of positive impact on either human rights, the rule of law and democracy. This is the case for instances of AI applications aimed at fraud/corruption detection, or better access to education. As a general remark, by amplifying the capacity of citizens to access, share and report information, some AI applications can contribute to the democratic oversight of public institutions and strengthen their accountability.

The examples of AI applications presented above fall into several different categories: some provide for fully automated decision-making, some enable decision-making advice, others provide decision-relevant information, and yet others have no apparent connection with decision-making.

In all cases, the adoption of AI solutions seems to be driven by an obvious promise of increased efficiency of the decision-making and other processes in the public sector, where efficiency would comprise speed, cost, and quality. These promise needs however to be measured against possible risks of violation of human rights, rule of law and democracy. Some of the examples above show that some AI applications are still in a trial stage, with tests being carried out at a small scale before large-scale deployment at the national level. In other member States which have deployed some solutions already at the national level, it is likely that processes of verification and weighting of potential benefits against potential risks have already taken place. It would be interesting to receive further information from member States as regards such processes (for instance audits and evaluation of the effects of the applications, including feedback from concerned stakeholders, which might have taken place).

This would be even more desirable considering that the examples of use of AI applications in member States listed above do not provide indications as regards possible risks associated to their deployment. However, as the CAHAI's feasibility study noted, there are also risks arising from the use of the AI in the public sector that need to be considered.

Risks resulting from a poor design and development of AI systems – such as embedding bias and discrimination into service delivery which would result in amplifying the scale of harmful outcomes, or issues of consent around the data sets being used to build AI systems, or of robustness and safety of AI systems need to be considered. If the use of AI systems can render public institutions more efficient, this could happen to the potential detriment of transparency, human agency and oversight. Furthermore, public authorities often depend on private actors to procure and deploy AI-systems, which creates a risk of further eroding public trust, as it exacerbates the challenges of accountability, independent oversight and public scrutiny that can be amplified by the use of non-transparent AI systems. An appropriate governance framework should hence enable AI developers and deployers to act responsibly and in compliance with relevant legal requirements, while allowing for proper remedies and intervention by state authorities when this does not happen.

On the other hand, if a particular AI application is proven to be robust when employed in a particular public sector process, and there are no superseding professional, ethical, legal or policy arguments, continuing to prefer the «AI-free» version of the

same process might be just as well considered unreasonable, if the latter process yields more errors.

However, any discussion with respect to these risks would only be helpful in the context of a concrete assessment of concrete applications, both in terms of the amount of potential threat and its probability.

As a matter of policy, it is therefore advisable to subject all AI applications in the public sector to a procedure of impact assessment on the basis of a risk-based approach and from the perspective of human rights, rule of law and democracy, and to repeat such an assessment on a regular basis to assess real impact on users. Adequate oversight safeguards should also be ensured, with a view to fostering compliance with principles of transparency, fairness, responsibility and accountability.

A number of specific, practical recommendations are formulated hereafter with a view to providing appropriate guidance to policy makers in the different stages of design, procurement, the development and the deployment of AI technologies in the public sector, as well as regards continuous education efforts to be deployed.

5. Policy Guidance for public actors seeking to adopt AI

In this section, some concrete policy guidance is provided for public actors when seeking to deploy AI systems, focusing in particular on the measures that can be taken to mitigate potential risks for human rights, democracy and the rule of law. Such measures hinge upon an appropriate underlying regulatory framework that includes clear benchmarks for the design, public procurement, development and deployment of AI-based systems. Importantly, this framework should apply regardless of whether the AI system is developed and deployed by public actors, or whether this occurs by private actors who are acting on behalf of public actors.

5.1 Design Phase

- Analyse the problem that requires a solution

AI systems are tools to enhance human welfare, not ends in themselves. While their use can generate important efficiencies, they are not suitable to solve every problem. It is hence essential to first analyse the specific problem that one intends to solve, and then to establish to which extent AI can provide a solution (and which AI technique would be the most appropriate). Indeed, sometimes other approaches or technologies might be better placed to tackle the specific problem at hand. Moreover, the utility of an AI-based solution often depends on the availability of accurate, reliable and representative data. In some instances, such data may not exist, or its collection might be very costly or not feasible. These potential difficulties should be calculated into the cost-benefit analysis of the potential solutions to the problem.

