COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Strasbourg, 11 March 2021

CAHAI-PDG(2021)03
Provisional

AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI)

POLICY DEVELOPMENT GROUP

(CAHAI-PDG)

Artificial Intelligence in Public Sector

Draft prepared by Sub-Working Group 2[1]

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

www.coe.int/cahai

Executive summary: Policy guidance - recommendations

- Public authorities which use algorithmic decision-making with specific consequences for individual citizens should be able to explain both the procedures followed by the algorithm and the specific decisions taken
- In public organisations explainability is key as decisions made based on AI should be fully understood and explainable for reasons of accountability.
- Public authorities should be able to provide accountability for their decisions
- Government should develop open and transparent accountability structures.
- Governments should make understandable how AI systems are used in decision-making processes.
- AI in the public sector should be transparent in the most possible way.
- Regulatory frameworks should include clear transparency, benchmarks for the public procurement, design, development and use of AI-based systems by them or for them by private actors.
- Member states should establish public registers to account for high-risk AI-based systems.
- Governments should establish measures that can ensure traceability, such as documenting methods for training the algorithm
- Regulatory frameworks should include clear transparency benchmarks for the public procurement, design, development and use of AI-based systems
- Datasets and decision-making processes should be documented to the extent possible to allow for transparency.
- AI systems impact lives should be able to demand an explanation of the decision-making process.
- Governments should establish measures that can ensure traceability, such as documenting methods for training the algorithm
- Governments should clearly communicate characteristics, limitations and potential shortcomings of the AI system
- Governments should develop a trustworthy, fair, and accountable approach to using AI, such as through establishing ethical frameworks,
- As with all policymaking and public service design, government AI interventions should be shaped by a constantly evolving understanding of the needs of the end users
- Governments should identify relevant enduser groups.
- The knowledge and understanding of AI should be promoted in government institutions, independent oversight bodies, national human rights structures, the judiciary and law enforcement, and with the general public.
- Governments 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
- Governments should take adequate measures to counter the use or misuse of AI systems for unlawful interference in electoral processes, for personalised political targeting without adequate transparency, responsibility and accountability mechanisms
- Political microtargeting should not exclude vulnerable or minority groups

1. Definition of AI

An international, commonly agreed definition of artificial intelligence (AI) does not exist. Following the feasibility study the term covers a wide variety of sciences, theories and techniques of which the aim is to have a machine reproduce the cognitive capacities of a human being (see Feasibility study, p 3) AI also includes different types of automated learning.

Countries around the world and international organisations such as the European Union, UNESCO or OECD have understood the tremendous economic potential of AI, which is considered as a strategic technology. The technology is fast evolving which makes it hard to assess its impact or to develop a common coordinated approach.

2. Use Cases of AI in Government

2.1. Overview of National AI Strategies

Council of Europe Member States are launching national AI strategies or similar initiatives to lay out their approach to the development and use of artificial intelligence, with a view to fully harness its benefits.

Most national strategies address the use 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. In fact, as mentioned above, 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).

AI is discussed in many strategic documents as part of technological progress. Governments generally have plans for investing into AI and they have responisbility for it. In Table.1 shows the countries which dedicate a section of national strategic documents to AI or have AI-specific strategic documents[2].

	EARLY PROJECTS	STRATEGY / SET AGENDA	GUIDELINES	PILOT ACTIVITIES AND TRIALS (TOP-DOWN)	WIDESPREAD ADOPTION
Countries	UK	Austria, Belgium, Finland, France, Germany, Hungary, Ireland, Italy, Lithuania, Malta, Netherlands, Portugal, Slovenia, Spain, Switzerland, UK	Austria, Sweden, UK	Ireland, Latvia, Italy, Lithuania, Portugal, UK	Finland, Germany
Total	**1**	**16**	**3**	**6**	**2**

[2] Exploring Digital Government Transformation in the EU, JRC Science for Policy Report, Exploring Digital Government transformation in the EU | EU Science Hub (europa.eu)

Table.1 Overview of Countries' AI perspective dating from 2019, NL must be added for pilot activities)

Member States recognise the fact that they need to invest in capacity building of civil servants and public sector officials. Some national strategies explicitly address "up-skilling" as an issue.

Access to more and better data is 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.

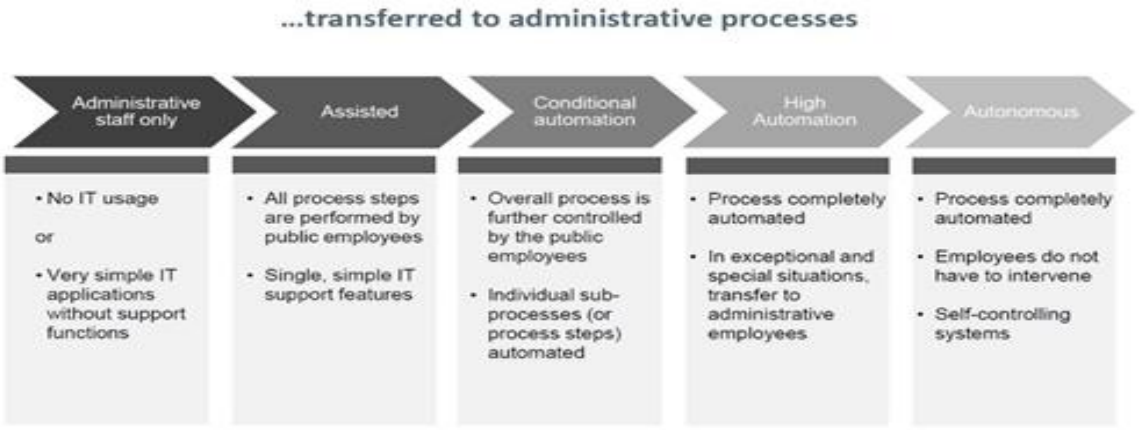
While most Member States mention an ethical framework, some also specifically mention the need to regulate AI and see the public sector in the regulatory role. As mentioned above, at the Council of Europe the Ad hoc Committee on Artificial Intelligence (CAHAI) is currently conducting a feasibility study regarding whether and how AI can be regulated.

