

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE

Strasbourg, 11 November 2020

CAHAI-PDG(2020)01rev2

**AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI)
POLICY DEVELOPMENT GROUP
(CAHAI-PDG)**

**Draft Feasibility Study
V.0.5.**

www.coe.int/cahai

1. GENERAL INTRODUCTION

1. The Council of Europe is the continent's leading human rights organisation and the guardian of the rights of some 830 million Europeans. Throughout the transformations of our society since 1949, the Council of Europe has constantly ensured that human rights, democracy and the rule of law guide development, including technological development, and some of its legal instruments have become recognised European or world standards, reconciling innovation and regulation for the benefit of human beings¹.
2. Specifically, in the digital domain, the advances of the last decades have fundamentally transformed society by providing new tools for communication, information consumption, public administration, education, and many other facets of daily life. Thanks to the detection of patterns and trends in large datasets using statistical methods, algorithmic systems now offer the possibility to recognise images or sound, streamline services or products and achieve huge efficiency gains in the performance of complex tasks. These services and products, commonly referred to as "artificial intelligence" (AI²), have the potential to promote human prosperity and individual and societal well-being by bringing progress and innovation. Member States agree that economic prosperity is an important objective of public policies and consider innovation as one of its key components. At the same time, concerns are rising in respect of harm resulting from different types of AI applications and their potential negative impact on human beings and society. Discrimination, the advent of a surveillance society, the weakening of human agency, information distortion, electoral interference, digital exclusion, attention economy, are just some of the concrete concerns that are being expressed.
3. It is therefore crucial that Council of Europe's standards on human rights, democracy and the rule of law are effectively anchored in appropriate legislative frameworks by member States. A legal response, aimed at filling legal gaps in existing legislation and tailored to the specific challenges raised by AI systems should be developed, as it has already happened in the past with innovative industrial processes such as pharmaceuticals, biomedicine or the automotive industry. Moreover, such a legal response could also foster and influence AI technologies in line with the above-mentioned standards.
4. Therefore, on 11 September 2019, the Committee of Ministers mandated an Ad hoc Committee on Artificial Intelligence (CAHAI) to examine, on the basis of broad multi-stakeholder consultations, the feasibility and potential elements of a legal framework for the development, design and application of artificial intelligence, based on Council of Europe standards in the field of human rights, democracy and the rule of law. This feasibility study takes into account Council of Europe standards for the design, development and application of digital technologies in the field of human rights, democracy and the rule of law, as well as existing relevant international - universal and regional - legal instruments. It also takes into account work carried out by other bodies of the Council of Europe as well as work in progress within other regional and international organisations (in particular within the United Nations – including UNESCO, ITU, WIPO and the WHO – the European Union, OECD, OSCE, the World

¹ See in this regard the [Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data \("Convention 108", ETS No. 108\)](#) and [its Protocol](#) ("Convention 108 +", CETS No. 223); [the Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine](#), ETS No. 164 ("Oviedo Convention") ; the [Convention on Cybercrime, ETS No. 185](#) ("Budapest Convention"); the [Convention on Elaboration of a European Pharmacopeia, ETS No. 50](#).

² In order to avoid any form of anthropomorphising and to include all technologies falling under the umbrella term of 'AI', the terms "AI systems,", "AI applications", "AI solutions" will be generally preferred in this feasibility study to refer to algorithmic systems based, indifferently, on machine learning, deep learning, rule-based systems such as expert systems or any other form of computer programming and data processing. The notion of "algorithmic systems" is to be understood as defined in the appendix to Recommendation CM/Rec(2020)1 of the Committee of Ministers, as "applications which, often using mathematical optimisation techniques, perform one or more tasks such as collecting, grouping, cleaning, sorting, classifying and deriving data, as well as selecting, prioritising, making recommendations and taking decisions. By relying on one or more algorithms to perform their tasks in the environments where they are implemented, algorithmic systems automate activities to enable the creation of scalable, real-time services".

Bank, and the World Economic Forum). Finally, this study takes into account a gender equality perspective and the building of cohesive societies and the promotion and protection of the rights of vulnerable people, including persons with disabilities and minors.

2. SCOPE OF APPLICATION OF A COUNCIL OF EUROPE LEGAL FRAMEWORK ON ARTIFICIAL INTELLIGENCE

5. To date, there is no single definition of AI accepted by the scientific community. The term, which has become part of everyday language, 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. The term can therefore cover any automation resulting from this technology, as well as precise technologies such as machine learning or deep learning based on neural networks.
6. Similarly, the various international organisations that have worked on AI have also not found a consensus on the definition of AI. The High-Level Group of Independent Experts mandated by the European Commission has therefore published a comprehensive document on the definition of AI³. The European Commission's AI Watch Observatory has also conducted a very thorough study on an operational definition and taxonomy of AI⁴. The OECD Council Recommendation on AI includes a preamble defining AI systems, the life cycle of an AI system, AI knowledge, AI actors and stakeholders⁵. UNESCO has produced a preliminary study referring to "AI-based machines" and "cognitive computing"⁶ as well as a draft Recommendation on the Ethics of Artificial Intelligence defining AI systems as "technological systems which have the capacity to process information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control"⁷.
7. As regards the non-binding instruments that have been published on this topic by the Council of Europe so far, no uniform definition of AI has been used. The Recommendation of the Committee of Ministers to member States on the impact of algorithmic systems on human rights⁸ defines the notion of "algorithmic systems" which may cover a broad range of AI applications. The Declaration of the Committee of Ministers on the Manipulation Capabilities of Algorithmic Processes⁹ does not include definitions and uses various concepts such as "technologies", "data-based systems", "machine learning tools", depending on the specific objects to be considered. The Commissioner for Human Rights¹⁰, the Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (T-PD¹¹) and the European Commission for the Efficiency of Justice (CEPEJ¹²) use a relatively similar generic definition referring to a set of sciences, theories and techniques.
8. In sum, it can be concluded that the term "AI" is used as a "blanket term" for various computer applications based on different techniques, which exhibit capabilities commonly and currently associated with human intelligence. These techniques can consist of formal models (or symbolic systems) as well as data-driven models (learning-based systems) typically relying on statistical

³ [AI HLEG, A Definition of AI: Main Capabilities and Disciplines, April 2019.](#)

⁴ [AI Watch, Joint Research Centre, Defining Artificial Intelligence: towards an operational definition and taxonomy of artificial intelligence, February 2020.](#)

⁵ [OECD, Council Recommendation on Artificial Intelligence, June 2019.](#)

⁶ [UNESCO, Preliminary study on the technical and legal aspects relating to the desirability of a standard-setting instrument on the ethics of artificial intelligence, March 2019](#)

⁷ [UNESCO, First draft of the Recommendation on Ethics of Artificial Intelligence, September 2020](#)

⁸ [Council of Europe, Committee of Ministers, Recommendation CM/Rec\(2020\)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, April 2020.](#)

⁹ [Committee of Ministers, Declaration of the Committee of Ministers on the manipulative capabilities of algorithmic processes, February 2019.](#)

¹⁰ [Commissioner for Human Rights, Unboxing AI: 10 steps to protect human rights - Recommendation of the Commissioner for Human Rights, May 2019.](#)

¹¹ [Data Protection Convention Advisory Committee on the Automated Processing of Personal Data \(T-PD\), Guidelines on Artificial Intelligence and Data Protection, January 2019.](#)

¹² [CEPEJ, European Ethical Charter for the use of artificial intelligence in judicial systems and their environment, December 2018.](#)

approaches, including for instance supervised learning, unsupervised learning and reinforcement learning. AI systems act in the physical or digital dimension by recording their environment through data acquisition, analysing certain structured or unstructured data, reasoning on the knowledge or processing information derived from the data, and on that basis decide on the best course of action to reach a certain goal. They can be designed to adapt their behaviour over time based on new data and enhance their performance towards a certain goal.

9. Whereas the CAHAI members, participants and observers have also indicated different approaches on the (need for a) definition of AI resulting from different legal traditions and cultures, a consensus has been found on the need to approach AI systems in a technologically neutral way, whereby all the various automated decision-making technologies that fall under this umbrella term are comprised, including their broader socio-technical context. Furthermore, a balance should be sought between a definition that may be too precise from a technical point of view and might thus be obsolete in the short term, and a definition that is too vague and thus leaves a wide margin of interpretation, potentially resulting in a non-uniform application of the legal framework.
 10. As a result, a future legal framework should adopt a simplified and technologically neutral definition of its purpose, covering those application cases where AI systems can impact on human rights, democracy and the rule of law, and taking into account all of the systems' socio-technical implications.¹³
- 3. OPPORTUNITIES AND RISKS ARISING FROM THE DESIGN, DEVELOPMENT AND APPLICATION OF ARTIFICIAL INTELLIGENCE ON HUMAN RIGHTS, THE RULE OF LAW AND DEMOCRACY. "GREEN" AND "RED" AREAS" - MEANING RESPECTIVELY POSITIVE AND PROBLEMATIC EXAMPLES OF ARTIFICIAL INTELLIGENCE APPLICATIONS FROM A HUMAN RIGHTS, THE RULE OF LAW AND DEMOCRACY PERSPECTIVE, WHILE CONSIDERING THE CONTEXT-SENSITIVE ENVIRONMENT FOR ARTIFICIAL INTELLIGENCE DESIGN, DEVELOPMENT AND APPLICATION IN EUROPE AND DEVELOPMENTS AT GLOBAL LEVEL.**

3.1 Introduction

11. As noted in various Council of Europe documents, including reports recently adopted by the Parliamentary Assembly (PACE)¹⁴, AI systems are substantially transforming individual lives and have a profound impact on the fabric of society and the functioning of its institutions. Their use has the capacity to generate substantive benefits in numerous domains, such as healthcare, transport, education and public administration, generating promising opportunities for humanity at large. At the same time, the development and use of AI systems also entails substantial risks, in particular in relation to interference with human rights, democracy and the rule of law, the core elements upon which our European societies are built.
12. AI systems should be seen as "**socio-technical systems**". The impact of an AI system depends not only on the system's design, but also on the way in which the system is developed and used within a broader environment, including the data used, its intended purpose, functionality and accuracy, the

¹³ The CAHAI Legal Framework Group should examine this issue in greater depth and propose an agreed solution as part of its mandate. It is worth noting in this respect that other legal instruments of the Council of Europe relating to scientific fields, the Convention on Human Rights and Biomedicine ("Oviedo Convention", ETS No. 164) does not define its subject matter either. The "Convention 108", as its modernised version (Convention 108+) defines the concept of "data processing", without mentioning specific technical objects such as algorithms and link such concept to the notion of "personal data", thus making it possible to determine whether or not a processing operation falls within its scope.

¹⁴ See e.g. the reports of the Parliamentary Assembly of the Council of Europe, in particular on [the need for democratic governance of AI](#); [the role of AI in policing and criminal justice systems](#); [preventing discrimination caused by AI](#); [ethical and legal frameworks for the research and development of neurotechnology](#); [AI and health care](#); [consequences of AI on labour markets](#); and [legal aspects of 'autonomous vehicles'](#). See also the Recommendation by the Commissioner for Human Rights, "Unboxing artificial intelligence: 10 measures to protect human rights"; the Committee of Ministers' Recommendation Rec/CM(2020)1,

scale of deployment, and the broader organisational, societal and legal context in which it is used.¹⁵ The positive or negative consequences depend also on the values and behaviour of the human beings that develop and deploy them, which leads to the importance of ensuring human responsibility. There are, however, some distinct characteristics of AI systems that set them apart from other technologies in relation to both their positive and negative impact on human rights, democracy and the rule of law.¹⁶

13. First, the **scale and reach of AI systems** can amplify certain risks that are also inherent in other technologies or human behaviour. AI systems can analyse an unprecedented amount of fine-grained data (including highly sensitive personal data) at a much faster pace than humans. This ability can lead AI systems to be used in a way that perpetuates or amplifies unjust **bias**¹⁷, also based on new discrimination grounds.¹⁸ The increased prominence of proxy discrimination in the context of machine learning may raise interpretive questions about the distinction between direct and indirect discrimination or, indeed, the adequacy of this distinction as it is traditionally understood. Moreover, AI systems are subject to statistical error rates. Even if the error rate of a system applied to millions of people is close to zero, thousands of people can still be adversely impacted due to the scale of deployment. On the other side, the scale and reach of AI systems also imply that they can be used to mitigate certain risks and biases that are also inherent in other technologies or human behaviour, and to monitor and reduce human error rates.
14. Second, the **complexity or opacity** of many AI systems (in particular in the case of machine learning) can make it difficult for humans, including a system's developers, to understand or trace the system's functioning or outcome. This opacity, in combination with the involvement of many different actors at different stages during the system's lifecycle, further complicates the identification of the agent(s) responsible for a potential negative outcome, hence reducing human accountability.
15. Third, certain AI systems can **re-calibrate** themselves through feedback and reinforcement learning. Given the system's complexity, if the newly generated data based on the recalibration leads to a discriminatory, erroneous or malicious functioning of the system, resulting from its own decisions, a vicious feedback loop may arise which can be difficult to detect.

3.2 Opportunities arising from AI

16. AI systems can have a highly positive impact across society. As a key driver for socio-economic development globally, they can contribute to alleviating some of the world's problems. AI systems can optimise agricultural processes, revolutionize transportation and urban living, help mitigate the effects of climate change or predict natural disasters. They can significantly improve the efficiency of existing industry practices, assist in the development of new industrial applications, and enhance their safety. AI systems can also lead to the creation of new services, products, markets and industries, which can significantly increase the well-being of citizens and society at large and be used to support socially beneficial applications and services.
17. Furthermore, AI systems can provide intelligent capability in many areas that are of value to individuals and society at large, and given their efficiency and large scale effects, be used to help overcome some of the barriers posed by the limited availability of human cognitive and decision-making capability.
18. One of the most significant attributes of AI systems is their potential impact on human health and healthcare systems. This includes the improvement of medical diagnosis and treatment, the improvement of foetal health, as well as the advanced prediction and monitoring of epidemics and chronic diseases. Some opportunities generated by AI systems can also be observed within the

¹⁵ See also Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems.

¹⁶ These three factors are interacting and mutually reinforcing. Given the rapid evolution of the technology and its unforeseen uses in future, this list is not conclusive but subject to constant development.

¹⁷ "Unjust bias" means a violation of the right to equality and non-discrimination in a context specific application of AI technology.