- Identify which datasets could be used for the AI system

Once the problem has been defined and it has been concluded that AI can contribute towards a solution, the design of the specific AI application can be worked out. The identification of the type of data that needs to be collected and analysed to solve the problem is an important step. Questions of data ownership and data protection will

need to be considered. Moreover, it should be assessed whether the data sets that can be used are sufficiently representative to provide a holistic image of the information that is needed. Often, use will be made of so-called data proxies – namely data points that provide only an indirect parameter of the information that is sought.

- Make the presumptions that underly the design of the system explicit

During the design phase of an AI system, a number of assumptions will be made that often remain implicit, yet which have an influence on the manner in which the AI system will be built and hence on its outcomes. For instance, it is based on the assumptions of the (causes) of the problem, and the nature of the problem, that specific data will be selected (or not), and that a specific solution will be proposed (or not). However, since these assumptions can sometimes turn out to be erroneous or otherwise problematic (for instance because they rely on certain unjust biases), it is important to make them explicit during the design phase and to document them. By making them explicit, assumptions can later be rectified or challenged if they turn out to be wrong, irrelevant or discriminatory.

- Involve system's intended users and consider their capabilities

As the above examples have shown, AI applications in the public sector are very diverse. The users of these applications can be civil servants (e.g. when the system is meant to operate internally to render certain admonistrative processes more efficient), but they can also be citizens (e.g. in case the system is designed to interact directly with individuals, like in the case of chatbots). The design choices around the AI application will need to take into consideration the intended user of the AI system, as well as the users' competences and capabilities.

To ensure the system's user-centricity, it is advisable to already involve the intended users of the AI system in the design phase and gather their input, so that the system can be built in a way that best meets their needs. For example, using open and participatory workshops to connect with frontline workers and discover which aspects of their jobs they would like to spend more time on, and those where AI might usefully step in. By actively listening and seeking to understand the users' needs, and focusing on using AI to alleviate or eliminate frustrating or repetitive tasks, the process not only gains legitimacy with the workforce, but it also becomes easier to discover new efficiencies and increase motivation in the workplace.

- Whenever possible, choose for an open and transparent design

Already from the design phase onwards, it is advisable to choose for a system design that allows for traceability to the extent possible. Especially when the system is meant to assist in public decision-making processes that can have a significant impact on people's lives, there is a strong argument in favour of using transparent techniques rather than black box systems, so as to ensure the possibility of tracing and explaining the decision-making process. In addition, where possible, reliance on open source software used should be open-source. This will not only enable third parties (such as researchers or civil society organisations) to carry out assessments, but it can also further enhance public trust in the system.

- Carry out a human rights, democracy and rule of law impact assessment

An important step to mitigate any potential adverse effects posed by AI systems, concerns the carrying out of an assessment of the system's impact on human rights (HR), democracy (D) and the rule of law (R) (a so-called HRDR impact assesment, or a HRDRIA). Such asssement can help anticipate potential risks, so that where possible these can be already prevented or mitigated in the design of the system. It should be noted that the CAHAI PDG is currently developing a model for such an impact assessment. While this assessment is part of broader human rights due diligence requirements that all organisations have when taking action that could have an adverse impact on human rights, this is especially salient in the context of the public sector, given that public authorities have a direct obligation to protect human rights.

5.2 Procurement Phase

Public procurement could be an important factor for the adoption of AI. Through AI-oriented public procurement schemes, public actors can stimulate the development of innovative AI solutions for the public good, while simultaneously setting steps towards more digitalised public services that put citizens at the centre. Since AI is an emerging technology, it can be difficult to define a route to market for requirements.[23] As many public actors today do not have the necessary know-how, data and competences to develop AI systems themselves, public procurement is one of the main avenues that the public sector currently relies on to deply the technology. However, when AI systems are procured from private actors rather than developed in-house, the standards that governments must meet in order to ensure their AI systems comply with human rights, democracy and the rule of law, should still be complied with. Due to fast technological developments and the lack of uniform standards for AI, existing procurement processes might fall short of ensuring that potential concerns arising from the use of AI in the public sector are duly taken into account. It is, therefore, essential that public actors take the following considerations into account when procuring AI systems.