In 2020, AI Watch (the European Commission knowledge service to monitor the development, uptake and impact of Artificial Intelligence for Europe) also published a report mapping the use of artificial intelligence in public services in EU Member States.

The study concludes that *"governments across the EU are exploring the potential of AI use to improve policy design and evaluation, while reorganising the internal management of public administrations at all levels. Indeed, when used in a responsible way, the combination of new, large data sources with advanced machine learning algorithms could radically improve the operating methods of the public sector, thus paving the way to pro-active public service delivery models and relieving resource constrained organisations from mundane and repetitive tasks".*

Furthermore, the authors stress that: *"There is a high expectation from the use of AI in government – but it is clear from our current exploration that positive impact is far from straightforward and should not be taken for granted."*

As discussed above, AI-enabled tools are increasingly being used by the public sector. Much of the debate centres around the use of algorithmic or automated decision making (ADM) systems.



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

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.

The legal bases for using ADM may vary in Member States. Examples are Germany and Netherlands.. According to German law, 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, the situation is as follows: the principle of legality demands a basis in the law for decision making (with legal consequences/when fundamental rights are in play), irrrespective of information systems are used or not. Of course, the processing of personal data is governed bij the GDPR and other privacy/data protection laws.

2.2.Types of AI Applications could be used in government

2.2.1. Natural Language Processing (NLP)

The field of NLP is also called computational linguistics and presents solutions in understanding human languages through computational models and processes[3].

2.2.2. Speech Recognition

Speech Recognition enables a computer to identify the words that a person speaks into a microphone or telephone and convert them into written text[4].

2.2.3. Computer Vision

[3] Otter,D.,W., Medina,J. ,R., Kalita,J.,K.(2021). A Survey of the Usages of Deep Learning for Natural Language Processing, IEEE Transactions on Neural Networks and Learning Systems, Vol.32,No.2. Available : IEEE Xplore Full-Text PDF: . Last Accessed : 09 March 2021

[4] Vimala,C.,Radha,V.(2012). A Review on Speech Recognition Challenges and Approaches. World of Computer Science and Information Technology Journal (WCSIT),Vol. 2, No. 1, 1-7,ISSN: 2221-0741. Available:A Review on Speech Recognition Challenges and Approaches (wcsit.org). Last Accessed : 09 March 2021.

AI applications from this category use some form of image, video or facial recognition to gain information on the external environment and/or the identity of specific persons or objects[5].

2.2.4. Machine Translation

Machine translation is a sub-field of computational linguistics that focuses on the use of software to translate text or speech from one language to another[6].

2.2.5. Robotics

Robotics is an interdisciplinary field integrating mechanical engineering, electrical engineering, information engineering, mechatronics, electronics, bioengineering, computer engineering, control engineering, software engineering, and that includes the designing, construction, operation, and use of robots[7].

2.2.6. Rules-based systems

Rule-based systems (also known as production systems or expert systems) are the simplest form of artificial intelligence. A rule-based system is a way of encoding a human expert's knowledge in a fairly narrow area into an automated system[8].

2.2.7. Machine Learning

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the systems which can learn from data, identify patterns and make decisions with minimal human intervention[9].

2.3. AI Applications to use in the public sector

Compared to the private sector, one might argue that the public sector is currently behind in adopting AI. However, governments are seeking to catch up and close the gap. According to OECD's initial mapping on AI, OECD identified 50 countries (including EU) that have partaken in introducing national AI strategies, 36 of these countries 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

[5] European Commission.(2020). AI Watch: Artificial Intelligence in Public Services.Available: AI Watch - Artificial Intelligence in public services | EU Science Hub (europa.eu). Last accessed: 09 March 2021.

[6] Wikipedia. Machine translation. Available : https://en.wikipedia.org/wiki/Machine_translation . Last Accessed:09 March 2021.

[7] Wikipedia.Robotics. Available: https://en.wikipedia.org/wiki/Robotics#cite_note-1 . Last Accessed : 09 March 2021

[8]Grosan C., Abraham A. (2011) Rule-Based Expert Systems. In: Intelligent Systems. Intelligent Systems Reference Library, vol 17. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21004-4_7

[9] SAS, Available : https://www.sas.com/en_us/insights/analytics/machine-learning.html#:~:text=Machine%20learning%20is%20a%20method,decisions%20with%20minimal%20human%20intervention. Last Accessed : 11 March 2021

citizens and business, or solve specific problems in critical fields such as health, transportation and security[10].