¹⁸ Affinity Profiling and Discrimination by Association in Online Behavioural Advertising, Wachter 2020.

response to the COVID-19 pandemic. AI systems are deployed to study the virus, accelerate medical research, develop vaccines, detect and diagnose infections, predict the virus' evolution, and to rapidly exchange information.

19. Finally, AI systems can foster and strengthen human rights more generally, and contribute to their effective application and enforcement, for instance by detecting biased (human or automated) decisions, monitoring representation patterns of different people or groups (for example women in the media) or analysing discriminatory structures in organisations. 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 the authorities and other 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 focused policies.

3.3 Impact on Human Rights, Democracy and the Rule of Law

20. Despite these benefits, the increasing use of AI systems in all areas of private and public life also carries significant challenges for human rights, democracy and the rule of law¹⁹. Examples of known cases for each is discussed below. Given that respect for human rights is an essential component of the rule of law and democracy, the review of the challenges posed by AI systems specifically to the rule of law and democracy is closely entwined with the analysis of the impact of AI systems on human rights.

3.3.1 Impact on Human Rights

21. The development and use of AI systems has an impact on a wide range of human rights. The main issues are briefly set out below, focusing in particular on the rights set out by the European Convention on Human Rights ("ECHR"), its Protocols and the European Social Charter ("ESC").

Liberty and Security; Fair Trial; No Punishment without Law (Art. 5, 6, 7 ECHR)

22. The above-mentioned risks of AI systems to facilitate or amplify unjust bias can pose a threat to the right to liberty and security combined with the right to a fair trial (Art. 5, 6, 7 ECHR) when AI systems are used in situations where physical freedom or personal security is at stake (such as justice and law enforcement). For instance, AI systems used to predict recidivism rely on characteristics of the suspect shared with others (such as address, income, nationality, debts, employment), which raises concerns as regards maintaining an individualised approach to sentencing and other fundamental aspects of the right to a fair trial. In addition, an AI system's opacity may render it impossible to understand the reasoning behind its outcomes, hence making it difficult or impossible to ensure the full respect of the principle of equality of arms, to challenge the decision or seek effective redress. If applied responsibly however, certain AI applications can also make the work of justice and law enforcement professionals more efficient and hence impact positively on these rights.

Private and Family Life; Physical, Psychological and Moral Integrity (Art. 8 ECHR)

23. Art. 8 ECHR encompasses the protection of a wide range of aspects of our private lives, which can be divided into three broad categories namely: (i) a person's (general) privacy, (ii) a person's physical, psychological and moral integrity and (iii) a person's identity and autonomy.²⁰ Various AI applications can impact these categories. This occurs most notably when personal data is processed (for instance to identify or surveil individuals), but it can also occur without the processing of personal data. Examples of invasive AI applications include in particular systems that track the faces or other biometrical data of individuals, such as micro-expressions, gait, (tone of) voice, heart rate or

¹⁹ See the report prepared by Catelijne Muller (CAHAI(2020)06-fin), The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law.

²⁰ See Guide on Article 8 of the ECHR, Council of Europe.

temperature data.²¹ Beyond identification purposes, such data can also be used to assess, predict and influence a person's behaviour, and to profile or categorise individuals for various purposes and in different contexts, from predictive policing to insurance.²² There is also ample evidence that the use of biometric recognition technology can lead to discrimination, notably on the basis of race and/or gender, when bias in the algorithm or underlying dataset is insufficiently addressed.²³

24. Furthermore, the widespread use of AI-based tracking techniques broadly affects 'general' privacy, identity and autonomy by creating a situation in which individuals are constantly watched, followed, identified and influenced, thereby also affecting their moral and psychological integrity. People might feel inclined to adapt their behaviour to a certain norm, thus shifting the balance of power between the state or private organisation who uses surveillance technologies on the one hand, and the (group of) individuals on the other.²⁴ The indiscriminate on- and offline tracking of all aspects of people's lives (through online behaviour, location data, data from smart watches and other Internet-of-Things (IoT) applications, such as health trackers, smart speakers, thermostats, cars, etc.), can have the same impact on the right to privacy, including to psychological integrity. A right to privacy implies a right to a private space free from AI-enabled surveillance as necessary for personal development and democracy.²⁵

Freedom of expression; Freedom of assembly and association (Art. 10, 11 ECHR)

25. The use of AI systems - both online and offline - can impact individuals' freedom of expression, as well as the freedom of assembly and association. AI applications can be used to intervene in media space with super-human efficiency, and substantively alter human interactions. The internet and social media platforms have shown huge potential for people organising themselves to exercise their right to peaceful assembly and association. At the same time, the use of AI-driven surveillance can jeopardise these rights by automatically tracking and identifying those (groups of) individuals or even excluding them from protests.²⁶ Moreover, the personalised tracking of individuals – in virtual and real life – may hamper these rights by diminishing the protection of 'group anonymity'. This can lead to individuals no longer partaking in peaceful demonstrations, and to more generally refraining from openly expressing their opinions, watching certain media or reading certain books or newspapers.
26. Furthermore, AI systems can affect the right to receive and impart information and ideas when used in online (social) media and news curations to pre-sort or display the content according to personal interests or habits. They can thereby also reinforce outdated social norms, including gender stereotypes.²⁷ Search engines, recommendation systems and news aggregators often are non-transparent and unaccountable, both concerning the data they use to select or prioritise content, but

²¹ The case law of the European Court of Human Rights (ECtHR) makes clear that the capture, storage and processing of such information, even briefly, impacts art. 8 ECHR.

²² It can be noted that no sound scientific evidence exists corroborating that a person's inner emotions or mental state can be accurately 'read' from a person's face or other biometric data. See also the study by Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019).

²³ See e.g. the MIT Study by Joy Buolamwini (2018): <https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212>; the US National Institute of Standards and Technology study on face recognition (2019): <https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf>.

²⁴ Study by Catelijne Muller, CAHAI(2020)06-fin, para 18; Examined Lives: Informational Privacy and the Subject as Object, Julie E. Cohen, 2000.

²⁵ The chilling effect describes the inhibition or discouragement of the legitimate exercise of a right. Studies have shown that, once people know they are being surveyed, they start to behave and develop differently. Staben, J. (2016). Der Abschreckungseffekt auf die Grundrechtsausübung: Strukturen eines verfassungsrechtlichen Arguments. Mohr Siebeck.

²⁶ Algorithms and Human Rights, Study on the human rights dimensions of automated data processing techniques and possible regulatory implications, Council of Europe, 2018.

²⁷ See e.g. a study by Carnegie Mellon researchers: <https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study>

also as concerns the purpose of the specific selection or prioritisation²⁸ which they can use for financial and political interest promotion. AI systems are routinely used to select and prioritise content that keeps people on the platform as long as possible, irrespective of whether the content is objective, factually true, diverse or relevant. Furthermore, content is increasingly being “faked” by producing synthetic media footage, e.g. by Mimicking real people’s appearance or voice using so called deep fakes. Such technology is already able to manipulate or generate visual and audio content with an unprecedented potential to deceive and to blur the line between real and fake content. This significantly affects the capacity of individuals to form and develop opinions freely, to receive and impart information and ideas, which might lead to an erosion of our information society.²⁹ Apart from that, online platforms are increasingly turning to AI systems to remove content which breaches their terms of service. Inaccuracies of the AI systems can lead to the consequence that legitimate content – protected by the right to freedom of expression – is flagged or removed in error.

Equality and Non-Discrimination (Art. 14 ECHR, Protocol 12)

27. The impact of AI systems on the prohibition of discrimination and the right to equal treatment is one of the most widely reported upon. This also has an impact on securing the availability of equal opportunities and (gender) equality. As noted above, AI systems can be used to detect and mitigate human bias. At the same time, the use of AI systems can also enable the perpetuation and amplification of biases, such as gender and other stereotypes, sexism, racism, ageism and other unjust discriminations, which creates a new challenge to non-discrimination and equal treatment. As AI systems clearly possess super-human capacity to detect patterns and discriminate, applications of AI create a real risk to equality and non-discrimination.
28. The risk of discrimination can arise in multiple ways, for instance due to biased training data (e.g. when the data-set is not sufficiently representative or inaccurate), due to a biased design of the algorithm or its optimisation function, due to exposure to a biased learning environment, due to the conscious or unconscious stereotypes or biases of developers or due to a biased use of the AI system. For instance, in light of past legal or factual discriminations against women, historical data bases can lack sufficiently gender-balanced data. When such a data base is subsequently used by AI systems, this can lead to equally biased decisions and hence perpetuate unjust discrimination. The same holds true for other traditionally vulnerable, excluded or marginalised groups. In addition, the gaps in representation of the above-mentioned groups in the AI sector might further contribute to the amplifying the risk of bias and unjust discrimination in AI systems.³⁰ Measures to improve the diversity of the AI workforce in terms of gender and ethnic/social origin could help mitigate some of those risks.
29. Given the lack of transparency in many AI systems, and the lack of mandatory reporting or auditability requirements, the existence of such biases can easily remain undetected or even be obscured, and thus marginalise the social control mechanisms that typically govern human behaviour.

Social and Economic Rights (Art. 2, 3, 5 and 20 ESC)

30. AI systems can have major benefits when used for hazardous, heavy, exhausting, unpleasant, repetitive or boring work. However, the wide adoption of AI systems in all domains of our lives also creates new risks to social and economic rights. AI systems are increasingly used to monitor and track workers, distribute work without human intervention and assess and predict worker potential and

²⁸ Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 2053951715622512.

²⁹ UN Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/73/348. Its effects on democracy are further discussed below.

³⁰ According to the study conducted by AI Now Institute “Discriminating systems, Gender, Race, and Power in AI” in 2019, women comprised only 15% of AI research staff at Facebook and 10% at Google. There is no public data on trans workers or other gender minorities. For black workers, the picture is even worse. For example, only 2.5% of Google’s workforce is black, while Facebook and Microsoft are each at 4%. Given decades of concern and investment to redress this imbalance, the current state of the field is alarming.

performance in hiring and firing situations. Furthermore, AI can also be used to detect and counter the unionisation of workers. These applications can jeopardise the right to just, safe and healthy working conditions, dignity at work as well as the right to organise. The discrimination capacity of AI systems that assess and predict the performance of job applications or workers can also undermine equality, including gender equality, in matters of employment and occupation.

3.3.2 Impact on Democracy

31. The development and use of AI systems can also impact the functioning of democratic institutions and processes, as well as the social and political behaviour of citizens and civil society at large.³¹ Where designed, deployed and used responsibly, AI systems could improve the quality of governance, namely accountability, responsiveness and efficiency of public institutions, to help fight corruption and to foster pluralism. They can help broaden the space for democratic representation by decentralising information systems and communication platforms. Moreover, they can improve the way the citizens and civil society at large receive and collect information about political processes and help them participate therein remotely by facilitating political expression and providing feedback channels with political actors. At the same time, AI systems can be used in ways that (un)intentionally hamper democracy.
32. A functioning democracy relies on open social and political discourse, as well as the absence of improper voter influence. As indicated above, AI technologies can be used by private interest to interfere in human media space for private financial or political gain. While propaganda and manipulation are not new, AI-based tools have amplified their scale and outreach, and facilitated rapid iteration to strengthen their capabilities to influence people. They enable large scale yet targeted misinformation campaigns, for instance through deep fakes, the micro-targeting of voters and the polarisation of public debate. Moreover, they can threaten to undermine the human agency and autonomy required for meaningful voter decisions, which are at the heart of the creation of legitimate institutions. As a consequence, certain uses of AI can undermine confidence in democratic institutions and hinder the electoral process.
33. More generally, the concentration of power in the hands of a few private platforms with limited regulation so far, while these platforms have de facto become part of the public sphere, can amplify these risks. Furthermore, public-private collaborations on the use of AI in sensitive fields, such as law enforcement or border control, can blur the boundaries between the interests and responsibilities of democratic states on the one hand, and of private corporations on the other. This raises *inter alia* questions as regards the accountability of public institutions for decisions taken through AI solutions provided by private actors.
34. Finally, AI's impact on the human rights set out above can more generally have a negative impact on democracy. AI systems can for instance be used by governments to control citizens, e.g. by automatically filtering information (which can amount to censorship) or by using AI-enabled (mass) surveillance. Such use of AI systems can hamper democratic values, curb the free will of the people, and might further lead to the erosion of political freedoms – such as freedom of expression, association and assembly.
35. So far, public institutions have predominantly used AI to support standardised administrative decisions. The prospective reliance on AI by public institutions to decide policies would be very problematic, as no decision taken through AI will be able to replace a dialogue between the majority and the minority. In addition, growing reliance on AI could substantially affect the nature of state powers (legislative, executive and judiciary) and affect the balance between each of them, both for better and for worse.

³¹ For further details on the impact of AI systems on democracy, see the report for the Parliamentary Assembly of the Council of Europe on the “Need for democratic governance of artificial intelligence” (Doc. 15150).

3.3.3 Impact on the Rule of Law

36. In addition to impacting human rights and democracy, AI systems can also affect the rule of law. The rule of law prescribes that all public authority act within the constraints set out by law, in accordance with the principles of democracy and human rights, and under the control of independent and impartial courts. When used responsibly, AI systems can be used to increase the efficiency of governance, including legal institutions such as the courts³², as well as law enforcement and public administrations.³³ Furthermore, AI systems can help agencies to identify corruption within public entities,³⁴ as well as detect and defend against cyberattacks.³⁵
37. The rule of law requires respect for principles such as legality, transparency, accountability, legal certainty, equality and effective judicial protection – which can be at risk when certain decisions are delegated to AI systems. In addition, AI systems can also negatively affect the process of law-making and of the application of the law by judges.³⁶ Concerns have also been expressed on the possible negative effects of some AI applications being used in judicial systems or connected areas³⁷, which could have a negative impact on key elements of the right to a fair trial enshrined in Article 6 of the ECHR, of which components such as the right to an independent and impartial judiciary, the right to a lawyer or the principle of equality of arms in judicial proceedings are key elements that are also of essential for the effective implementation of the rule of law.
38. Moreover, there is increased pressure, including through regulation, for companies to make decisions around the legality of content. Because of social media platforms becoming the new “public square”, their own terms of service essentially set the rules of how freedom of expression manifests itself online, but with fewer safeguards than in more traditional public settings. It is however essential that states can and do continue to fulfil their responsibility for the protection of the rule of law.