- Examine applicable legislation and policy measures

Virtually every member state already has existing legislation in place around the public procurement of technology. In addition to legislation dealing with public procurement, there may also be other regulatory requirements or policy documents that apply to the use of AI systems (or technology more broadly) by public authorities. Whenever public actors seek to start a public procurement process for an AI system, it is hence important that applicable legislation and policies in this context are duly mapped, and taken into account for the specific procurement process.

- Adapt public procurement processess where necessary and establish public procurement guidelines for AI

As the Feasibility Study described, AI systems can exacerbate existing risks or pose new ones. Accordingly, public bodies should build further on their existing public

[23] Guidelines for AI Procurement.(2020). Available: https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement , Last accessed : 08 March 2021

procurement policies and adapt them to the specific problems that can be raised by AI. This adaptation can take the form of a change in public procurement legislation, but it can also concern the adoption of guidelines for AI-specific public procurement requirements. Such guidance has for instance been adopted by the UK. AI-specific procurement requirements could be based on socio-technical standards and certifications, and are ideally developed together with affected stakeholder groups. They should ensure conformity with Council of Europe standards on human rights, democracy and the rule of law, and can thereby stimulate the adoption of these standards also by the private companies that participate to AI-related public tender procures.

Besides general AI procurement guidance, the public procurement of the AI system that will be procured, should also take into account the specific risks arising from the AI application in question. As stressed above, not all public AI systems pose the same level of risk. This will depend on factors such as the system's purpose and intended use. Besides the abovementioned HRDRIA, also risk assessment tools can be helpful in this regard.

- Ensure a multidisciplinary and multi-stakeholder approach[24]

When establishing AI-specific public procurement requirements, it is advisable to organise an approach that involves the views from experts of various disciplines, as well as input from different stakeholders. Involving various perspectives and angles increases the chance that any assessment of procurement opportunities and risks is carried out in a more comprehensive manner, and that the potential impact on various stakeholder groups is taken into account.

- Consider the impact on public accountability

Government authorities are responsible for the algorithms they use, even if they are created by third parties, and even if they cannot be explained in detail. During the public procurement process, this accountability should hence be anticipated. Governments should hence ensure that they receive all necessary information about the procured AI system so as to allow them to take responsibility for these systems, and can be held accountable for potential adverse consequences. While public actors can negotiate contractual liability regimes with the organisation from which they procure the AI system, this only affects the relationship between the public actor and the contractant, and may not lead public actors to escape their accountability towards the public at large.

For this reason, the feasibility study already indicated that public authorities should not acquire AI systems from third parties that do not comply with legal information obligations as regards their AI systems, or are unwilling to waive information restrictions (such as confidentiality or trade secrets) where such restrictions impede the process of carrying out human rights impact assessments (including carrying out

[24] as required by the necessity to assess impact on human rights, rule of law and democracy – and also as recommended by Council of Europe Commissioner For Human Rights, Unboxing Artificial Intelligence: 10 steps to protect Human Rights – Recommendation, May 2019, Step 2 (Public consultations), Para 1: "State use of AI systems should be governed by open procurement standards, applied in a transparently run process, in which all relevant stakeholders are invited to provide input."

external research/review and making these assessments available to the public.[25] Indeed, reference can be made to the guidance point formulated above, namely the importance of carrying out a human rights, democracy and rule of law impact assessment of the system in advance. This enables the anticipation of potential risks, which should be taken into account in the procurement requirements.

5.3 Development Phase

Whether AI systems are developed in-house or procured, to truly deliver public benefits they need to meet certain requiremenst to ensure they do not negatively impact human rights, democracy and the rule of law. Given the risks that can arise in the context of the public use of AI systems – and through the use of AI systems more generally – the following guidance should be taken on board during the development phase of AI systems.

- Establish documentation and logging processes

Public actors that develop AI systems (whether in house or outsourced) should establish measures that can ensure the transparency and traceability of the system, such as documention of the development processes followed, selection and curation choices for the data sets used and methods for training and testing the algorithm. Such information should not only be documented, but ideally also be rendered public for those citizens and stakeholders wishing to make use of their access to information rights. By documenting the relevant processes and rendering this information transparent, public trust in the use of AI systems can be augmented.