3. AI Transformation in Public Sector

3.1. Benefits/Opportunities

Artificial Intelligence solutions in public services provide public organisations making more efficient by improving job satisfaction of public servants while increasing the quality of services offered. To give examples for the applications of AI in the public sector ; AI could assist citizens by answering their frequently asked questions through chatbots or virtual assistants or AI could help public servants by making welfare payments and immigration decisions, detecting fraud, planning new infrastructure projects..etc. The point where technology has come allows to automate many services of bureaucracy. According to statistics of Statista ; it is shown the adoption of artificial intelligence in the public sector by comparing with other organizations, by sector (Please see Fig.1).

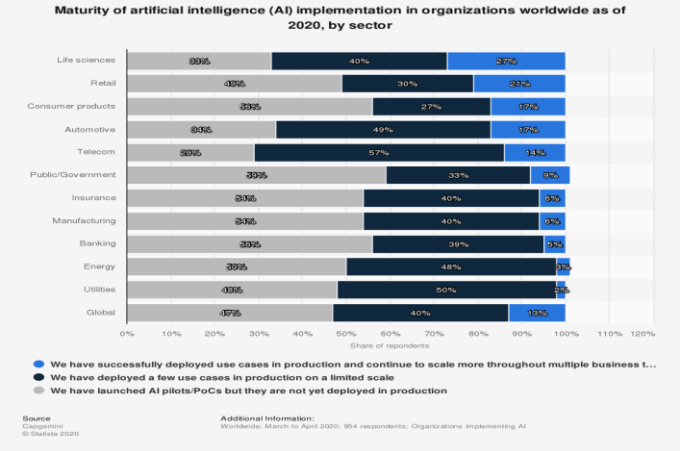


Fig.1. Source (• Maturity of AI implementation by sector worldwide 2020 | Statista)

3.2. Challenges/Risks

While there are many potential benefits of AI , risks must be governed by respecting democratic values and human rights. Considering how AI technologies have an impact on societies, It gains more importance to base AI on human dignity and privacy concerns[11].

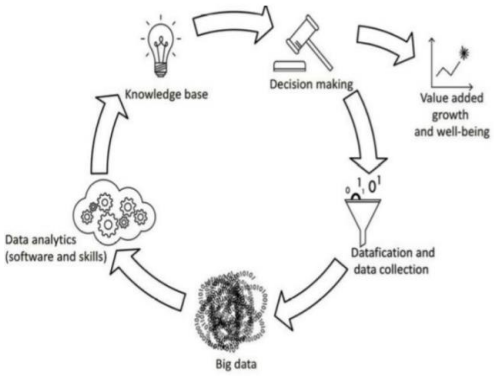
3.2.1. Data Governance

During the life cycle of AI which is based on data, it must be guaranteed data protection. Once developed and deployed, AI is also an algorithmic system. As such, it is used for

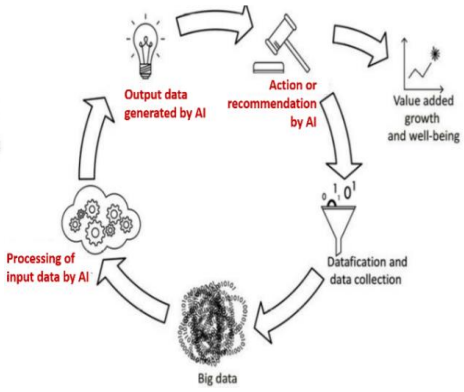
[10] https://www.oecd-ilibrary.org/docserver/726fd39d-en.pdf?expires=1614757709&id=id&accname=guest&checksum=4E57037FE90FC8105CF2F3ACC49AF395

[11] On Artificial Intelligence - A European approach to excellence and trust,European Commission(2020), Available: commission-white-paper-artificial-intelligence-feb2020_en.pdf (europa.eu) ,Last accessed : 07 March 2021

processing input data in order to obtain particular output data, such as a classification, prediction or recommendation[12]. While AI is fed by digital records of humans, It can make inferences about individuals' sexual orientation, health condition, religious or political views.



Data value cycle[13]



Data and AI - as input and output data[14]

This can lead to people questioning their trust in the processing of their data. To prevent this, It must be ensured their data not be used to harm their privacy with the help of technological measure[15]. 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 be managed in a proper way and should be developed audit mechanisms to control the process point-to-point. Data governance is built with the following questions[16]:

- ➢ Who owns the data and what do these data rights entail?
- ➢ Who is allowed to collect what data?
- ➢ What are the rules for data aggregation?
- ➢ What are the rules for data rights transfer?

[12] GPAI.(2020). Data Governance Working Group A Framework Paper for GPAI's work on Data Governance. Available :https://gpai.ai/projects/data-governance/gpai-data-governance-wg-report-november-2020.pdf . Last Accessed:11 March 2021.

[13] 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.

[14] 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.