3.4 Green and Red lines to be drawn, within a contextual and risk-based approach

39. The above demonstrates that applications of AI systems pose a range of risks to human rights, democracy and the rule of law, yet these risks depend on the application context, technology and actors involved. In order to counter any stifling of socially beneficial AI innovation, and to ensure that the benefits of this technology can be reaped fully while adequately tackling its risks, the CAHAI Policy Development Group (CAHAI-PDG) has recommended that a future legal framework should pursue a risk-based approach targeting specific application context. This means not only that the risks posed by AI systems should be assessed and reviewed on a systematic and regular basis, but also that any

³² AI systems can support legal professionals’ work, for instance by assisting with complex tasks like analysing and structuring information on legal cases and legal documents, transcribing the minutes of court proceedings, promoting automated document classification hence eliminating a lot of processing time for the courts, civil registries and territorial offices, or providing legal information via chatbots.

³³ Danaher, J. (2016). The Threat of Algocracy: Reality, Resistance and Accommodation. *Philosophy & Technology*, 29(3), 245–268.

³⁴ West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers & Security*, 57, 47–66. Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods. *Knowledge-Based Systems*, 128, 139–152.

³⁵ Taddeo, M., & Floridi, L. (2018a). Regulate artificial intelligence to avert cyber arms race. *Nature*, 556(7701), 296–298.

³⁶ By favouring the emergence of quantitative trends of analysis of judicial decisions, the traditional process of application of the law by the judge, with the j could be jeopardised. See the CEPEJ European Ethical Charter on the use of AI in judicial systems and their environment, §35. See e.g. G. Buchholtz, “Artificial Intelligence and Legal Tech: Challenges to the Rule of Law” in T. Wischmeyer, T. Rademacher (eds.), *Regulating Artificial Intelligence*, Springer (2020).

³⁷ See the [CEPEJ European Ethical Charter on the use of AI in judicial systems and their environment](#), which refers specifically to risks arising from systems of anticipation of judicial decisions in civil, administrative and commercial matters, from risk-assessment systems in criminal matters, and from the use of AI systems without appropriate safeguards in the framework of non-judicial alternative dispute resolution. Among those risks the CEPEJ notes the risks of “performative effect” and of delegation of responsibility, and of lack of transparency of judicial decision-making.

mitigating measures, that are further elaborated under Chapter 7, should be specifically tailored to these risks. In addition to the risk-based approach, where relevant - for instance where a certain AI system in a certain context poses a significant level of risk coupled with a high level of uncertainty - a precautionary approach should be considered.

40. AI applications that promote, strengthen and augment the protection of human rights, democracy and the rule of law, would generally be considered to fall within “green lines” as long as their risks are known and clearly minimal. However, where based on a context-specific risk assessment it is found that an AI application can pose “significant” or unknown risks to human rights, democracy or the rule of law, and no appropriate mitigation measures exists within existing legal frameworks to adequately mitigate these risks, states should consider the introduction of strong restrictions for the exceptional and controlled use of the application, and where essential, a (temporary) ban or moratorium (red lines).³⁸ Examples of such applications are remote biometric recognition systems – or other AI-enabled tracking applications – that risk leading to mass surveillance or to social scoring, or AI-enabled covert manipulation of individuals, each of which significantly impact individuals’ autonomy as well as fundamental democratic principles and freedoms. Exceptional use of such technologies should be specifically foreseen by law, necessary in a democratic society and proportionate to the legitimate aim, and permissibly only in controlled environments and (if applicable) for limited periods of time. On the other hand – where a certain application of an AI system does not pose any risk to human rights, democracy or the rule of law – it should be exempted from any additional regulatory measures. The CAHAI-PDG notes that, both as concerns red lines and green lines, a contextual and periodical assessment is necessary, in light of the context-specific nature of the benefits and risks related to the application of AI. As a transversal technology, the same AI technology can be used for different purposes and in different contexts, and the positive or negative consequences of the technology will depend heavily thereon.

4. THE COUNCIL OF EUROPE'S WORK IN THE FIELD OF ARTIFICIAL INTELLIGENCE TO DATE

41. The significant impact of information technologies on human rights, democracy and the rule of law has led the Council of Europe to develop relevant binding and non-binding mechanisms, which complement and reinforce one another. They will be examined below, along with the case law on new technologies of the European Court of Human Rights.

4.1. Work in the field of protection of personal data

42. "Convention 108³⁹", modernised by an amending protocol in 2018⁴⁰ ("Convention 108+"), sets global standards on the rights to privacy and data protection of individuals, regardless of technological evolutions. In particular, it requires that the processing of sensitive data⁴¹ only be allowed where appropriate safeguards are enshrined in law, complementing those of the Convention, and creates a right for everyone to know that their personal data are processed and for which purpose, with a right of rectification. The amending protocol added new principles, such as transparency, proportionality, accountability, respect for privacy by design. As regards the rights of individuals, the right not to be subject to a decision significantly affecting him or her based solely on an automated processing of data without having his or her views taken into consideration, and the right to obtain knowledge of the reasoning underlying the processing of data, where the results of the processing are applied, have

³⁸ One of the intentions of building international agreement on red lines is to prevent competitive disadvantages. Red Lines in the form of moratoria could in some instances be overcome when provisions can be set out to secure appropriate methods to develop trustworthy (legal, ethical and robust AI), for instance where prior evaluation, continuous monitoring, certification procedures or standardised development processes can ensure appropriate guarantees to safeguard human rights, democracy and the rule of law.

³⁹ [Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data \("Convention 108", ETS No. 108](#)

⁴⁰ [Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, CETS No. 223](#)

⁴¹ See Article 6 of Convention 108+ for the full list of sensitive data.

been introduced (Article 9). Those new rights are of particular importance in relation to the profiling of individuals and automated decision-making⁴².

43. Even if it is not specific to AI applications, the legal framework built around Convention 108 remains fully applicable to AI technology as soon as the processed data fall within the scope of this text. Guidelines and a report in 2019 specified the guiding principles to be applied, both for legislators and decision makers and for developers, manufacturers and service providers⁴³. A legal instrument on AI applications will therefore have to take full account of this acquis to supplement it by including in its scope such processing operations that do not only involve personal data and by extending its scope to the prevention of societal (and not only individual) harm.

4.2. Work in the field of cybercrime

44. Various AI systems may entail major risks in the field of cybercrime and are already much used to perpetrate such crimes, from automated and coordinated distributed denial of service attacks to scanning systems for vulnerabilities, social engineering and identity theft, and autonomous cybercrime by machines. The Convention on Cybercrime ("Budapest Convention") is the international reference instrument for criminalising offences against and by means of computers, for procedural powers to investigate cybercrime and secure electronic evidence in relation to any crime subject to rule of law safeguards, and for effective international co-operation.⁴⁴ A new Protocol to the Budapest Convention on enhanced co-operation on cybercrime and electronic evidence is in preparation and may become available in 2021. The Budapest Convention and its provisions are fully applicable to acts carried out or facilitated by AI systems.

4.3. Work in the field of algorithmic systems

45. The Committee of Ministers adopted a Declaration on the Manipulation Capabilities of Algorithmic Processes⁴⁵ in February 2019 and a Recommendation on the Human Rights Impacts of Algorithmic Systems⁴⁶ in April 2020. Studies and reports on the human rights' dimensions of automated data processing techniques⁴⁷ and accountability and AI⁴⁸ have also been developed by specialised committees and expert bodies, while the development of a standard-setting instrument (or other form

⁴² See in this respect [Recommendation \(2010\)13 on the protection of individuals with regard to automatic processing of personal data in the context of profiling, and its explanatory memorandum](#). The Committee of Convention 108 is currently working on updating this important Recommendation.

⁴³ [Advisory Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data \(T-PD\), Guidelines on Artificial Intelligence and Data Protection, January 2019](#) and [Advisory Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data \(T-PD\), Report on Artificial Intelligence \(Artificial Intelligence and Data Protection: Challenges and Possible Solutions\), January 2019](#)

⁴⁴ [Convention on Cybercrime, ETS No. 185](#) – By October 2020 it had 65 Parties and 12 States had signed it or been invited to accede. The only Council of Europe member State not having signed it to date is the Russian Federation due to apprehensions regarding Article 32.b. The Cybercrime Convention Committee (T-CY) had addressed this concern through a [Guidance Note](#) adopted in December 2014.

⁴⁵ [Committee of Ministers, Declaration on the manipulation capabilities of algorithmic processes - Decl\(13/02/2019\)1, 13 February 2019](#) The Declaration draws, inter alia, member States' attention to "properly assess the need for stricter regulatory or other measures to ensure appropriate and democratically legitimate oversight of the design, development, deployment and use of algorithmic tools, with a view to implementing effective protection against unfair practices and abuses of economic power".

⁴⁶ [Committee of Ministers, Recommendation to member States on the human rights impacts of algorithmic systems - CM/Rec\(2020\)1, 8 April 2020](#) The Recommendation, for its part, invites member States to "review their legislative frameworks and policies, as well as their own practices with regard to the ongoing acquisition, design, development and deployment of algorithmic systems to ensure that they are in line with the guidelines set out in the Appendix to this Recommendation".

⁴⁷ See the study produced by the Committee of experts on Internet Intermediaries (MSI-NET) under the authority of the Steering Committee on the Media and the Information Society (CDMSI) on the human rights' dimensions of automated data processing techniques and possible regulatory implications [DGI\(2017\)12](#).

⁴⁸ [MSI-AUT, Accountability and AI: Study on the impact of advanced digital technologies \(including artificial intelligence\) on the notion of accountability, from a human rights perspective - DGI\(2019\)05, September 2019](#)

of guidance from the Committee of Ministers to member States) on the impacts of digital technologies on freedom of expression⁴⁹ is underway.

4.4 Work in the field of justice

46. The European Commission for the Efficiency of Justice (CEPEJ) adopted in December 2018 the European Ethical Charter for the use of artificial intelligence in judicial systems⁵⁰ which sets five key principles (respect of fundamental rights, non-discrimination, quality and security, transparency, impartiality and fairness, "under the control" of the user) for the use of AI systems in this field. The CEPEJ is currently studying the advisability and feasibility of a certification or labelling framework for artificial intelligence products used in judicial systems. The European Committee on Legal Cooperation (CDCJ) is preparing guidelines to ensure the compatibility of these mechanisms with Articles 6 and 13 of the Convention on Human Rights. The European Committee on Crime Problems (CDPC) is currently studying the topic of AI and criminal law and may propose the creation of a new specialised legal instrument⁵¹.

4.5 Work in the field of Good Governance and elections

47. The European Committee on Democracy and Governance (CDDG) is preparing a study on the impact of digital transformation – including AI – on democracy and governance. The study looks at the impact of AI on elections, civil participation, and democratic oversight. In the chapter devoted to governance, it maps the use of AI by public administrations in Europe and analyses its use through the lens of the 12 Principles of Good Democratic Governance.

48. The Venice Commission has also published a report on digital technologies and elections⁵² and is currently drafting principles on a human-rights compliant use of digital technologies in electoral processes.

4.6 Work in the field of gender equality and non-discrimination

49. The Committee of Ministers Recommendation CM/Rec(2019)1 on preventing and combating sexism comprises measures to prevent and to condemn sexism in relation to artificial intelligence.

50. Work is underway in the field of equality and non-discrimination, following the comprehensive study prepared by the European Commission against Racism and Intolerance (ECRI) on "discrimination, artificial intelligence and algorithmic decision making".

4.7 Work in the field of education, youth and culture

51. The Committee of Ministers' Recommendation on developing and promoting digital citizenship education⁵³ invites member States to adopt regulatory and policy measures on digital citizenship education, assess their impact at regular intervals, provide or facilitate the provision of appropriate initial and in-service education and training on digital citizenship education to teachers and other professionals in education, to name but a few recommended measures. Building on this, the Committee of Ministers mandated the Steering Committee for Education Policy and Practice (CDPPE) to explore the implications of artificial intelligence and other emerging technologies for education

⁴⁹ See the ongoing work of the Committee of Experts on Freedom of Expression and Digital Technologies (MSI-DIG)

⁵⁰ [CEPEJ, European Ethical Charter on the use of artificial intelligence in judicial systems and their environment - CEPEJ\(2018\)14, December 2018](#)

⁵¹ See CDPC(2020)3Rev, [Feasibility Study on a future Council of Europe instrument on a future Council of Europe instrument on AI and criminal law](#).

⁵² [Venice Commission and DG1, Joint Report of the Venice Commission and the Directorate for the Information Society and Action against Crime of the Directorate General for Human Rights and the Rule of Law \(DG1\) on Digital Technologies and Elections - CDL-AD\(2019\)016, 24 June 2019](#)

⁵³ [Committee of Ministers' Recommendation on developing and promoting digital citizenship education](#), CM/Rec(2019)10.

generally and more specifically for their use in education. A Committee of Ministers Recommendation addresses specifically the rights of the child in the digital environment⁵⁴.

52. In October 2018, the Division of Culture and Cultural Heritage organised a seminar on culture, creativity and artificial intelligence in the framework of the Croatian Chairmanship of the Committee of Ministers of the Council of Europe⁵⁵. Eurimages published a study on the impact of predictive technologies and AI on the audio-visual sector, including possible specific measures to be put in place to guarantee freedom of expression and cultural diversity⁵⁶.

4.8 The work of the Parliamentary Assembly of the Council of Europe

53. The Parliamentary Assembly of the Council of Europe (PACE) adopted, on 28 April 2017, a Recommendation on "Technological convergence, artificial intelligence and human rights⁵⁷". On 22 October 2020, the PACE adopted 7 reports, focusing on: the need for democratic governance of AI; the role of AI in policing and criminal justice systems; discrimination caused by AI; threats to fundamental freedoms; medical, legal and ethical challenges in the field of health care; consequences on labour markets; and legal aspects of 'autonomous vehicles'. The reports were accompanied by Recommendations to the Committee of Ministers and Resolutions. The report on the need for democratic governance of artificial intelligence proposed, in particular, that the Committee of Ministers supports the drafting of a legally binding instrument governing AI applications, possibly in the form of a Convention⁵⁸.