Insight should also be given into parameterisation and choices pertaining to the training data. Where methods are used that require parameters to be defined in advance or that make use of training data, a description should be provided of the way in which the parameterisation and choice of training data was established, accompanied by an exploration of the potential discriminatory factors. Important questions in this respect include: how was the model developed; which data and algorithms were used; how were they obtained; how were they reviewed internally; and in what format are the results of data analytics published? Besides keeping records of the assumptions that were used, documentation should also be ensured of the analytic method that was used and the way in which its accuracy was measured. Furthermore, records should be kept of the input data to be used (source data/datasets), and of the manner in which the quality of the data source(s) used was verified.
Finally, it should be noted that automated logging processes can be incorporated into the AI system during the design and development phase. These can facilitate record-keeping by ensuring automatic recording of the processes or events taking place in the context of the AI system.

- Put in place adequate test- and validation processes

Testing and validation processes are a key component of an AI system's development cycle. The more automatised the process is, and/or the more authority is being delegated to the AI system, the more emphasis should be put on adequate testing

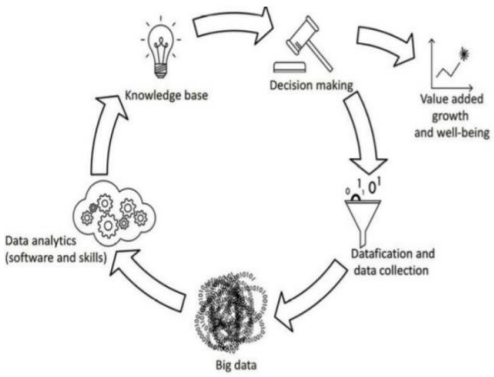
[25] CAHAI(2020)23

methodologies prior to the system's use. This is especially relevant whenever the system might have an impact on individuals. The system should not only be tested on accuracy and robustness, but it must also be established whether the chosen parameters and data sets might adversely and disproportionally affect a given set of the population.

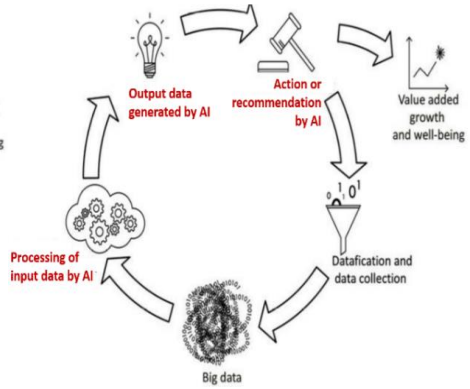
Some specific measures in the system's development can help in this process. For instance, the code is ideally organised into modules that can be evaluated separately and in combination. This can facilitate the testing of these modules for correct functionality, both separately and in combination.

- Set up data governance mechanisms

During the life cycle of AI, which is often primarily based on data, data protection must be guaranteed. Once developed, AI is also an algorithmic system. As such, it is used for processing input data in order to obtain particular output data, such as classification, prediction, or recommendation. While AI is fed by digital records of humans, it can make inferences about individuals' traits and characteristics, including also sensitive information. This can lead to people questioning their trust in the technology, and in the way their data is being processed and used. To secure their legitimate trust, it must be ensured their data is not used in a way that infringes their right to personal data protection, or in a way that infringes other rights, by using adequate technological protection measures and data governance mechanisms[26]. The Council of Europe Convention 108 and its modernised version 108+ provide important personal data protection mechanisms that should be respected in this context.



Data value cycle[27]



Data and AI - as input and output data[28]

Aside from data privacy, the quality of data sets is one of the parameters to develop a high quality AI system. The data gathered should not include biases, inaccuracies and mistakes. The integrity of the data is also important. The process of data access should

[26] COM(2019) 168 final,Available: COM(2019)168/F1 - EN (europa.eu), Last accessed : 07 March 2021.

[27] OECD.(2015). Data-Driven Innovation, Big-Data for Growth and Well-Being.Available :https://read.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en#page4 , Last Accessed:11 March 2021.

[28] OECD.(2015). Data-Driven Innovation, Big-Data for Growth and Well-Being.Available :https://read.oecd-ilibrary.org/science-and-technology/data-driven-innovation_9789264229358-en#page4 , Last Accessed:11 March 2021.

be managed in a proper way and audit mechanisms should be developed to control the process point-to-point. The following can for instance advance the data governance process[29]: Who are the data subjects and what rights do they have? Who is allowed to collect what data, for what purposes, on which legal ground and for how long? How will the data be erased when it is no longer required? Under which circumstances, if any, could the data be re-used, by whom, for which purpose and on which legal fround? Which techniques, if any, are being used to aggregate/anonymise the data?