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

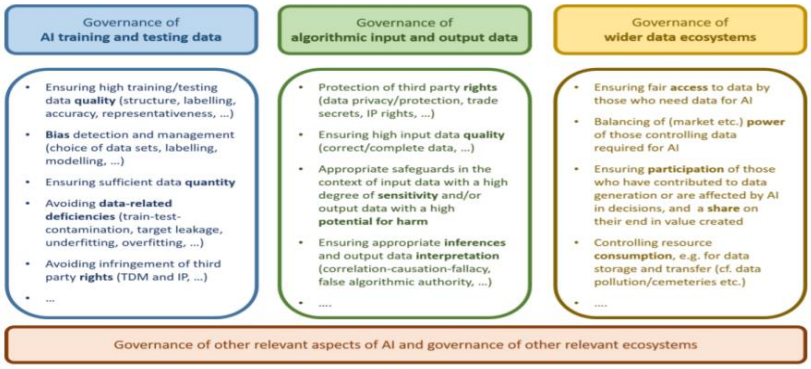
[16] Medhora, P. R.Centre for International Governance Innovation (2018). Data Governance in the Digital Age, Special Report,Available: Data Series Special Reportweb.pdf (cigionline.org) ,Last accessed : 07 March 2021.

Terminology[17]

- *'data'* means any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audiovisual recording;
- *'re-use'* means the use by natural or legal persons of data held by public sector bodies, for commercial or non-commercial purposes other than the initial purpose within the public task for which the data were produced, except for the exchange of data between public sector bodies purely in pursuit of their public tasks;
- *'non-personal data'* means data other than personal data as defined in point (1) of Article 4 of Regulation (EU) 2016/679;
- *'data sharing'* means the provision by a data holder of data to a data user for the purpose of joint or individual use of

Since the data is one of the important parameters to develop AI technology efficiently, It should be used by the developers and researchers with the limitation (, the GDPR states the principles of purpose limitation, data minimisation and storage limitation). In that case, governing the data which provides re-using it gains importance. The data needs protection on the grounds of commercial confidentiality, statistical confidentiality, protection of personal data or protection of intellectual property rights of third parties[18].

A framework for data governance should provide the sensitive data to be analyzed privacy-friendly through privacy-enhancing approaches such as anonymisation, pseudonymisation, differential privacy, generalisation, or suppression and randomisation. Development of audit mechanisms will be useful for the sustainability of data governance.



The need for data governance in the AI context[19]

[17] COM(2020) 767 final.

[18] European Comission, Regulation of The European Parliament and of the Council on European Data Governance(Data Governance Act).COM(2020) 767 final. Available : https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0767&from=EN . Last Accessed:10 March 2021.

[19] GPAI.(2020). Data Governance Working Group A Framework Paper for GPAI's work on Data Governance. Available :https://gpai.ai/projects/data-governance/gpai-data-governance-wg-report-november-2020.pdf . Last Accessed:11 March 2021.

The biased data sets can lead to discriminatory outputs by AI. Good data governance should ensure that the data used is appropriate for the intended purpose.

Effective governance measures should include secure data spaces having transparent conditions.

The structure of data governance should be risk-based including on society and the economy.

The way of data governance should be friendly in terms of inclusiveness and diversity[20][21][22].

- Parameters for Data Governance should include[23] ;
  - Categories of data,
  - Data ecosystems,
  - Rights with regard to Data.

3.2.2. Procurement

Public procurement could be an important factor for the adoption of AI. In case not to build the AI system in house, It should be contemplated how to procure the technology. Since AI is an emerging technology, it can be difficult to define a route to market for requirements[24]. To manage the procurement process healthy, steps to follow should be defined. Due to the fast technological developments and lack of the standards in the AI area, it can not be efficient to use existing procurement processes.

AI procurement framework should include :

- references to existing legislation and policy measures of other technologies' procurement processes,
- multidisciplinary approach,
- Multi-stakeholder approach[25],

[20]The Rockefeller Foundation, Announcing the Lacuna Fund: Closing Data Gaps to Enable Equitable Machine Learning, Available : https://www.rockefellerfoundation.org/blog/announcing-the-lacuna-fund-closing-data-gaps-to-enable-equitable-machine-learning/ , Last Accessed : 11 March 2021.

[21] European Comission, Ethics guidelines for trustworthy AI, Available : https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai , Last Accessed:11 March 2021.

[22] Diversity.AI. A think tank for inclusive Artificial Intelligence, Available : http://diversity.ai/#firstpage , Last Accessed: 11 March 2021.

[23] GPAI.(2020). Data Governance Working Group A Framework Paper for GPAI's work on Data Governance. Available :https://gpai.ai/projects/data-governance/gpai-data-governance-wg-report-november-2020.pdf . Last Accessed:11 March 2021.

[24] 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

[25] 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."

- ❖ a sustainable language which provides investing in building responsible practices for how they procure AI[26],
- ❖ impact assessments on human rights, rule of law and democracy[27]
- ❖ 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 (e.g., confidentiality or trade secrets) where such restrictions impede the process of[28]
  - ➢ carrying out human rights impact assessments (including carrying out external research/review and
  - ➢ making these assessments available to the public.

Existing applications used in public sector and examples for risk assessment tools[29] :

Applications

Spam-filters in email programs – detect and block unwanted emails.

AI in cybersecurity solutions – protect networks, programs, and data from attack, damage, or unauthorized access.

Chatbots – converse with people via voice interfaces or text messages.

Fraud detection – detects, prevents and manages fraudulent patterns in the data.

AI in policing or social services – support and/or drive decisions in fields such as law enforcement, crime prevention, public safety, children welfare, social programs.