4.9 The work of the Commissioner for Human Rights

54. The Commissioner for Human Rights issued on 14 May 2019 a Recommendation "Unboxing artificial intelligence: 10 measures to protect human rights⁵⁹". It proposes a series of practical recommendations to national authorities on 10 main areas for action: human rights impact assessment; public consultations; human rights standards in the private sector; information and transparency; independent monitoring; non-discrimination and equality; data protection and privacy; freedom of expression, freedom of assembly and association, and the right to work; avenues for redress; and promoting knowledge and understanding of AI.

4.10. The case law of the European Court of Human Rights relating to information technology

55. The European Court of Human Rights (ECtHR) has not yet developed any specific case law on AI systems, hence the CAHAI could not rely on any ECtHR decisions specifically on AI technology. At the moment there are no known relevant cases pending before the Court either.
56. Existing case law in connection with this topic concerns algorithms in general and violations of Article 8 of the Convention (privacy) or Article 10 (freedom of expression) and, in a more indirect way, Article 14 (non-discrimination) on cases dealing with e.g. mass surveillance⁶⁰, the editorial responsibility of platforms⁶¹ and electoral interference⁶².

⁵⁴ See the Committee of Ministers Recommendation (2018) 7 on Guidelines to respect, protect and fulfil the rights of the child in the digital environment.

⁵⁵ [Division of Culture and Cultural Heritage, Conclusions of the Expert Seminar on Culture, Creativity and Artificial Intelligence, 12-13 October 2018](#) and see also the [proposals for action](#) arising from it.

⁵⁶ [Eurimages, Study on the impact of predictive technologies and AI on the audiovisual sector, including possible specific measures to be put in place to ensure freedom of expression and cultural diversity, December 2019](#)

⁵⁷ [Recommendation 2102\(2017\)](#)

⁵⁸ [Parliamentary Assembly of the Council of Europe, Political Affairs and Democracy Committee, Report on the need for democratic governance of artificial intelligence, Doc. 15150, 24 September 2020](#)

⁵⁹ [Commissioner for Human Rights, Recommendation "Unboxing AI: 10 steps to protect human rights", May 2019](#)

⁶⁰ [ECtHR, Big Brother Watch and others v. the United Kingdom, 13 September 2018 \(Chamber judgment\) - case referred to the Grand Chamber in February 2019](#)

⁶¹ [ECtHR, Delfi AS v. Estonia, 16 June 2015 \(Grand Chamber\)](#)

⁶² [ECtHR Court, Magyar Kétfarkú Kutya Párt v. Hungary, 23 January 2018 - case referred to the Grand Chamber in May 2018](#)

57. In *Sigurður Einarsson and others v. Iceland*⁶³, a prosecuting authority used statistical data processing techniques to process large amounts of information and establish evidence in an economic and financial case. The question raised in this case concerned access by the defence to the data from which incriminating evidence was inferred.
58. Other decisions of the Court have dealt with the consequences of algorithmic mechanisms used to prevent the commission of infringements. In 2006, the Court held its *Weber and Saravia v. Germany* judgment⁶⁴ that any potential abuse of the state's supervisory powers was subject to adequate and effective safeguards and that, in any event, Germany had a relatively wide margin of appreciation in the matter.
59. With regard to mass surveillance of the population using algorithms, which could potentially include AI tools, two potentially relevant cases are pending before the Grand Chamber: *Centrum För Rättvisa v. Sweden*⁶⁵ and *Big Brother Watch and others v. the United Kingdom*⁶⁶. The last hearings in these cases took place on 10 July 2019.

5. MAPPING OF INSTRUMENTS APPLICABLE TO ARTIFICIAL INTELLIGENCE

5.1 International legal instruments applicable to artificial intelligence

59. General international and regional human rights instruments, including the ECHR, are applicable in all areas of life and are therefore also applicable in the context of AI systems. The question is, however, whether these instruments, separately or applied together, can sufficiently meet the challenges posed by AI systems and ensure adherence to the Council of Europe's standards on human rights, democracy and the rule of law throughout their life cycle. Currently, no international legal instrument exists that specifically applies to the challenges raised by AI systems – or by automated decision making more generally – for democracy, human rights and the rule of law in a comprehensive way. There are, however, a number of international legal instruments that partially deal with certain aspects pertaining to AI systems indirectly.
60. In this regard, the CAHAI took note, during its 2nd plenary meeting, of the analysis of relevant international binding instruments made by an independent consultant.⁶⁷ This analysis was based on a review of binding and non-binding instruments in four core areas (data protection, health, democracy and justice) and was complemented by an overview of the Council of Europe's instruments in other fields. It noted that various international legal instruments already exist to safeguard human rights more generally⁶⁸, to safeguard the rights of specific groups in light of vulnerabilities that are also relevant in an AI context⁶⁹, and to safeguard specific human rights that can be impacted by AI. The latter

⁶³ [ECtHR, *Sigurður Einarsson and Others v. Iceland*, 4 June 2019 \(2nd section\)](#)

⁶⁴ [ECtHR, Dec. 29 June 2006, *Weber and Saravia v. Germany*, no. 54934/00](#)

⁶⁵ [ECtHR, Dec. 19 June 2018, *Centrum För Rättvisa v. Sweden*, no. 35252/08 referred back to the Grand Chamber in February 2019 - hearing held on 10 July 2019](#)

⁶⁶ [ECtHR, Dec. 13 September 2018, *Big Brother Watch and others v. the United Kingdom*, nos. 58170/13, 62322/14 and 24960/15 referred back to the Grand Chamber in February 2019 - hearing held on 10 July 2019](#)

⁶⁷ See CAHAI (2020)08-fin, Analysis of internationally legally binding instruments, report by Alessandro Mantelero, University of Turin.

⁶⁸ Such as e.g. the European Convention on Human Rights (ETS No. 5) and its Protocols; the European Social Charter (ETS No. 163); the International Covenant on Civil and Political Rights; the International Covenant on Economic, Social and Cultural Rights; and the EU Charter of Fundamental Rights.

⁶⁹ See e.g. the Convention on the Rights of the Child and the Convention on the Rights of Persons with Disabilities. See also the European Charter for Regional or Minority Languages (ETS No. 148) which can indirectly help ensure attention to minority languages when developing AI-applications.

encompass, for instance, the right to non-discrimination⁷⁰ and the right to the protection of privacy and personal data⁷¹, particularly in the context of automated personal data processing.

61. Of particular importance is the Protocol amending the original Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, which was already mentioned above. This protocol not only modernised but also aligned this Convention more closely to the EU General Data Protection Regulation. It introduced, for instance, requirements of fairness and transparency, as well as protective rights for data subjects who are subjected to automated decision-making processes. This Protocol, however, has not entered into force yet.⁷²
62. Furthermore, in addition to horizontally applicable instruments, a number of international legal instruments deal with specific sectors or domains that may indirectly pertain to AI or automated decision-making processes. These instruments cover areas as diverse as cybercrime⁷³, (bio)medicine⁷⁴ and aviation.⁷⁵ Finally, some legal instruments concern procedural rights – such as transparency⁷⁶ and access to justice⁷⁷ – that might be helpful to monitor and safeguard the protection of substantive rights, or to address aspects relating to liability for certain harms.⁷⁸
63. The CAHAI-PDG acknowledges that these different legal instruments may be relevant in the context of AI regulation. However, the CAHAI-PDG also supports the conclusions drawn in the analysis that these instruments do not always provide adequate safeguards to the challenges raised by AI systems. This will be the subject of further analysis under sub-section iv) below.
64. The growing need for a more comprehensive and effective governance framework to address the new challenges and opportunities raised by AI has been acknowledged by a number of intergovernmental actors at international level. To date, most of these initiatives have been limited to non-binding recommendations.⁷⁹ It is worth mentioning that the European Commission has announced the preparation of a legislative proposal to tackle fundamental rights challenges related to ensuring trustworthy AI, which is scheduled for publication in the first quarter of 2021.⁸⁰

5.2 Ethics Guidelines applicable to artificial intelligence

⁷⁰ See e.g. the International Convention on the Elimination of All Forms of Racial Discrimination, the Convention on the Elimination of All Forms of Discrimination against Women, and the Convention on Cybercrime and its Additional Protocol.

⁷¹ See e.g. the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (ETS No. 108), the EU General Data Protection Regulation (2016/679) and the EU Law Enforcement Directive (2016/680).

⁷² The protocol will only enter into force when ratified by all Parties to Treaty ETS 108, or on 11 October 2023 if there are 38 Parties to the Protocol at this date. Protocol (ETS No. 223) amending the Convention (ETS No. 108) for the Protection of Individuals with regard to Automatic Processing of Personal Data.

⁷³ See e.g. the Convention on Cybercrime (ETS No. 185). As regards the EU, see e.g. the Cybersecurity Act (Regulation 2019/881) and the NIS Directive (2016/1148).

⁷⁴ See e.g. the Convention on Human Rights and Biomedicine (ETS No. 164). See also the EU's Medical Device Regulation (2017/745) and Regulation on in vitro diagnostic medical devices (2017/746).

⁷⁵ See e.g. the Chicago Convention on International Civil Aviation.

⁷⁶ See e.g. the Council of Europe Convention on Access to Official Documents (ETS No. 205).

⁷⁷ See e.g. the European Convention on the Exercise of Children's Rights (ETS No. 160) and the European Convention on Mutual Assistance in Criminal Matters (ETS No. 30).

⁷⁸ See for instance the European Convention on Products Liability in regard to Personal Injury and Death (ETS No. 91) and the European Union's Product Liability Directive and Machinery Directive.

⁷⁹ For instance, the OECD adopted a Council Recommendation on AI listing a number of ethical principles (see <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>), which provided inspiration for the human-centered AI principles endorsed by G20 in a Ministerial Statement in June 2019 (see <https://www.mofa.go.jp/files/000486596.pdf>). Also UNESCO is preparing a (non-binding) Recommendation on ethical AI (see UNESCO, First Draft of the Recommendation on the Ethics of Artificial Intelligence, September 2020, <https://unesdoc.unesco.org/ark:/48223/pf0000373434>.) While UNESCO's current draft mention AI's impact on human rights and the rule of law, it does not focus on AI's challenges to democracy.

⁸⁰ The European Commission particularly emphasises risks for fundamental rights, safety and the effective functioning of the liability regime. See the EC White Paper on Artificial Intelligence, published in February 2020, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

65. In recent years, private companies, academic and public-sector organisations have issued principles, guidelines and other soft law instruments for the ethical use of AI⁸¹. In this regard, the CAHAI took note, during its 2nd plenary meeting, of the mapping work by two independent consultants⁸² who reviewed 116 documents on “ethical AI”, primarily developed in Europe, North America and Asia. This mapping revealed that current AI ethics guidelines converge on some generic principles, but they sharply disagree over the details of what should be done in practice. Notably as regards transparency, the most frequently identified principle, it was not clear whether transparency should be achieved through publishing source code, the algorithmic training data (while considering the applicable data protection laws) or some other means. Thus, resolving the challenge of applying these principles in practice and considering potential interdependencies and trade-offs with other desirable properties was considered an important issue to be addressed by policy makers.
66. According to the mapping, compared to the rest of the world, soft law documents produced within Council of Europe’s member States appear to place greater emphasis on the ethical principles of solidarity, trust and trustworthiness, and refer more sporadically to the principles of beneficence and dignity. The principles of privacy, justice and fairness showed the least variation across Council of Europe’s member States, observers and the rest of the world, and hence the highest degree of cross-geographical and cross-cultural stability.
67. In terms of key policy implications, it was noted that ethics guidelines are useful tools to exert some influence on public decision making over AI and to steer its development towards social good. However, it was also underlined that soft law approaches cannot substitute mandatory governance. Due to the fact that the interests of those developing and commercialising the technology and those who might suffer negative consequences thereof are not fully aligned, there is a particular risk that self-regulation by private actors can bypass or avoid mandatory governance by (inter)governmental authorities. Soft law instruments can however play an important role in complementing mandatory governance.
68. The CAHAI-PDG agrees with the general findings of the mapping study and finds that the common principles identified in the study on relevant ethics guidelines could be considered for inclusion in the legal framework to be developed by the CAHAI. Respect for human rights, which was mentioned in just over half of the soft law documents reviewed, should be the focus of any future legal instrument on AI based on the Council of Europe’s standards. In addition, the mapping study could be used as a practical foundation for implementing ethical frameworks in Member States in a harmonized fashion.

5.3 Overview of national instruments, policies and strategies related to artificial intelligence

69. The analysis of the electronic consultation carried out among CAHAI members, observers and participants on this issue⁸³ indicated that three member and two observer States have adopted specific legal frameworks on AI systems. In all three member States, these frameworks concern the testing and use of autonomous cars. Two member States are developing legal frameworks on the use of AI systems in the fields of recruitment and automated decision making by public authorities.
70. Domestic ethics charters and soft law documents appear to be more widespread and regard issues such as robotics, facial recognition, the use of “ethical AI” in the public service and in electoral processes, and the use of personal and non-personal data. In one member State, a voluntary AI certification programme was launched. Two member States have formally endorsed international or European non-binding AI ethics frameworks. A total of 11 member and four observer States have adopted one or more of the above-mentioned instruments. Different types of institutions such as

⁸¹ Amongst recent initiatives feature the Ethics Guidelines for Trustworthy AI published in April 2019 by the Independent High-Level Expert Group on Artificial Intelligence, set up by the European Commission, and its “Assessment List for Trustworthy AI” (ALTAI) for self-assessment published in July 2020.

⁸² See CAHAI (2020)07-fin, AI Ethics Guidelines: European and Global Perspectives, report prepared by Marcello Ienca and Effy Vayena.

⁸³ See the document CAHAI (2020) 09 rev 2, on the electronic consultation of CAHAI members, observers and participants, which includes replies until 30 September 2020.

national councils, committees, public institutions specialised in AI and government entities have been responsible for their development.

71. Strategies and policies on AI systems have been put in place in thirty member and four observer States. Built on multi-annual action plans, accompanied in some cases by ambitious funding programmes, they pursue the objectives of increasing the trust in this technology and promoting its uptake, strengthening skills for its design and development, supporting research and boosting business development. States have very often involved experts from the public and private sectors, as well as academia, in the preparation of these plans. In most cases, AI systems are the subject of targeted strategies, whilst in other cases they have been integrated into broader sector policies concerning the economy and digital technologies. The development and use of AI systems has also been considered in sectorial strategies on agriculture, e-justice, public services, health, environment, education, security and defence, mobility and data.
72. Finally, the need to promote the development of AI in line with ethical requirements and international human rights standards has been underlined in seven national strategies.