- Consider the risk of unequal access or treatment

As has been explained in the feasibility study, the use of AI systems risks exacerbating biases and prejudices, whether it concerns sexism, racism, ageism or discrimination based on other unjust discrimination grounds (including discrimination based on proxies or intersectional grounds). This can occur through various causes, such as a biased training data set, a biased design of the algorithm or its optimisation function, or exposure to a biased environment. By the time the system is developed, it is often already too late to counter this risk. It is therefore important to anticipate and mitigate it from the outset, by verifying for instance whether the data set used is sufficiently representative and accurate, whether there is no disproportionate impact on a certain population group, whether the assumptions based on which the system is developed are not based on unjust biases. This is especially important since gaps in representation of vulnerable and marginalised groups in the AI sector might further amplify this risk.

When the AI system's intended user concerns citizens, it is also important to ensure equal access to the AI system, particularly in light of the existing digital divide, or other potential obstacles to accessibility, such as disability, high age, or knowledge gaps.

Importantly, human beings also have biases and prejudices that can negatively affect public decision-making. It can hence be pointed out that AI systems can also be used to try to counter existing human biases in the public sector. However, also when AI is used for this purpose, it remains crucial to take the above guidance into account and ensure transparent procedures to ensure both the trust of the public at large and of the public officials who will be deploying the AI system.

- Assess the impact on gender equality

One of the main causes of gender bias in AI-based tools, stems from biases in the data they are trained on – or the lack of sufficiently representative data. In data used for machine learning, some groups tend to be under or over-represented, thereby (mostly unintentionally/unnoticed) either leaving aside or over-emphasising their presence, background markers or life experience,. Unrefined and uncritical use of historical or contemporary data for training AI can hence cause or perpetuate sex-based discrimination and gender stereotypes (for example concerning gender marketing, job profiling, recruitment tools or image searching).

Sexism and gender stereotypes can also be reproduced and sustained through AI-based tools that possess gender stereotypical characteristics. For instance,

[29] Medhora, P. R.Centre for International Governance Innovation (2018). Data Governance in the Digital Age, Special Report,Available: Data Series Special Reportweb.pdf (cigionline.org) ,Last accessed : 07 March 2021.

most virtual assistants are given women's names and personalities associated with sexist/stereotypical 'women's reactions', whereas some physical robots (e.g. rescue robots) were given men's shapes. There is, however, no reason to associate these systems to either sex and doing so risks reproducing or even exacerbating gender stereotypes. The gender equality aspect can also impact the technological side of AI solutions. Voice, speech and face recognition systems have been found to be performing worse for women than for men, with face recognition often worse for women from some groups. In addition, the lack of gender balance in AI policy/decision- making, development and research – also when it concerns AI in the public sector – futher amplifies the risk that AI systems sustain existing structural gender inequalities.

It is hence important that deveopers of public AI systems are aware of the potential impact on gender equality that their AI system can have, by anticipating and mitigating any such risk in advance. In this regard, addressing the gender data gap is an important step, especially in light of the heavy reliance on data.[30] In addition, improving transparency and raising awareness about the potential biases based on gender (stereotypes) is likewise essential.

5.4 Deployment Phase

Once the public AI system has been developed and duly tested to minimise any adverse effects, the work to ensure the responsible use of AI is not done. Also during the deployment phase, public actors should consider a number of steps that are important both to enhance public trust in AI and to mitigate potential negative effects on human rights, democracy and the rule of law.

- Maintain risk-management and mitigation frameworks

Already during the design and development phase of the AI system, public actors intending to use AI systems should anticipate and address potential risks that the AI system can generate. Risk-management and mitigation is, however, also an essential element that should be in place when the AI system is being deployed. Indeed, risk management concerns an iterative process, which is even more important to maintain when the AI system can further be altered and learn during its deployment. The risk-management process should be documented, and entail at least the following steps: analysis and evaluation of potential (foreseeable) risks, of their likelihood and extent; mechanisms for risk mitigation and control; fall-back plan in case the system can no longer be used; and a communication plan – tailored to various relevant stakeholders – of the risks that can ensue.