AI in HR – takes on key HR tasks including hiring, retaining talent, training, benefits and employee satisfaction.

Tools:

AI risk assessment tool – helps deciding on a proportional approach to AI procurement. It sets out examples for decision criteria to include in a risk assessment of any potential solutions that contain AI capabilities. The tool outlines some of key questions you should consider when deciding your procurement strategy, considering what questions to ask in your RFP and assessing a solution.

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. The ethical platform includes; a list of values that orient you in deliberating about the ethical permissibility and impact of a prospective AI project; a set of principles that all members of your project delivery

[26]World Economic Forum.(2020). AI Procurement in a Box: AI Government Procurement Guidelines, Available: gx-wef-ai-government-procurement-guidelines-2020.pdf (deloitte.com), Last Accessed : 09 March 2021
[27] CAHAI(2020)23
[28] CAHAI(2020)23
[29] World Economic Forum.(2020). AI Procurement in a Box: AI Government Procurement Guidelines, Available: gx-wef-ai-government-procurement-guidelines-2020.pdf (deloitte.com), Last Accessed : 09 March 2021

team should be well-acquainted with and a framework that operationalizes these values and principles in an end-to-end workflow governance model.

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.

3.2.3. Employment

One of the concerns brought by artificial intelligence is losing jobs for people. According to a Eurobarometer survey published by the European Commission, 72% of respondents believe robots steal the jobs of people[30]. To reduce these worries, awareness of AI technologies should be stimulated among civil servants. So, they can discover the advantages of AI approaches for their work. There are a number way to raise awareness such as organizing regular meetings between civil servants either in institutions or in specialized innovation hubs or by participating in policy events organised by European institutions or other relevant parties[31].

3.2.4. AI Biases/gender equality-discrimination-racism

While further attention should be given to AI biases in relation to gender equality, discrimination and racism, And as the use of AI technologies increases, evidence and existing studies show that biases exist at different levels and in different sectors.

AI-based tools can inherit gender bias from the data they train on. 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, life experience, etc. Unrefined and uncritical use of historical or contemporary data for teaching AI can 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 (ex: most virtual assistants are given women's names and personalities associated with sexist/stereotypical „women's reactions" -

[30] European Commission(2017). Attitudes Towards the Impact of Digitisation and Automation on Daily Life, Available: https://ec.europa.eu/digital-single-market/en/news/attitudes-towards-impact-digitisation-and-automation-daily-life , Last accessed : 07 March 2021

[31] European Comission.(2020). AI Watch Artificial Intelligence in public services. Overview of the use and impact of AI in public services in the EU. Available : jrc120399_Misuraca-AI-Watch_Public-Services_30062020_DEF_0.pdf (europa.eu). Last Accessed : 10 March 2021.

some physical robots ex rescue robots) were given men's shapes. There is no reason to associate these systems to either sex and doing so reproduces 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 creates the risk that AI systems sustain existing structural gender inequalities. Women make up only 21.5% of the digital workforce in the EU and are also under-represented in ICT leadership.

Since the AI uses the data collected from objects, social networks or other connected sources; It hardly includes information about women. Since AI for everyone, the same goes for gender equality as well. The AI regulatory framework is recent and deserves to be completed and communicated widely. It should be subject to enforcement monitoring measures. Each stakeholder in the AI system must feel concerned by the subject and know their duties, rights, action or reaction plan. It is also important that all these stakeholders work together: companies with researchers, and with users, and with suppliers[32].

A gender equality perspective should be integrated in all policies, programmes and research in relation to artificial intelligence to avoid the potential risks of technology perpetuating sexism and gender stereotypes and examine how artificial intelligence could help to close gender gaps and eliminate sexism[33]. This includes:

- addressing gender imbalances in AI policy/decision- making, development and research;
- factoring in gender-based dynamics when designing data-driven instruments and algorithms;
- improving transparency and raising awareness about the potential biases based on gender and other factors in big data and the reproduction of sexism and gender stereotypes; and
- improving monitoring & accountability.

3.2.5. Explainability

In order to be able to effectively monitor the government's decision-making on the basis of data analytics, the relevant supervisory authorities must have sufficient insight into data processing processes. It is important here to distinguish between 'technical transparency' and 'explainability.

What happens during the training of neural networks is one of the main challenges of AI. AI in its various types is often known as a system acting black-box manner, since it is not understood why the algorithms make specific decisions while producing the results. Explainability could be a more important notion in public sector use cases because of accountability and transparency. In this manner, what is expected from an explainable AI is whose actions can be understood by humans. Explainability brings two notions with itself:

[32] Laboratoire de l'Egalité.An Egalitarian Artificial Intelligence Between Women and Men. Available: 1680a0577f (coe.int). Last Accessed : 10 March 2021.
[33] Council of Europe.(2019). Preventing and Combating Sexism. Recommendation CM/Rec(2019)1. Available : 168094d894 (coe.int) . Last Accessed: 10 March 2021.

interpretability and transparency. Principles of these notions are important in the content of AI, especially when the system is human-centered. Even though there are studies in the academic literature, there is no exact consensus regarding explainability yet[34]. But as long as AI services for humans, It should be defined with required parameters to provide explainability.