5.4 Advantages, disadvantages and limitations of existing international and national instruments and ethical guidelines on artificial intelligence

73. The above overview has shown that a number of more broadly applicable provisions already extend to the development and use of AI systems. In the absence of a specific legal response in terms of international legally-binding instruments focused on AI, significant efforts have been put into interpreting existing legal provisions in the light of AI, and/or in formulating non-binding rules to contextualise the principles embedded in existing instruments.⁸⁴ However, the fact that existing legal instruments have been adopted prior to the wide-spread use of AI systems often tends to reduce their effectiveness to provide an adequate and specific response to the challenges brought by AI systems, as they are not tailored to its specificities. For instance, an ECRI study on "[Discrimination, artificial intelligence, and algorithmic decision-making](#)" has highlighted the limitations of existing international and domestic legal instruments in the field of non-discrimination.⁸⁵ The independent expert's analysis prepared for the CAHAI regarding the impact of AI on human rights, democracy and the rule of law provided similar conclusions as regards other rights.⁸⁶
74. Furthermore, despite being overlapping and mutually reinforcing, the number and diversity of instruments render it difficult to interpret and apply them to the AI context in a consistent and comprehensive manner, leading to uneven protection levels. While certain soft law instruments (e.g. ethics guidelines) set out more tailored principles on the development and use of AI systems, these are non-binding and hence limited in their effectiveness. Moreover, they do not clearly define binding obligations for member States and private actors with regards to the respect of human rights, democracy and the rule of law. Furthermore, ethics guidelines do not have the same universal dimension as human rights-based standards and are characterised by a variety of theoretical approaches⁸⁷, which limits their utility. The CAHAI-PDG therefore notes that, while there is no legal vacuum as regards AI regulation, a number of substantive and procedural legal gaps nevertheless exist.⁸⁸

⁸⁴ See for instance T-PD(2019)01 Guidelines on Artificial Intelligence and Data Protection; CEPEJ. 2019. European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment.

⁸⁵ These instruments do not apply if an AI system invents new classes which do not correlate with protected characteristics under such instruments (i.e. gender or race), to differentiate between people. Such differentiation can nevertheless be unfair. By way of illustration, AI-driven price discrimination could lead to certain groups in society consistently paying more.

⁸⁶ See CAHAI(2020)06-fin, report prepared by Catelijne Muller.

⁸⁷ As pointed out in the independent expert's report CAHAI (2020)08-fin, cited above.

⁸⁸ Note in this regard also the text adopted by the Standing Committee, acting on behalf of the Assembly, on 22 October 2020 (see Doc. 15150, report of the Committee on Political Affairs and Democracy,

75. First, the rights and obligations formulated in existing legal instruments tend to be articulated too broadly or generally to secure their effective application to the challenges raised throughout the life cycle of an AI systems. It has been indicated that a translation or concretisation of existing human rights to the context of AI systems⁸⁹, through more specific provisions, could help remedy this issue.⁹⁰ This could be done by specifying more concrete rights that fall under a broader human right and that could be invoked by those subjected to AI systems. For instance, the *right to a fair trial* could be further concretised in terms of a right to challenge and get insight into any evidence based on an AI system. This could also be done by deriving specific obligations that should be complied with or requirements that should be met by those who develop or deploy AI systems. For instance, the *right to non-discrimination* could be further concretised in terms of a due diligence obligation to mitigate, throughout AI systems' life cycle, the risk of unjust bias. Without such concretisation of existing rights in the context of impact by AI applications, individuals may fail to obtain the full and effective protection thereof.⁹¹ The CAHAI-PDG believes that Council of Europe's standards on human rights, democracy and the rule of law could provide an appropriate basis for the elaboration of more specific provisions to secure effective protection against the risks posed by the practical application of certain AI systems.
76. Secondly, a number of essential principles that are relevant for the protection of human rights, democracy and the rule of law in the context of AI, are currently not legally assured. These gaps concern, for instance, the necessity to ensure *human control and oversight*⁹² over AI applications, to ensure their *technical robustness*, and to secure their *transparency*⁹³ and *explainability*⁹⁴, in particular when they produce legal or other significant effects on individuals. A lack of enough comprehensive legal provisions in existing legal instruments to safeguard these principles was pointed out in several studies.⁹⁵ Importantly, safeguarding these principles is also a necessary precondition to safeguard substantive rights, given the opacity of AI systems and of the human choices to design and use them.⁹⁶

⁸⁹ As it is done by European General Data Protection Regulation with regard to the protection of personal data.

⁹⁰ See CAHAI(2020)06-fin and CAHAI (2020)08-fin, cited above. See also Karen Yeung, Andrew Howes, and Ganna Pogrebna (University of Birmingham), 'AI Governance by Human Rights-Centered Design, Deliberation, and Oversight: An End to Ethics Washing', in *The Oxford Handbook on Ethics of AI* (eds. M. D. Dubber, F. Pasquale, and S. Das), 2020, DOI: 10.1093/oxfordhb/9780190067397.013.5; Nathalie A. Smuha (KU Leuven), 'Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea', in *Philosophy and Technology*, 2020, <https://doi.org/10.1007/s13347-020-00403-w>.

⁹¹ Moreover, those who are responsible to apply and interpret existing rights (for instance judges) may lack guidance as to how to do so in the context of AI, potentially leading to unequal / inadequate standards of protection depending on the jurisdiction.

⁹² This could also include provisions to minimise the risks that may arise through the unqualified tampering or interference with AI systems.

⁹³ As regards the transparency of automated decision-making processes, it should be noted that the limited protection offered by Convention 108+ in this regard only applies to the processing of personal data, while AI systems can negatively affect individuals and societies also based on non-personal data. Moreover, the right not to be subjected to solely automated decision-making is currently formulated very restrictively, as it only applies when it can be proven that an individual is significantly impacted by the decision, and if it can be proven that the decision was taken 'solely' by an AI system. Hence, the risk exists that a very limited review by a human being – even if subjected to automation bias, severe time pressure or lack of information when reviewing the decision – makes this right moot.

⁹⁴ Convention 108+ does not contain a right to an explanation for the data subject, and it is highly contested whether and to which extent the EU General Data Protection Regulation embodies this right. See e.g. Sandra Wachter et al., 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation', 7 *International Data Privacy Law* 2, 76-99 (2017), doi:10.1093/idpl/ix005.

⁹⁵ See CAHAI(2020)06-fin and CAHAI (2020)08-fin, cited above. See also Automating Society Report 2020, AlgorithmWatch, <https://automatingsociety.algorithmwatch.org/report2020/european-union/>; Moreover, see also the European Commission White Paper on AI, 19 February 2020, COM(2020) 65 final, at p 9: "A key result of the feedback process is that, while a number of the requirements are already reflected in existing legal or regulatory regimes, those regarding transparency, traceability and human oversight are not specifically covered under current legislation in many economic sectors".

⁹⁶ It is only when the traceability of AI is ensured, for instance through the documentation or logging of relevant information, that a system can be audited and that it can be verified to which extent it may for instance infringe the right to non-discrimination. Furthermore, the lack of explanation of the decision-making process hinders the possibility for individuals to challenge a decision and seek redress. In this regard, the European Commission White Paper on AI noted more generally, at p12, that "the specific characteristics of many AI technologies, including

Without the transparency or explainability of an impactful AI-enabled decision, it cannot be assessed whether a right – such as the right to non-discrimination – is actually ensured. Moreover, it hinders an individual's capability to challenge the decision. The existence of asymmetries of information between those negatively impacted by AI systems and those developing and using them also stresses the need to reinforce mechanisms of *accountability* and *redress*, and to render AI systems *traceable* and *auditable*. If these gaps are not filled, for instance by securing the protection of these principles through the establishment of concrete rights and obligations, those negatively impacted – as well as other stakeholders, including regulators and law enforcers – will not be able to assess the existence of human rights and other infringements.

77. Current instruments also lack sufficient attention to the steps that developers and deployers or AI systems should take to ensure the *effectiveness* of these systems whenever they can impact human rights, democracy or the rule of law, and to ensure that AI developers and deployers have the necessary *competences or professional qualifications* to do so. Moreover, the *societal dimension* of AI's risks that surpasses the impact on individuals, such as the impact on the electoral process and the democratic institutions or the legal system, is not yet sufficiently considered. While a number of national and international mechanisms allow individuals to seek redress before a court when a human right is breached in the context of AI, this mechanism is currently underdeveloped as regards an interference with democracy or the rule of law, which concern broader societal issues. Their protection necessitates public oversight over the responsible design, development and use of AI systems whenever such risks exist, by setting out clear obligations or requirements to this end.
78. These legal gaps also lead to uncertainty for stakeholders, and in particular AI developers, deployers and users, who lack a predictable and sound legal framework in which AI systems can be designed and implemented. This uncertainty risks hampering beneficial AI innovation, and can hence stand in the way of reaping the benefits provided by AI for citizens and society at large. A comprehensive legal framework for AI systems, guided by a risk-based approach, can help provide the contours in which beneficial innovation can be stimulated and enhanced, and AI's benefits can be optimised, while ensuring – as well as maximising – the protection of human rights, democracy and the rule of law via effective legal remedies.
79. Finally, the various gaps in existing legal instruments - as well as the fragmented approach of applying these instruments to the context of AI across Europe - also raise uncertainty as regards the manner in which the transboundary nature of the impact generated by the development and use of AI applications can be tackled. A lack of common norms at international level might hamper the cross-border trade of AI products and services, as the lack of shared norms can stand in the way of mutual trust, hence potentially also preventing that the benefits of AI applications can reach across national borders.
80. Based on the above, it can be concluded that a regulatory approach to AI should aim to address those gaps. Beyond existing legal frameworks, such an approach could contain binding provisions to safeguard human rights, democracy and the rule of law in the context of AI, to ensure a more comprehensive level of protection regardless of the sector concerned.⁹⁷ This can be done by clarifying or broadening the scope of existing rights and/or obligations and mandating the protection of additional principles or requirements to this end. In addition to such a binding approach, consideration can also

opacity ('black box-effect'), complexity, unpredictability and partially autonomous behaviour, may make it hard to verify compliance with, and may hamper the effective enforcement of, rules of existing EU law meant to protect fundamental rights". This also applies to the human rights provisions in other existing legal instruments, as they are currently not tailored to the specific challenges raised by AI. However, several examples of "supplementary models" and other methods for understanding how a decision has been reached do exist. Supplementary models are becoming more common, as is the use of more interpretable AI systems (see <https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-ai/>).

⁹⁷ In this regard, it can be noted that many AI systems can be repurposed for use in other sectors. Therefore, an approach that sets out certain safeguards across sectors can be desirable, potentially coupled with complementary safeguards or guidelines that are more sector-specific where needed.

be given to the elaboration of sector-specific guidance and ethical guidelines for issues that are only or particularly relevant in a given field or application.⁹⁸ In this regard, reference can be made to Chapter 8 of this Feasibility Study, which sets out various options for a Council of Europe legal framework for the design, development and application of AI.

81. As mentioned in the CAHAI progress report, the work undertaken by the CAHAI provides an opportunity to contribute and complement other international initiatives in this area (e.g. by the OECD, the European Union, the UNESCO and the United Nations in general, with whom coordination and synergies are being sought on a regular basis⁹⁹) by enacting an instrument based on the Council of Europe's standards on human rights, the rule of law and democracy, as part of a global legal mechanism for the regulation of digital technologies. In this regard, the CAHAI-PDG underlined that part of the added value that the Council of Europe can provide when elaborating a legal instrument on AI is that, besides the protection of human rights, it can also address the societal and environmental challenges posed by AI to democracy and the rule of law.¹⁰⁰ Developing a legally-binding instrument based on Council of Europe standards— should this option be supported by the CAHAI – would contribute to making the CAHAI initiative unique among other international initiatives, which either focus on elaborating a different type of instrument or have a different scope or background.

5.5 International legal instruments, ethical guidelines and private actors

82. Council of Europe instruments are typically addressed to the member States rather than to private actors. Nevertheless, private actors can be addressed indirectly, by virtue of the rights granted to, and obligations assumed by, states under such instruments. States have a duty¹⁰¹ to ensure that private actors act in line with certain human rights and freedoms by implementing and enforcing them in their national laws, and by making sure that effective legal remedies are available at national level. Additionally, private actors, in line with the UN Guiding Principles on Business and Human Rights, have a corporate responsibility to respect the human rights of their customers and all stakeholders.¹⁰²
83. A number of international instruments directly focus on the need for businesses to comply with human rights and ensure responsible technological research and innovation.¹⁰³ Over the past years, private actors have shown a strong interest in advancing the responsible development and use of AI systems, acknowledging not only the opportunities but also the risks raised thereby. Private actors have not only contributed to the proliferation of AI ethics guidelines, but some also explicitly argued in favour of a regulatory framework to enhance legal certainty in this domain.¹⁰⁴

⁹⁸ In this regard, the CAHAI-PDG recognized the context-specificity of certain risks. The wide-scale use of AI-based remote biometric identification, for instance, does not raise the same impact on human rights as the use of an AI-based system to recommend a song.

⁹⁹ During its second plenary meeting, the CAHAI heard updates from the FRA, the European Union, the OECD, the United Nations High Level Panel on Digital Co-operation and UNESCO. See the report of the second plenary meeting of the CAHAI, paragraphs 78-84.

¹⁰⁰ It can be noted that, while the European Commission White Paper on AI focuses on the impact of AI on fundamental rights, it does not explicitly address AI's impact on democracy and the rule of law.

¹⁰¹ UN Guiding Principles: States have a duty to protect against human rights abuse within their territory and/or jurisdiction by third parties, including business enterprises.

¹⁰² See Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic system), <https://rm.coe.int/09000016809e1154>. See also Recommendation CM/Rec(2016)3 of the Committee of Ministers to Member States on human rights and business, at <https://rm.coe.int/human-rights-and-business-recommendation-cm-rec-2016-3-of-the-committee/16806f2032>.

¹⁰³ Most notably the UN Guiding Principles on Business and Human Rights, particularly Art. 18 and 19. See also the OECD Due Diligence Guidelines for Multinational Enterprises and the OECD Due Diligence Guidelines for Responsible Business Conduct.