- Carry out (independent) audits of the AI system

AI systems that are deployed for public tasks should be regularly audited, and the audit results should be rendered publicly available. Where an AI application can negatively affect human rights, democracy or the rule of law, it is important that such audits can be carried out independently. Importantly, to ensure the possibility of conducting audits, it is imperative that the abovementioned guidance on making the AI system traceable through documentation and logging is followed. The results of the audit

[30] Council of Europe.(2019). Preventing and Combating Sexism. Recommendation CM/Rec(2019)1. Available : 168094d894 (coe.int) . Last Accessed: 10 March 2021.

reports should be made public to enhance public transparency (without, however, affecting legitimate interests such as national security or information protected by intellectual property rights).

Ideally, audits are carried out based on agreed-upon standards or processes that are widely accepted, so that the reports prepared by the auditors – which reflect the audtit results and outcomes – can be compared and evaluated based on common underlying norms. Where such standards do not exist, or are insufficient to meet the specific concerns arising in the context of AI, the development of new standards in cooperation with other countries is advisable.

- Set up a public register for public AI systems

Member states should establish public registers in which they list AI systems used in the public sector. These registers can contain basic information about the system, such as the purpose, technique or type of data used. Making such registers available for the public to consult can increase trust in the use of AI systems in the public sector. Furthermore, such registers can also promote an inclusive and diverse democratic debate around AI systems, and enhance their democratic oversight. While such register could encompass a list of all AI systems used by public actors, it is especially important to make this information available for systems that can have an adverse impact on human rights, democracy and the rule of law.

- Establish feedback mechanisms

Users of AI systems – as well as those affected therbey – should ideally be granted the oportunity to provide feedback on the system's deployment. By establishing a feedback mechanism, input can be collected on ways in which the system can be improved, for instance by making it more user-friendly, increasing its accuracy or mitigating non-anticipated negative effects. The feedback mechanism should ideally be accessible to a wide range of stakeholders, and any follow-p to the feedback received should occur in a transparent manner.

- Ensure the possibility for citizens to interact with a human being

Given the importance of citizens to have access to public services, whenever an AI system is used in this contex, it is improtant tha individuals have the ability to instead inetract with a human interlocutor. This possibility is especially important to provide when an AI-based service does not run adequately, or is not accessible to people certain disabilities or a lack of Access to the system, or when the system's outcome can have a significant impact on the individual.

- Ensure transparency and communication towards users and citizens

Once deployed, AI systems in the public sector should be as transparent as reasonably feasible. Transparency is important to ensure both for the system's direct user (for instance, the public official that relies on the system's output and needs to assess its justifiability), but also for the public at large, which can be affected by the system. This can be done by providing essential information about the system's purpose, capabilities, limitations, underlying logic and assumptions, and potential risks. The information provided should be tailored to the intended audience, and take into acount

the audience's knowledge level, age and other characteristics that are relevant to ensure that the information is communicated in an understandable manner. Where relevant, this also includes existing information obligations in the context of personal data processing, and the specific requirements that should be cosnidered in this respect (for instance, such information should be concise, provided in clear and simple terms, in a transparent, intelligible and easily accessible format).

Whenever an AI-based system is used by the public sector for interaction with individuals in the context of public services, selection processes or identification purposes, either directly or via private sector actors, the users or targeted individuals need to be notified in clear and accessible terms of such use, how the process takes place, how decisions are reached and how they can be reviewed. Reference can also be made to the existing access to documents and information obligations resting upon authorities at national and international level. For example, pursuant to Convention 108+, data subjects have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

Public authorities that use AI systems for public decision-making that can have with specific consequences for individual citizens, should furthermore be able to explain to the citizen both the procedures followed by the algorithm, and the (reasons for the) specific decision that was taken as regards the individual. Indeed, the system's decisions should be explainable to the data subject in the case of any automated decision. This also implies that, as a starting point, government organisations should not, in principle, use algorithms that are too complex to explain.