As seen, data protection law, may require models to be explainable in the sense of offering an explanation for an automatically made decision. These explanations ensure that algorithms can be introspected to check whether the decisions they made are fair and unbiased.[35] All in all, while interpretable machine learning certainly still has a long way to go, the advent of explainability requirements will likely inject even more force into these models. Their development and deployment will likely benefit both decision makers and data subjects in the long run.

In cases where 'technical transparency' does not always achieve the desired degree of clarity, explainability can play an important role. Explainability is about explaining the use, the purpose and the outcome of an algorithm in terms that people can understand.

How algorithms are used

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':

1. Descriptive – 'What's happening?'
2. Diagnostic – 'Why is it happening?'
3. Predictive – 'What will happen?'
4. Prescriptive – 'What needs to be done?'

The algorithm's explainability is about being able to explain the outcome of the algorithm in clear, understandable language. Technical transparency can contribute to explainability to a certain extent, but has its limits when dealing with very complex algorithms. For inherently opaque models, however, techniques have been developed or are being developed to determine afterwards which information an algorithm used as the basis for its outcome. Against this background, explainability of algorithms focuses on describing the purpose for which an algorithm is used, which variables are decisive factors in the outcome, the type of data used (including data quality and how the data is combined), and any decision-making rules involved.[36] As a result, explainability is much more meaningful than technical transparency in many cases.

Public authorities that use algorithmic decision-making with specific consequences for individual citizens must be able to explain both the procedures followed by the algorithm and

[34] Sokol,K.,Flach,P. (2020). Explainability Fact Sheets: A Framework for Systematic Assessment of Explainable Approaches.Fairness, Accountability, and Transparency (FAT* '20), January 27–30, 2020, Barcelona, Spain. ACM, New York, NY, USA, 22 pages. https://doi.org/10. 1145/3351095.3372870 . Available: https://arxiv.org/abs/1912.05100 , Last Accessed : 08 March 2021

[35] Hacker, Philipp, Ralf Krestel, Stefan Grundmann, and Felix Naumann. "Explainable AI under Contract and Tort Law: Legal Incentives and Technical Challenges." *Artificial Intelligence and Law* 28, no. 4 (December 2020) p. 16

[36] Parliamentary Papers, House of Representatives, 2018/19, 26 643, no. 570, p. 4.

the specific decisions taken. This implies that, as a starting point, government organisations should not, in principle, use algorithms that are too complex to explain.

Insight into parameterisation and choices pertaining 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?

These specific measures are relevant here:

- Organise the code into modules that can be evaluated separately and in combination.
- Test these modules for correct functionality, both separately and in combination.
- Explain the analytic method used and measure its accuracy.
- Keep and retain records of the input data to be used (source data/datasets) and use relevant data only.
- Describe and check the quality of the data source(s) used.
- Keep records of the assumptions used.

3.2.6. Accountability

In public organisations, explainability is key as decisions made based on AI must fully be understood and explainable for reasons of accountability.

Moreover, decisions made by public stakeholders may have a strong impact on the lives of citizens.

States should develop open and transparent accountability structures.

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 make understandable how AI systems are used in decision-making processes. o It should be possible to assess algorithms, data and design processes. Organisations should seek to identify, assess, document and minimise the potential negative impacts of AI systems.

- Methods should be put in place to negotiate, evaluate and document
- Mechanisms should be in place that ensure individuals have the right to redress when an unjust adverse impact occurs.

Some questions:

Where application affects fundamental rights can the AI system be audited independently?

Are processes for third parties or workers to report potential vulnerabilities, risks or biases established?

Did the government establish an adequate set of mechanisms that allows for redress in case of the occurrence of any harm or adverse impact?

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.

Government authorities are responsible for decisions made by the algorithms they use, even if the algorithms are created by third parties, and even if they cannot be explained in detail. Those authorities should therefore be able to provide accountability for those decisions.

3.2.7 Transparency

AI in the public sector should be transparent in the most possible way. Algorithms are, due to their technical nature, not always transparent. Sometimes they are opaque due to their complexity. In other cases, the way an algorithm works is deliberately not revealed in an attempt to keep people from 'gaming the system', causing the algorithm to become ineffective. A lack of transparency in how algorithms work may also be caused by the creators' desire to protect their proprietary interests, or by the incredible complexity of algorithms, which can be extremely dynamic and unfathomable by nature. As an algorithm becomes more complex, 'technical transparency' (see below) provides less and less understanding of the algorithm.

- ❖ Regulatory frameworks should include clear transparency benchmarks for the public procurement, design, development and use of AI-based systems by them or for them by private actors.
- ❖ Member states should establish public registers to account for high-risk AI-based systems that are being procured, designed, developed or used by/for the public sector both at national and local level. Such registers should provide information on the AI-based systems' scope, their way of functioning, the datasets used for their training, the entities responsible for their development, the results of the regular human rights impact assessments conducted and the stakeholders involved in such assessments.
- ❖ Governments should establish measures that can ensure traceability, such as documenting methods for training the algorithm

- ❖ The general information provided to the public on data analytics should meet a number of requirements as far as the processing of personal data is concerned[37]:
    - ➢ It must be concise and transparent, comprehensible and easily accessible.
    - ➢ Including the information in a privacy statement on the organisation's website with a clear link on the homepage can contribute to easy access.
    - ➢ The information should also be provided in clear and simple terms, especially if it is intended for children.
    - ➢ It should not leave any room for different interpretations and should in any case clarify the purpose and legal basis of the data analytics.