¹⁰⁴ Besides the statements of individual companies, such as e.g. Microsoft (<https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/>) or IBM (<https://www.ibm.com/blogs/policy/ai-precision-regulation/>), the Policy Recommendations of the European Commission's High-Level Expert Group on AI, including over 20 companies, ask to consider the adoption of new legislation, e.g. at p. 40: "For AI-systems deployed by the private sector that have the potential to have a significant impact on human lives, for example by interfering with an individual's fundamental rights at any stage of the AI

84. Should a regulatory approach that combines a binding instrument with soft law tools be supported by the CAHAI, private actors, civil society organisations, academia and other stakeholders would have an important role not only in assisting states in the development of a binding legal instrument, but also in contributing to the development of sectorial soft law instruments that can complement as well as aid in the implementation of the binding provisions in a context-specific manner (for instance through sectorial guidelines, certifications and technical standards). An effective regulatory framework for AI systems will require close co-operation between all stakeholders, from states and public entities who must secure public oversight, private actors who can contribute their knowledge and secure socially beneficial AI innovation, and civil society organisations who can represent the interests of the public at large. The CAHAI-PDG acknowledges that the Council of Europe is uniquely positioned to lead this effort and – by building further on existing frameworks – to guide the alignment of AI systems with its standards on human rights, democracy and the rule of law.

6. MAIN CONCLUSIONS OF THE MULTI-STAKEHOLDER CONSULTATIONS

The multi-stakeholder consultation is planned to take place in 2021, under the aegis of the Working Group on Consultations and Outreach (CAHAI-COG), which is currently working in close co-operation with the CAHAI-PDG to determine the scope, the target groups and the modalities of the consultation. The CAHAI will take a decision on these issues during its third plenary meeting in December 2020. The findings of the consultation, which could feed the work of elaboration of the main elements of a legal framework that the CAHAI is mandated to develop, will be first reviewed by the CAHAI and then presented to the Committee of Ministers as part of the process of reporting of CAHAI activities.

7. MAIN ELEMENTS OF A LEGAL FRAMEWORK FOR THE DESIGN, DEVELOPMENT AND APPLICATION OF ARTIFICIAL INTELLIGENCE

7.1 Key values, rights and principles deriving - in a bottom-up perspective - from sectoral approaches and ethical guidelines; in a top-down perspective - from the requirements of human rights, democracy and the rule of law.

86. In line with the CAHAI's mandate, a legal framework on AI should ensure that the development, design and application of this technology complies with the Council of Europe's standards on human rights, democracy and the rule of law. Following a risk-based approach, it should provide an enabling regulatory setting in which beneficial AI innovation can flourish, all the while addressing the risks set out in Chapter 3, and the substantive and procedural legal gaps identified in Chapter 5, to ensure both its relevance and effectiveness amidst existing instruments.

87. This can be done by formulating key principles that must be secured in the context of AI and, on that basis, identifying concrete rights that individuals can invoke (whether existing rights or newly tailored rights to the challenges and opportunities raised by AI) as well as requirements that developers and deployers of AI systems should meet. In addition, Chapter 3 has shown that – on a context-specific basis – green lines can be considered for AI systems that promote, strengthen and augment the protection of human rights, democracy and the rule of law, and that red lines (the latter in the form of strong restrictions for the exceptional and controlled use of the application, and where essential, a (temporary) ban or moratorium) should be considered for AI systems or applications that pose significant risks for human rights, democracy and the rule of law. In what follows, the main principles are discussed which should be considered for inclusion in a potential Council of Europe legal instrument

system's life cycle and for safety-critical applications, consider the need to introduce: a mandatory obligation to conduct a trustworthy AI assessment (including a fundamental rights impact assessment which also covers for example the rights of children, the rights of individuals in relation to the state, and the rights of persons with disabilities) and stakeholder consultation including consultation with relevant authorities; traceability, auditability and ex-ante oversight requirements; and an obligation to ensure appropriate by default and by design procedures to enable effective and immediate redress in case of mistakes, harms and/or other rights infringement". It also stresses the need for legal certainty.

on AI, including the concrete rights and obligations attached thereto.¹⁰⁵ While these principles, rights and requirements are described in a horizontally applicable manner, as noted above, they could be combined with a sector-specific approach that provides (more detailed) contextual requirements in the form of soft law instruments, such as sectoral guidelines or assessment lists.

7.1.1 *Human dignity*

88. Human dignity is the foundation of all human rights. It recognises that all individuals are inherently worthy of respect by mere virtue of their status as human beings. Human dignity is inviolable. Hence, even when a human right is restricted – for instance when a balance of rights and interests must be made – human dignity must always be safeguarded. In the context of AI, this means that the design, development and use of AI systems must respect the dignity of the human beings interacting therewith or impacted thereby. Humans should be treated as moral subjects, and not as mere objects that are categorised, scored or manipulated.
89. AI applications can be used to foster human dignity and empower individuals, yet their use can also challenge it and (un)intentionally run counter it. To safeguard human dignity, it is essential that human beings are aware of the fact that they are interacting with an AI system and are not misled in this regard. Moreover, they should in principle be able to choose not to interact with it, and to not be subject to a decision by an AI system whenever this can significantly impact their lives. Furthermore, the allocation of certain tasks may need to be reserved for humans rather than machines given their potential impact on human dignity. More generally, AI systems should be developed and used in a way that secures and promotes the physical and mental integrity of human beings.

❖ Key substantive rights:

- The right to human dignity, the right to life, and the right to physical and mental integrity.
- The right to know that one is interacting with an AI system rather than with a human being, including the right to be informed of the fact that one is interacting with an AI whenever confusion may arise.

❖ Key obligations:

- Where tasks will negatively affect human dignity if carried out by machines rather than human beings, these tasks should be reserved for humans.
- AI deployers should inform human beings of the fact that they are interacting with an AI system rather than with a human being whenever confusion may arise
- Where certain uses of AI systems, in certain contexts, are deemed incompatible with human dignity, red lines should be drawn.
- Where tasks that can promote and enhance human life or dignity are demonstrably better performed by AI systems than by human beings, green lines should also be drawn.

7.1.2 *Human Freedom and Human Autonomy*

90. Human freedom and autonomy are core values which are reflected in various human rights of the ECHR. In the context of AI, they refer to the ability of humans to act self-determinedly, by deciding in an informed and autonomous manner on the use of AI systems and on the consequences thereof on themselves and others. This also includes the decisions if, when and how to use AI systems. As noted in Chapter 3, human freedom and autonomy can be impacted by AI in different ways, such as by AI-driven (mass) surveillance or targeted manipulation – whether by public or private entities – for instance through the use of remote biometric recognition or online tracking.

¹⁰⁵ The principles in this chapter are derived from the identified principles in the CAHAI (2020)07-fin report of M. Ienca and E. Vayena and from further CAHAI-PDG discussions. They are not stated in any specific order.

91. In general, AI systems should not be used to subordinate, coerce, deceive, manipulate or condition humans, but rather to complement and augment their capabilities. Human oversight mechanisms must be established, ensuring that human intervention is possible whenever needed to safeguard human rights, democracy and the rule of law. As noted in Chapter 5, the establishment of adequate human oversight mechanisms is not yet secured by law. The extent and frequency of oversight should be tailored to the specific AI application context¹⁰⁶ and the autonomy of such human interventions should be preserved¹⁰⁷. It must, however, be ensured that when human intervention is required, for instance to avoid that a decision is taken solely by an AI system, this occurs by someone with the truly autonomous ability to override the system's decision¹⁰⁸ (without hindrance of automation bias or lack of time for review).¹⁰⁹

❖ Key substantive rights:

- The right to human autonomy and self-determination, and the right to liberty. The right not to be subject to a decision based solely on automated processing when this produces legal effects on or similarly significantly affects individuals.
- The right to effectively contest and challenge decisions informed and/or made by an AI system and demand that such decision be reviewed by a person (right to opt out).

❖ Key obligations:

- Member States should require AI developers and deployers to establish appropriate human oversight mechanism that safeguard human autonomy:
 - An adequate level of human involvement should be ensured in the operation of AI systems, based on a contextual risk assessment taking into account the system's impact on human rights, democracy and the rule of law.
 - Whenever necessary and possible, based on a thorough risk assessment, a qualified human being should be able to disable any AI system or change its functionality.
 - Those developing and operating AI systems should have the adequate competences or qualifications to do so, to ensure appropriate oversight that enables the protection of human rights, democracy and the rule of law.

○

7.1.3 Non-Discrimination, Gender Equality, Fairness and Diversity

92. As noted in Chapter 3, AI systems can negatively impact the right non-discrimination and the right to equal treatment and equality. Various studies have pointed to the fact that the use of these systems can perpetuate and amplify discriminatory or unjustifiable biases and harmful stereotypes, which has an adverse impact not only on the individuals subjected to the technology, but on society as a whole.¹¹⁰

¹⁰⁶ See in this regard the distinction made between a human-on-the-loop (HOL), human-in-the-loop (HIL) and human-in-command approach (HIC) in the Ethics Guidelines for Trustworthy AI at p. 16, accessible at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419.

¹⁰⁷ E.g. by ensuring – where appropriate and feasible – that the person, who intervenes should not know the decision taken by the machine.

¹⁰⁸ Regarding the overreliance on the solutions provided by AI applications and fears of challenging decisions suggested by AI applications, which risk altering the autonomy of human intervention in decision-making processes see also the *T-PD Guidelines on AI and Data Protection (T-PD(2019)01)* where it is said that « The role of human intervention in decision-making processes and the freedom of human decision makers not to rely on the result of the recommendations provided using AI should therefore be preserved. »

¹⁰⁹ Care must be taken that to ensure that the 'human in the loop' does not become a moral or legal 'crumple zone', which can be used to describe how responsibility for an action may be misattributed to a human actor who had limited control over the behaviour of an automated or autonomous system.

¹¹⁰ See e.g. the CoE study by F. Zuiderveen Borgesius, *Discrimination, artificial intelligence, and algorithmic decision-making*, 2018, at: <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>; Joy Buolamwini, Timnit Gebru; Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:77-91, 2018. See also CAHAI-PDG 1st meeting report, p.5.

Indeed, reliance on unjustly biased AI systems could increase inequality, thereby threatening the social cohesion and equality required for a thriving democracy.

93. While the right to non-discrimination and equality is already set forth in numerous international legal instruments, as noted in Chapter 5, it requires contextualisation to the specific challenges raised by AI so that its protection can be secured. In particular, the increased prominence of proxy discrimination in the context of machine learning may raise interpretive questions about the distinction between direct and indirect discrimination or, indeed, the adequacy of this distinction as it is traditionally understood. Similarly, there may be interpretive questions about the meaning of traditional justifiability standards for discrimination in the context of machine learning. Special attention should be given to the impact of the use of AI systems on gender equality, given the risk that gender-based discrimination, gender stereotypes and sexism might be (inadvertently) perpetuated thereby. Caution is also needed for the potential amplification of discrimination against vulnerable groups more generally. The current lack of diversity among the people developing and making decisions in the AI sector is a source of concern, and diverse representation in consultative processes regarding AI system applications in sensitive areas should be encouraged. This would help prevent and mitigate adverse human rights impacts, notably in relation to equality and non-discrimination. It is equally important to consider duly the risk of intersectional discrimination arising from the use of AI systems¹¹¹, as well as discrimination based on new (non-protected) differentiation grounds or erroneous associations.¹¹²

❖ Key substantive rights:

- The right to non-discrimination and the right to equal treatment.
 - The right to non-discrimination on the basis of the protected grounds set out in Article 14 of the ECHR and Protocol 12 to the ECHR, including intersectional discrimination.
 - As AI systems can give rise to unjust discrimination based on new types of differentiation that are not traditionally protected, this right should extend to all unjust differentiation grounds.
 - This right must be ensured in relation to the entire lifecycle of an AI system (design, development, implementation and use), as well as to the human choices around the AI system's use, whether used in the public or private sector.

❖ Key obligations:

- Member States should impose requirements to effectively counter the potential discriminatory effects of AI systems deployed by both the public and private sectors and protect individuals from the negative consequences thereof.
 - These requirements should cover the entire lifecycle of an AI system and should concern, *inter alia*, filling existing gender data gaps, the representativeness, quality and accuracy of data sets¹¹³, the design and optimisation function of algorithms, the use of the system, and adequate testing and evaluation processes to verify and mitigate the risk of discrimination.

¹¹¹ Intersectional discrimination takes place on the basis of several personal grounds or characteristics that operate and interact with each other at the same time in such a way as to be inseparable. Current AI systems are particularly susceptible to such discrimination as they merely look for correlations between different features. A Council of Europe legal framework should take a special interest in this issue, as intersectional discrimination is rarely covered by national discrimination law which tends to focus on one discrimination ground at a time.

¹¹² See e.g. the example in the CoE Study on AI and discrimination cited above, at p.35: "Suppose an AI system finds a correlation between (i) using a certain web browser and (ii) a greater willingness to pay. An online shop could charge higher prices to people using that browser. Such practices remain outside the scope of non-discrimination law, as a browser type is not a protected characteristic."

¹¹³ This could also include the mandatory use of intersectional training data sets, the creation of intersectional benchmarks and the introduction of intersectional audits. The basis to assess whether these requirements are met can also be the results produced by the AI system, which means that access to the training, test and evaluation as such is not always necessary. This requires, however, suitable procedures to enable the meaningful review of the system's results in terms of e.g. representativeness, accuracy and quality.

- the transparency and auditability of AI systems must be ensured to enable the detection of discrimination throughout the lifecycle of an AI system (see below).
- Member States should encourage a gender balance and diversity in the AI workforce and periodic feedback from a diverse range of stakeholders. Awareness of the risk of discrimination and bias in the context of AI should be fostered.
- Member States are obliged to ensure that the AI systems they deploy do not result in unlawful discrimination, harmful (gender) stereotypes and wider social inequality, and should therefore apply the highest level of scrutiny when using or promoting the use of AI systems in sensitive public policy areas, including law enforcement, justice, asylum and migration, health, social security and employment.
- Member States should include non-discrimination and promotion of equality requirements in public procurement processes for AI systems, and ensure that the systems are independently audited for discriminatory effects prior to deployment.
- Member States should encourage the deployment of AI systems where they could effectively counter existing discrimination in human and machine-based decision-making.