- Set up accountability and redress mechanisms

Establishing fair and accountable processes and structures helps governments realise the potential of AI to transform public services and administration and build public confidence in their ability to do so. In order for accountability to work effectively, governments should not only make understandable how AI systems are used in decision-making processes, but they should also enable the possibility to assess algorithms, data and design processes, for instance through audits. Furthermore, public actors should seek to identify, assess, document and minimise the potential negative impacts of the AI systems they deploy. To strengthen accountability and oversight, for certain applications that entail a particularly high risk for human rights, democracy and the rule of law, it may be advisable to establish a system of prior authorization from a judicial authority or independent administration before the system can be used.

Accountability also means that a specific actor can be held accountable in case something goes wrong, and that a potentially hamed individual can seek redress. Avenues for redress should be clearly communicated and provided, and the right to an effective remedy against any AI-related harm should be ensured, whereby the public actor deploying the AI system remains accountable. Where necessary, this may also mean that liability regimes may need to be adapted to ensure that individuals can seek compensation in case they sufered unjust harm through a publicly used AI system.

- Ensure the regular (re-)evaluation of the AI system

Assessing the trustworthiness of AI systems, and their alignment with human rights, democracy and the rule of law, is an iterative process. It is therefore important that the AI system's impact assessment is regularly evaluated and updated where needed. The continuous evaluation of the system ideally occurs in a multidisciplinary team, and with the wide involvement of stakeholders. Peridic impact assessments could for instance be rendered public, with the possibility for citizens, çivil society organisations and other interested parties to provide feedback on them.

5.5 Continous Education

The above policy guidance will be difficult to execute without adequate knowledge and understanding of AI, not only in the public sector but also and with the general public[31]. It is therefore essential that the increased use of AI systems in the public sector is accompanied by an increased awareness of AI's benefits, risks, capabilities and limitations for all those who use AI or will be impacted thereby.

- Set up education schemes for public servants

Education on AI and big data should be provided for public officials working in government institutions and administrations, the judiciary and law enforcement, as well as independent oversight bodies and national human rights structures. Given the importance of data for the development and depoyment of AI, public servants should also be educated on what data governance implies, how data quality can be assessed, and how data can be processed in full compliance with privacy and data protection rules. Moreover, by educating and training public officials about the possibilities and limitatiosn of AI and big data, they will not only be able to use AI more responsibily, but they will also be emposered to reflect on which uses of AI could advance or faciliate their tasks in a manner that can best improve public services.

- Set up education schemes for the public at large

Member states should invest in the level of literacy on AI with the general public through robust awareness raising, training, and education efforts, including (in particular) in schools. This should not be limited to education on the workings of AI, but also its potential impact – positive and negative – on human rights. Particular efforts should be made to reach out to marginalised groups, and those that are disadvantaged as regards IT literacy in general[32].

[31] Council of Europe. Commissioner For Human Rights. Unboxing Artificial Intelligence: 10 steps to protect Human Rights-Recommendation. Available : https://www.coe.int/en/web/commissioner/-/unboxing-artificial-intelligence-10-steps-to-protect-human-rights . Last Accessed: 11 March 2021

[32] Council of Europe. Commissioner For Human Rights. Unboxing Artificial Intelligence: 10 steps to protect Human Rights-Recommendation. Available : https://www.coe.int/en/web/commissioner/-/unboxing-artificial-intelligence-10-steps-to-protect-human-rights . Last Accessed: 11 March 2021

6. Conclusions

Through the analysis of different sources and of examples from some Council of Europe's member states, this paper has shown that the exploration of the use of AI in the public sector in very different fields is a growing phenomenon, in the light of the possible benefits expected from the use of this technology for individuals and the public administrations. AI systems are expected to contribute to the rationalisation of public sector operating methods and to greater effectiveness of administrative action, relieving resource-constrained organisations from mundane and repetitive tasks and paving the way to pro-active public service delivery models. However, the possible benefits associated with the use of such technologies must be balanced by an appropriate assessment of the risks of negative impacts on human rights, democracy and the rule of law – risks already noted in the CAHAI feasibility study - for a sector that plays a key role in protecting these values within member States.

This report has provided further guidance to the Member States in order to adequately address these risks, by formulating recommendations concerning the measures that should accompany the design, procurement, the development and the deployment of AI technologies in the public sector, as well as the efforts to be deployed when it comes to continuous education, with a view to making the public sector more accountable, transparent and respectful of European standards on human rigths, rule of law and democracy.