3.2.8. Liability

Artificial intelligence has the potential to transform our societies and economies for the better as other emerging digital Technologies while one of the main concerns regarding the use of it is the so-called black box effect. The algorithm runs through the data and leads to a result. However, often neither the programmers nor the public officials can explain how or why the algorithm came up with this particular result. The reasoning and decision-making happen in a black box. The specific characteristics of these technologies are complexity, modification through updates or self-learning during operation, limited predictability, and vulnerability to cybersecurity threats. In order to rise to the challenges emerging digital technologies bring with them, liability regimes should be designed[38].

3.2.9 Trustwothy & Legitimacy

Develop a trustworthy, fair, and accountable approach to using AI, such as through establishing ethical frameworks, clarifying existing binding norms and - where useful and necessary – creating new binding norms. Also shoul be maintained a focus on individuals who may be affected, clarifying the role of humans in AI-driven processes, pursuing the explainability of AI outcomes and developing open accountability structures.

As with all policymaking and public service design, government AI interventions should be shaped by a constantly evolving understanding of the needs of the end users, who might be frontline workers or citizens themselves. Governments should identify relevant enduser groups.

[37] See also article 12 (1) of the GDPR and the Guidelines on Transparency under Regulation 2016/679 (GDPR) of the Article 29 Working Party, paragraphs 6 to 18.

[38] European Comission.(2019). Liability for Artificial Intelligence and Other Emerging Digital Technologies. Available : index.cfm (europa.eu). Last Accessed:12 March 2021

Involving users

Involving codesigners from the very beginning of the process will help to identify the building block tasks effectively. 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 building active listening and understanding of user needs into the heart of design and focusing on using AI to alleviate or eliminate frustrating or repetitive tasks, the process not only gains legitimacy with the workforce, it becomes easier to discover new efficiencies and increase motivation in the workplace.

3.2.10 Efficiency

The main motivation for digitalisation in the public sector is to increase efficiency and thus reduce costs. In addition, it is believed that digitalisation will free public official from routine activities that can best be automated, thus potentially increasing the quality of service delivery.

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.

Establishment of public registers for ADM systems used within the public sector; to develop and establish approaches to effectively audit algorithmic systems; and to promote an inclusive and diverse democratic debate around ADM systems. The question of democratic oversight over AI systems remains pertinent and, so far, no independent bodies or processes exist.

To better ensure that automated decision making (ADM) systems currently deployed and those about to be implemented throughout Europe are consistent with human rights and democracy Impact on democracy

3.2.11 Impact of democracy

Use of AI enabled systems might have multiple impacts, depending also on the intent of the producers, providers and users.

By amplifying the capacity of ordinary people to access, share and report information, digital transformation can contribute to the democratic oversight of public institutions and strengthen their accountability.

The impact of artificial intelligence on participatory tools is not clear and might vary greatly, depending to a large extent on who is using the AI systems and for which purposes.

The use of AI enabled technologies in participatory tools raises questions of transparency (are people aware that an AI system is being used; do people know who is behind the AI system; are people aware what data goes into the system and what algorithms are used?) and accountability (who is held to account in case of false results, data breaches or misuse of data?).

Special care needs to be taken to ensure that democratic principles are not undermined, and that participation is enabled and equal for all (and thus issues are not co-opted by vocal and digital-savvy interest groups). This would require for example addressing the exisiting gaps in relation to access to, use and knowledge of digital technologies based notably on age, gender and social status. In general, digital tools for participation, including AI systems, must avoid creating new barriers. At the same time, the tools are vulnerable to misuse and manipulation. Therefore, measures must be taken to minimize these risks with full respect to the demands of data-protection and the right to privacy as well as transparency and accountability.

The impacts on elections are important. Free and fair elections are the cornerstone of representative democracy. They require independent public opinion formation is vital. Digital technologies including artificial intelligence and machine-learning algorithms form an integral and important part of the information eco-system that voters rely on. Digital technologies have reshaped the ways in which people express their will through votes and representation and they have to a large extent changed political campaigning.

Digital technologies, including artificial intelligence and machine-learning algorithms, have been used by various actors to influence democratic processes and outcomes. For instance, the recourse to *bots* and *trolls* during electoral campaigns with a view to manipulating voter behaviour has received great attention by media, governments and organisations working in the area of democracy, including the Council of Europe.

Democracy, governance and digital transformation are fast-moving processes that require continuous vigilance to identify and address emerging risks. They also require continuous investment, adjustment and the introduction of enabling measures to allow individuals, democratic actors and society as a whole to fully benefit from them.

In these times, the 12 Principles of Good Democratic Governance from the Council of Europe will continue to be a very important compass to steer public action at all levels of government. This report has explored the impact of digital transformation on democracy and good governance in the light of the 12 Principles, highlighting risks, opportunities, mitigating and enabling factors, sometimes providing examples of policies draw from the experience of Council of Europe member States. Some important questions, however, are left outside the scope of this report. They revolve around data: who owns it, who uses it, who and how limits its ownership and use and, above all, how to ensure that data is used for the public good. The answers to these questions will determine the evolution of democracy in the foreseeable future.