7.1.4 Principle of Transparency and Explainability of AI systems

94. A lack of transparency on whether a product or service uses an AI system, and if so, based on which criteria it operates, often renders it difficult or impossible to ascertain whether the system impacts human rights, democracy and the rule of law. Further, without such information, a decision cannot be effectively contested, nor can the system be improved or fixed when causing harm. Transparency is hence crucial to ensuring the enforcement of other principles and rights, including the right to an effective remedy if they are violated, which includes the right to challenge an AI-informed decision and to seek redress. Therefore, the principles of transparency and explainability are essential in the context of AI, especially when a system can impact human rights, democracy or the rule of law. As noted in Chapter 5, however, these principles are not yet adequately protected in existing legal instruments.
95. Transparency entails that AI processes are rendered traceable, for instance by documenting or logging them, and that meaningful information is provided on the system's capabilities, limitations and purpose. This information must be tailored to the context and intended audience. Measures should also be taken to enable the independent and effective audit of AI systems, allowing for a meaningful assessment of its impact. Those affected by a decision solely or significantly informed or made by an AI system should be notified and promptly provided with the aforementioned information. Moreover, they should receive an explanation of how decisions that impact them are reached. While an explanation as to why a system has generated a particular output is not always possible,¹¹⁴ in such a case, the system's auditability should be ensured. While business secrets and intellectual property rights must be respected, they must be balanced against other legitimate interests. Public authorities must be able to audit AI systems (either systematically, randomly or at the request of a party¹¹⁵) to verify compliance with existing legislation.

- ❖ Key substantive rights:
 - The right to be promptly informed that a decision which produces legal effects or similarly significantly impacts an individual's life is informed or made by an AI system.
 - The right to a meaningful explanation of how such AI system functions, what optimisation logic it follows, what type of data it uses, and how it affects one's interests, whenever it generates legal effects or similarly impacts individuals' lives. The explanation must be tailored to the context, and provided in a manner that is useful and comprehensible for an individual, allowing individuals to effectively protect their rights.

¹¹⁴ It should be noted that, in some situations, a higher standard of explainability can only be obtained by reducing the system's performance and accuracy.

¹¹⁵ In a way that is tailored to the specific AI system and context.

❖ Key obligations:

- Obligation to provide adequate communication:
 - When an AI system is used to interact with individuals, in particular in the context of public services, the user should have a right to be assisted by a human being. The user should be clearly informed of this right and of how to request such assistance.
- Requirements regarding traceability and information:
 - Persons with a legitimate interest (e.g. consumers, citizens, supervisory authorities or others) should have easy access to relevant information on AI systems.
 - This information should be comprehensible and could, *inter alia*, include the types of decisions or situations subject to automated processing, criteria relevant to a decision, information on the data used, a description of the method of the data collection. A description of the system's potential legal or other effects should be accessible for review/audit by credible bodies with necessary competences.
- Requirements regarding documentation
 - AI systems that can impact human rights, democracy or the rule of law should be traceable and auditable. The data sets and processes that yield the AI system's decisions, including those of data gathering, data labelling and the algorithms used, should be documented, hence enabling the ex post auditability of the system.
 - Qualitative and effective documentation procedures should be established.

7.1.5 Prevention of harm

96. AI systems can be used in security and protection systems to avoid harm to individuals, to the environment and even to other systems. At the same time, AI systems can also be used in a manner that harms individuals, societies and the environment. The prevention of harm is a fundamental principle that should be upheld, in both the individual and collective dimension, especially when such harm concerns a negative impact on human rights, democracy and the rule of law. Accordingly, human dignity as well as mental and physical integrity must be protected, with adequate safeguards for persons and groups who are more vulnerable. Particular attention must also be paid to situations where AI systems can cause or exacerbate adverse impacts due to asymmetries of power or information, such as between employers and employees, businesses and consumers or governments and citizens. Preventing harm also entails consideration of the natural environment and all living beings, and the manner in which the AI systems can have an adverse impact thereon.
97. Member states should therefore ensure that adequate safeguards are put in place to minimise and prevent harm stemming from the development and use of AI, whether this concerns physical, psychological, economic, social or legal harm. Attention must be given to the safety and security of AI systems, including safeguards for their technical robustness, reliability, and measures that prevent the risk of adversarial attacks or malicious uses. The above-mentioned safeguards are particularly important in the context of public procurement procedures. When implementing measures to prevent harm, member States should pursue a risk-based approach. Moreover, where relevant given the specific circumstances, for instance in case of a high level of uncertainty coupled with a high level of risk, a precautionary approach, including potential prohibitions, should be taken. In addition, member States could also use AI-based safeguards to minimise and prevent harm stemming from the actions of humans.

❖ Key substantive rights:

- The right to life, and the right to physical and mental integrity.
- The right to the protection of the environment.
- The right to sustainability of the community and biosphere.

❖ Key obligations:

- Member States should ensure that developers and deployers of AI systems take adequate measures to minimise any physical or mental harm to individuals, society and the environment.
- Member States should ensure the existence of adequate (by design) safety, security and robustness requirements and compliance therewith by developers and deployers of AI systems.
 - These requirements should include, *inter alia*, resilience to attacks, accuracy and reliability, and the necessity to ensure data quality and integrity. Moreover, AI systems should be duly tested and verified prior to their use as well as throughout the entire life cycle of the AI system including by means of periodical reviews to minimise such risks.
- Member States should ensure that AI systems are developed and used in a sustainable manner, with full respect for applicable environmental protection standards.
- Where relevant, member States could foster the use AI systems to avoid and mitigate harm from the actions of human beings and of other technological systems, while safeguarding the standards of human rights, democracy and the rule of law.

7.1.6 Data protection and privacy

98. Privacy is part of the right to private and family life under Article 8 of the ECHR and is afforded specific protection in the context of the automatic processing of personal data in Convention 108. It is also fundamental to the enjoyment of other human rights. Thus the design, development, training, testing and use of AI systems that rely on the processing of personal data must fully secure a person's right to respect for private and family life under Article 8 of the ECHR, including the "right to a form of informational self-determination" in relation to their data. Individuals should be able to exercise control over their data. Consent – while not the only legal basis to process personal data – is central in this regard. However, in order to be valid, consent needs to be informed, specific, freely given and unambiguous (if not "explicit" when the processing concerns sensitive data). Situations of asymmetry of power or information can affect the freely given requirement of consent, hence implying certain limitations to its protective function in certain situations and the need for a more appropriate legal basis for the processing.
99. Member States should effectively implement the modernised Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data ("Convention 108+") as well as other binding international instruments on data protection and privacy that is binding on the member State. Not all AI systems process personal data. But even where AI systems are not designed to process personal data and instead relying on anonymised, anonymous, or non-personal data, the line between personal data and non-personal data is increasingly blurred. The interplay between personal and non-personal data must hence be further examined, to close any potential legal gaps in protection. Machine learning systems in particular can infer personal information about individuals, including sensitive data, from anonymised or anonymous data, or even from data about other people. In this regard, special consideration must be given to protecting people against inferred personal data.¹¹⁶
100. Finally, regardless of the benefits that the use of a particular AI system could bring, any interference by a public authority with the exercise of the right to privacy shall be in accordance with the law and necessary in a democratic society. To establish whether a particular infringement on this right is "necessary in a democratic society", the European Court of Human Rights has clarified that "necessary" does not have the flexibility of such expressions as "useful", "reasonable", or "desirable", but instead

¹¹⁶ See, for instance, S. Wachter and B. Mittelstadt, A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI, *Columbia Business Law Review*, 2019(2).

implies the existence of a “pressing social need” for the interference in question. It is for national authorities to make the initial assessment of the pressing social need that the use of an AI system could meet in each case, subject to review by the Court. National authorities are encouraged to consult a wide range of stakeholders in the context of this assessment and ensure its periodic review.

Key substantive rights and obligations:

- ❖ Key substantive right:
 - The right to respect for private and family life, and the protection of personal data.
- ❖ Key obligations:
 - Member States must ensure that the right to privacy and data protection are safeguarded throughout the entire lifecycle of AI systems that they deploy, or that are deployed by private actors. The processing of personal data at any stage of an AI system’s lifecycle must be based on the principles set out under the Convention 108+, (including fairness and transparency, proportionality, lawfulness of the processing, quality of data, right not to be subject to purely automated decisions, data security, accountability, privacy by design).
 - When procuring or implementing AI systems, member States should assess and mitigate any negative impact thereof on the right to privacy and data protection as well as on the broader right to respect for private and family life, by particularly considering the proportionality of the system’s invasiveness in light of the legitimate aim it should fulfil, as well as its necessity to achieve it.
 - Member states should develop measures to protect data and AI systems whose benefits can contribute to the promotion and protection of human rights, such as the right to life (for instance in the context of evidence-based medicine).

7.1.7 Accountability and Responsibility

101. Persons developing or deploying AI systems must take responsibility for these systems, and they should be held accountable whenever the principles mentioned above are not respected or any unjust harm occurs to end-users or to others. This means that appropriate mechanisms must be put in place to ensure that AI systems, both before and after their development, deployment and use, comply with the Council of Europe’s standards on human rights, democracy and the rule of law. It is essential that potential negative impacts of AI systems can be identified, assessed, documented and minimised, and that those who report on such negative impacts (e.g. whistle-blowers) are protected. Effective public oversight must be guaranteed, to ensure that AI developers and deployers act responsibly and in compliance with relevant legal requirements, while allowing for intervention by state authorities when it does not happen.

102. In turn, those who might be negatively impacted by AI systems must be ensured an effective remedy against the developers or deployers of AI systems who are responsible. The availability of such remedy should be clearly communicated to them, with special attention to vulnerable persons or groups. Effective remedies should involve redress for any harm suffered, and may include measures under civil, administrative, or, where appropriate, criminal law. Moreover, because AI has a myriad of applications, any remedies need to be tailored towards those different applications. This should include the obligation to terminate unlawful conduct and guarantees of non-repetition, as well as the obligation to redress the damage caused.

❖ Key substantive rights

- The right to an effective remedy whenever the development or use of AI systems causes unjust harm or breaches an individual's legally protected human right.
- The right of an individual to challenge a decision that produces legal effects on individuals or similarly significantly affects them, and to an effective remedy, regardless of whether the decision is based on personal or non-personal data.

❖ Key obligations

- Member States must ensure that effective remedies are available, including where relevant civil liability and criminal responsibility, and that accessible redress mechanisms are put in place for individuals whose rights are negatively impacted by the development or use of a AI applications. In this regard, they should also consider the introduction of class actions in the context of harm caused by the use of AI systems.
- More generally, they should establish public oversight mechanisms for AI systems that may adversely affect human rights, democracy or the rule of law.
- Member States should ensure that developers and deployers of AI systems:
 - provide clear information on the availability of effective remedies in the case of harm caused through use of an AI system
 - identify, document and report on potential negative impacts of AI systems on human rights, democracy and the rule of law;
 - put in place adequate mitigation measures to ensure accountability for any caused harm.
- Member States should put in place measures to ensure that public authorities are always able to audit AI systems used by private actors¹¹⁷, so as to assess their compliance with existing legislation and to hold private actors accountable.

7.1.8 Democracy

103. In order to properly address the risks to democracy highlighted in Chapter 3, effective, transparent and inclusive democratic oversight mechanisms are needed to ensure that the democratic decision-making processes and the related values of pluralism, access to information, and autonomy are safeguarded.
104. Where relevant, member States should ensure a meaningful participatory approach and an involvement of different stakeholders (from civil society, the private sector, and the media) in the decision-making processes concerning the deployment of AI systems in the public sector, with special attention to the inclusion of under-represented and vulnerable individuals and groups, which is key to ensuring trust in the technology and its acceptance by all stakeholders.
105. In order to preserve the good functioning of democratic institutions and their representative character, in addition to participation, special safeguards need to be introduced as regards the way in which they are legitimately established and operating, by ensuring their legitimacy, transparency and accountability.
106. AI may also influence electoral processes negatively, through information disorder, that is misinformation, disinformation and malinformation, which can lead to the violation of the principles of free and fair elections, by unlawful interference in the equality of opportunities and the freedom of voters to form an opinion. It is crucial to ensure that electoral processes are in conformity with Council of Europe and other applicable international standards.

¹¹⁷ While business secrets and intellectual property rights must be respected, they must be balanced against other legitimate interests.

107. The use of AI systems can render public institutions more efficient, yet at the potential cost of providing less 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 such 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. The challenge is thus to establish a governance framework that allows 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. Including criteria such as equality, fairness, accountability and transparency in AI-related procurement processes is key, and introducing safeguards can serve two purposes. Firstly, it ensures that governments strictly only use systems that are compatible with the human rights, democracy and the rule of law, and secondly it also creates economic incentives for the private sector to develop and use systems that comply with this. When procuring AI systems for use in public institutions, member States have a special responsibility to ensure that this occurs in line with human rights, democracy and the rule of law. Since the use of AI systems in public services should be held to higher standards of transparency, public authorities should hence not acquire AI systems from third parties unwilling to waive information restrictions (e.g. confidentiality or trade secrets) where such restrictions impede the process of (i) carrying out human rights impact assessments (including carrying out external research/review) and (ii) making these assessments available to the public.

❖ Key substantive rights

- The right to vote and to be elected, the right to free and fair elections, and in particular universal, equal and free suffrage, including equality of opportunities and the freedom of voters to form an opinion. In this regard, individuals should not to be subjected to any deception or manipulation.
- The right to information, freedom of expression, freedom of assembly and association.
- The right to good governance.

❖ Key obligations

- 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, to shape voters' political behaviours or to manipulate public opinion.
- Member States should foster the use of AI solutions and other tools that can strengthen the informational autonomy of citizens, improve the way they collect information about political processes and help them participate therein.
- Member States should foster the use of AI solutions that can help fight corruption and economic crime, and that can enhance the legitimacy and functioning of democratic institutions. This can contribute to the positive impact of AI systems within the democratic sphere and enhance trust.
- Member States should ensure that their public procurement processes are subject to legally binding requirements that ensure the responsible use of AI in the public sector by safeguarding compliance with the above-mentioned principles, including transparency, fairness, and accountability.
- Member States should put in place measures to increase digital literacy and skills in all segments of the population, especially where AI systems are used in democratic processes.
- Member States should subject the procurement and application of AI systems in the public sector to adequate oversight mechanisms. This could include redress to ombudspersons and the courts.
- Member States should make public and accessible all relevant information on AI systems (including their functioning, optimisation functioning, underlying logic, type

of data used) that are used in the provision of public services, while safeguarding legitimate interests such as security.