Most national strategies address the use 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; in fact, as mentioned above, 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).

Member States recognise the fact that they need to invest in capacity building of civil servants and public sector officials. Some national strategies explicitly address “up-

skilling" as an issue. Furthermore, investment in the education sector to ensure that qualified workforce will be available in the future.

Member States should take adequate measures to counter the use or misuse of AI systems for unlawful interference in electoral processes, for personalised political targeting without adequate transparency, responsibility and accountability mechanisms, or more generally for shaping voters' political behaviours or to manipulate public opinion in a manner that can breach legal norms safeguarding human rights, democracy and the rule of law. (CAHAI study, page 40)

Member States should adopt strategies and put in place measures for fighting disinformation and identifying online hate speech to ensure fair informational plurality.

3.2.12 Information for citizens about AI systems

Digital transformation affects all levels of the public sphere. Individuals spend an increasing time of their lives online: to socialise, work, have entertainment and gather information. The way in which people decide to vote is increasingly influenced by the information that reaches them online, which is often targeted to them individually and/or can be manipulated by malicious actors.

Influencing the ability of individual voters to make a well-informed political decision on such a large scale as it is made possible by the joint use of algorithms and social media can affect the integrity of the electoral process and ultimately the legitimacy of representative institutions.

Digital technologies, including artificial intelligence and machine-learning algorithms, have been used by various actors to influence democratic processes and outcomes. For instance, the recourse to *bots* and *trolls* during electoral campaigns with a view to manipulating voter behaviour or coordinated disinformation campaigns.

The issues of misinformation and disinformation are further exacerbated by technological developments such as deep fakes. These technological advancements make it harder for all stakeholders, including the service providers themselves as well as citizens and civil society at large, public authorities, and media specialists to identify the truth content and thus separate fact from fiction.

The Council of Europe has addressed issues related to these trends, among others the Study on the use of internet in electoral campaigns ([39]) in 2018; Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries ([40]) and the Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes ([41]).

The Venice Commission notes: "*The "democratisation" of content production and the centralisation of online distribution channels have had as unintended consequence the proliferation of false information, private and public disinformation tactics. The advent of every means of communication (a) expands the dissemination of and the access to information*

[39] https://edoc.coe.int/en/internet/7614-internet-and-electoral-campaigns-study-on-the-use-of-internet-in-electoral-campaigns.html
[40] https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680790e14
[41] https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b

(freedom of communication); (b) implies the risk of abuses (malicious content); (c) opens the way to censorship and (d) to manipulation by the powerful public and private actor." ([42])

in this case also microtargeting should be mentioned

Digital microtargeting in the context of political campaigning is a technique by which political parties analyse large datasets in order to better understand the behaviour, opinions and feelings of potential voters. This allows political parties to cluster voters into groups which in turn receive messages that speak to their concerns and resonate with their opinions. Instead of one central message for all, political parties can disseminate a multiplicity of targeted messages in various formats and channels to carefully chosen audiences.

In the context of AI-enabled microtargeting by political parties, one of the key challenges is to define what actually constitutes political advertising. In addition, who should decide the definition? Are internet intermediaries and platforms entitled to provide a definition? Should political parties decide what information they categorise as a political advertisement? Should independent bodies be set up to decide? What happens when decisions are challenged? Who is in charge?

For the public as well as for relevant oversight bodies, it is difficult to track who receives what message. As such digital microtargeting brings campaigning to a whole new level of sophistication and can be used both positively and negatively. It might enhance or undermine democratic values and strengthen or amplify either democratic or populists' voices. Machine-learning algorithms are also being used by political parties to refine their message.

The EU General Data Protection Regulation (GDPR) establishes strict guidelines based on individual consent for the collection and processing of personal data, placing limitations on the use of digital microtargeting for parties. Furthermore, "political opinions" are defined as sensitive forms of data in Convention 108+ and thus enjoy special protection. Profiling – for the purposes of microtargeting for example – is highlight problematic and should only be considered when appropriate safeguards for the subjects concerned are in place.

While the GDPR plays a key role in the context of microtargeting, it is only a piece of the puzzle and it is insufficient for mitigating the risks caused by microtargeting.

Another danger of political microtargeting is that the public debates and democratic processes are captured by narrow interests, are fragmented or systematically exclude vulnerable or minority groups.

Lack of transparency as to the authors of the political ads and their source of financial backing may result in an uneven playing field, which further undermines trust in elections in particular and democratic processes in general. The challenges for election administration authorities in charge of monitoring the electoral process and political campaigning are manifold. They include fragmentation of enforcement and oversight, distribution of responsibilities amongst different agencies and regulators and the cross-border nature of online campaigning.

[42] https://www.venice.coe.int/webforms/documents/?pdf=CDL-AD(2019)016-e

When considering regulation of microtargeting, different rights need to be balanced. The question is where and how to draw the line, especially keeping in mind the scale and speed of information in the digital age. At the same time, civil society organisations in more fragile democracies warn of the danger that regulation of online spaces might be abused to curb political rights of the opposition.

The knowledge and understanding of AI should be promoted in government institutions, independent oversight bodies, national human rights structures, the judiciary and law enforcement, and with the general public[43].

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[44].

[43] 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

[44] 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