- Member States should foster AI solutions that can help in the provision of public services while safeguarding human rights, democracy and the rule of law.

7.1.9 Rule of Law

108. The use of AI systems can increase the efficiency of judicial systems, but as noted in Chapter 3, it can also create significant challenges for the rule of law.¹¹⁸ According to the European Ethical Charter on the use of AI in the judicial systems and their environment¹¹⁹, when AI tools are used to resolve a dispute, or when they are used as tool to assist in judicial decision making or to give guidance to the public, it is essential to ensure that they do not undermine the guarantees of the right of access to a judge and the right to a fair trial.

109. In particular, this means that the principle of equality of arms and respect for the adversarial process must be safeguarded. Moreover, the use of AI systems should not undermine judicial independence and impartiality. To ensure this, the CEPEJ has underlined the importance of securing the quality and security of judicial decisions and data, as well as the transparency, impartiality and fairness of data processing methods. In addition, safeguards for the accessibility and explainability of data processing methods, including the possibility of external audits, must likewise be introduced. Member States should therefore subject the use of AI systems within the judicial system to thorough checks and ensure their compliance with all the above principles.

110. Whenever a legal dispute arises in the context of the use of an AI system – whether by the private or public sector – persons who file a claim related to a violation or harm caused through the use of an AI system must have access to relevant information in the possession of the defendant or a third party that is necessary to allow for an effective remedy. Access to relevant information by parties in judicial proceeding is also critical when AI systems have been used to support judicial decision-making, as this represents an important condition for preserving the equality of arms between the parties. This may include, where relevant, training and testing data, information on how the AI system was used, meaningful and understandable information on how the AI system reached a recommendation, decision or prediction, and details of how the AI system’s outputs were interpreted and acted on. In this regard, a fair balance must be sought between intellectual property and other rights of the defendant and the legitimate interests of those seeking redress. Moreover, individuals who seek redress for alleged violations of human rights in the context of AI systems should not be held to a higher standard of proof.

❖ Key substantive rights

- The right to a fair trial, including the possibility to get insight into and challenge an AI-informed decision in the context of law enforcement or justice, including the right to review of such decision by a human.
- The right to judicial independence and impartiality and the right to legal assistance.
- The right to an effective remedy in case of unlawful harm or breach an individual’s human rights in the context of AI systems.

❖ Key obligations

- Member States must ensure that AI systems used in the field of justice and law enforcement are in line with the essential requirements of the right to a fair trial. To this end, they should pay due regard to the need to ensure the quality and security of judicial decisions and data, as well as the transparency, impartiality and fairness

¹¹⁹ The analysis of the CEPEJ concerns the challenges arising from the use of AI systems also in the field of online dispute resolution and law enforcement.

of data processing methods. Safeguards for the accessibility and explainability of data processing methods, including the possibility of external audits, should be introduced to this end.

- Member States must ensure that effective remedies are available and that accessible redress mechanisms are put in place for individuals whose rights are violated through the development or use of AI systems in contexts relevant to the rule of law.
- Member States should provide meaningful information to individuals on the use of AI systems in the field of justice and law enforcement, both as concerns the role of AI systems within the process, and the right to challenge the decisions informed or made thereby.

7.2 Role and responsibilities of member States and private actors in the development of applications complying with these requirements

111. AI systems can affect human rights, democracy and the rule of law when being developed and used by private and public actors alike. As noted in Chapter 5, in addition to an obligation to protect human rights in the public sphere, member States may also have the positive obligation to ensure that private actors respect human rights' standards. Moreover, several international frameworks also oblige or encourage private actors may have a responsibility to respect human rights both under international instruments (such as the UN Guiding Principles on Business and Human Rights) and/or under their relevant national legal frameworks.

112. In the section above, the obligations of member States to ensure conformity with the Council of Europe's standards on human rights, democracy and the rule of law in the context of AI systems were already pointed out. More generally, national authorities should carry out evidence-based assessments of domestic legislation to verify its compliance with – and ability to protect duly – human rights and adopt measures to fill potential legal gaps. Moreover, they should establish control mechanisms and ensure effective judicial remedies for redress whenever the development and use of AI leads to violations of law. To this end, national oversight authorities should be able to audit and assess the functioning of (public or private) AI systems. Such oversight should complement existing oversight obligations in the context of existing legislation, including data protection law (the accountability principle, impact assessment, prior consultation with supervisory authorities, etc) to increase transparency. There may be limited circumstances where, due to concerns around privacy or intellectual property, a degree of confidentiality is required.

113. It should be noted that many public actors procure AI systems from private actors and thus obtain relevant data to deploy it, and access to the underlying infrastructure on which AI systems can operate from them. Accordingly, given the essential role of private actors in this field, they have a responsibility to ensure that their systems are developed and used in line with the above principles, rights and requirements. As the interests of commercial private actors on the one hand, and of individuals and society on the other hand, are not always aligned, a legal structure that would oblige private actors to comply with specific rights and requirements in the context of AI may be appropriate. Moreover, this would secure access to justice should they fail to meet these obligations.¹²⁰

114. As noted above, when member States take measures to safeguard the listed principles, rights and requirements in the context of AI, a risk-based approach – complemented with a precautionary approach where needed – is recommended. Such approach acknowledges that not all AI systems pose an equally high level of risk, and that regulatory measures should take this into account. Moreover, it

¹²⁰ C. Muller, p. 16; this means going beyond merely referring to the Recommendation CM/Rec(2016)3 on human rights and business of the Committee of Ministers of the Council of Europe (and the UN Guiding Principles on Business and Human Rights).

requires that the risks posed by AI systems to human rights, democracy and the rule of law, are assessed on a systematic basis and that mitigating measures are specifically tailored thereto.

115. When implementing a risk-based approach and assessing the type of regulatory intervention needed to mitigate risks, Member States can be guided by a number of factors that are commonly used in risk-impact assessments. These risk-factors include, for instance, the potential extent of the adverse effects on human rights, democracy and the rule of law; the likelihood or probability that an adverse impact occurs; the scale and ubiquity of such impact; its geographical reach; its temporal extension; and the extent to which the potential adverse effects are reversible. In addition, a number of AI-specific factors that can influence the risk level (such as the application's level of automation, the underlying AI technique, the availability of testing mechanisms, the level of opacity) can also be considered.

7.3 Liability for damage caused by artificial intelligence

116. The development and use of AI systems raises new challenges in terms of safety and liability. Views differ, however, as to whether existing liability regimes should apply, or whether specific regimes should be developed for the context of AI. Nevertheless, it can be noted that the widespread use of AI systems may raise some challenges to the interpretation and implementation of existing liability legislation. For example, the Council of Europe Convention on Products Liability (ETS No. 91) for instance only applies to AI systems that are considered to be movable products (hardware) rather than software, and only applies to AI systems offered as a product rather than a service. Therefore, a clarification that stand-alone software can be qualified as a product within the meaning of existing product liability law might be advisable. The opacity of some AI systems, coupled with the asymmetry of information between AI developers and producers on the one hand and individuals who may be negatively impacted by AI systems on the other hand, may in certain cases make it difficult for the latter to meet the standard of proof required to support a claim for damages. However, in general, the existing assignment of the burden of proof can bring about appropriate and reasonable solutions with regard to AI systems.

117. If the CAHAI decides to address the question of liability in a future legal framework at the level of the Council of Europe, the CAHAI recommends that the following aspects be considered:

- A proper and balanced liability regime is important for both consumers and manufacturers, and can contribute to legal certainty.
- It is essential to guarantee the same level of protection to persons harmed through the use of an AI systems as those harmed through the use of traditional technologies.
- Liability for any unjust harm should be able to arise from any unjust harm occurring throughout the entire life cycle of the AI system.
- There should be a clear allocation of liability between actors involved in the development and operation of AI (creators, developers, deployers, operators, utilizers and users), as well as certification bodies where applicable.
- The issue of trans-border responsibility should be taken into account. This is particularly relevant when, for instance, a company using an AI system is registered in one state, the developer of that system in another state, and a user suffers harm is habitually resident in a third state.
- The rules for liability may be supplemented, in some sector specific applications, by industry ethical codes of conduct which would serve the purpose of enhancing public trust in sensitive areas of AI.

8. POSSIBLE OPTIONS FOR A COUNCIL OF EUROPE LEGAL FRAMEWORK FOR THE DESIGN, DEVELOPMENT AND APPLICATION OF ARTIFICIAL INTELLIGENCE BASED ON HUMAN RIGHTS, DEMOCRACY AND THE RULE OF LAW

116. In order to fill the gaps in legal protection identified in Chapter 5, a number of different options for a legal framework are available within the Council of Europe, including binding and non-binding legal instruments. These instruments, and their advantages and disadvantages, are outlined below. Whereas the previous chapter focused on the *substance* of the legal framework, this chapter focuses on its *format*.

8.1 Modernising existing binding legal instruments

117. A first option that could be considered is to amend existing binding legal instruments, to complement and/or adapt them in light of the particularities of AI systems.

118. An additional protocol to the ECHR could be adopted to enshrine new or adapt existing human rights in relation to AI systems. These could be drawn from Chapter 7 above, and for instance include¹²¹: a right to human autonomy, agency and oversight over AI; a right to transparency / explainability of AI outcomes, including the right to an explanation of how an AI system functions; a separate right to physical, psychological and moral integrity in light of AI-profiling and affect recognition; a right to refuse to be subjected to profiling, to have one's location tracked, to be manipulated or influenced by a "coach"; a strengthened right to privacy to protect against AI-driven mass surveillance; the right to have the opportunity, in the context of care and assistance provided to elderly people and people with disabilities, to choose to have contact with a human being rather than a robot. It is not unlikely that, under the dynamic and evolutive interpretation adopted by the ECtHR, existing ECHR rights, such as the right to private life, freedom of thought and of expression, and the right to non-discrimination may be interpreted so as to include some of the aforementioned rights. The advantage of an additional protocol, however, is that the recognition of certain rights in relation to AI would not depend on a judgment of the European Court of Human Rights (ECtHR), and hence, would offer more clarity and legal certainty (also avoiding possible criticism of the ECtHR for interpreting Convention rights too expansively). While the adoption of an additional protocol would affirm, in the strongest possible manner, the member States' commitment to protecting key substantive rights of citizens, the rule of law and democracy, in relation to AI-systems, it would not be an appropriate instrument to lay down specific requirements or obligations. It should also be noted that additional protocols to the ECHR are binding only upon those states that ratify them, which may result in only some member States being bound and a fragmentary oversight by the ECtHR.

119. Modernising existing vertical instruments, such as the Budapest Convention on Cybercrime ([CETS No.185](#)) or "[Convention 108+](#)", could be another plausible scenario. An important advantage of this approach – compared to drafting a new convention (see below) – is that existing networks for monitoring and enforcement (like in the case of Convention 108+ the national data protection independent authorities, whose scope of regulatory activities could be expanded to artificial intelligence) could be mobilised. The drawback of this approach, however, in addition to the length and complexity of adoption, lies with the limited scope of each existing vertical or sectoral instrument, which necessitates multiple interventions in order to tackle the various concerns discussed in previous chapters. Modernising "Convention 108+", for example, would not capture all concerns in relation to AI systems, given its (current) specific focus on the protection of individuals, and the processing of personal data; at the same time, it should be noted that many of the high level principles so far identified to face the challenges raised by AI systems (e.g. accountability, transparency, automated decisions) are already included in Convention 108+. Moreover, since "Convention 108+" was open for ratification in 2018, it might be difficult to modernise it again in the short term¹²².

¹²¹ See CAHAI (2020)06-fin, § 77.

¹²² To note in this respect that entry into force of amendments or of an amending protocol would normally require the acceptance/ratification of all Parties to the Convention, which is a lengthy process. Convention 108 was amended in October 2018 by Protocol 223. There are two options for when Convention 108+ will come into force. One involves ratification by all existing 55 Parties; the other could see it in force between ratifying Parties as early as 2023 if at least thirty-eight Parties have expressed their consent to be bound by the Protocol. At the time being, 8 Parties to the Convention have ratified the Protocol.

120. The two concerns expressed for each option could be addressed by combining both ideas, i.e. of an additional protocol to the ECHR with modernising (certain) vertical instruments, like “Convention 108+”. Whereas the first would lay down overarching principles and values, the latter could elaborate on the positive and negative obligations of states and establish an effective network of independent competent authorities to ensure the effective implementation of those safeguards. These authorities could deal with acts or omissions of states as regards AI systems and engage the state’s responsibility under the Convention under some circumstances. The lengthy character of a combined process remains however an issue, against the background of the fast-paced rollout of AI systems.

8.2 Adoption of a new binding legal instrument: Convention or Framework Convention

121. A second option to be considered is the adoption of a new binding legal instrument, which could take the form of a convention or framework convention. It is worth noting that both conventions and framework conventions are multilateral treaties, they have the same legal nature and are both subject to the usual rules for international treaties as set out in the Vienna Convention on the Law of Treaties (1969). Moreover, both may include a system of governance (see for more details Chapter 9.5) and can be complemented by additional protocols. The difference between the two is that a convention tends to regulate a specific issue or area in a more concrete way, typically by setting out certain rights and obligations, whereas a framework convention rather tends to set out broader principles and areas for action which have been agreed between states party.

122. A framework convention typically only foresees a general duty for state parties to undertake certain actions, achieve certain objectives, or to recognise certain rights, without attributing such rights directly to natural or legal persons. Hence, the national ratification of a framework convention would not suffice for natural and legal persons to be able to invoke certain rights, and additional legislative action by the individual states would be needed. There is consequently a considerable margin of discretion for states as to how they implement the broader principles and objectives.

123. A convention could more exhaustively regulate the design, development and application of AI systems or of algorithmic decision making more generally, building further on Recommendation CM/Rec(2020)1.¹²³ It could list certain rights and obligations that could help safeguard the protection of human rights, democracy and the rule of law in the context of AI systems, and thereby offer legal protection to both natural and legal persons once it has been ratified and has entered into force. It could stress the importance of a speedy accession by the maximum number of Parties to facilitate the formation of a comprehensive legal regime for AI systems as specified under the convention, and it would urge member States and other Parties to the convention to initiate the process under their national law leading to ratification, approval or acceptance of the Convention. It is worth noting that, in October 2020, the Parliamentary Assembly of the Council of Europe (PACE) recommended “that the Committee of Ministers support the elaboration of a “legally binding instrument” governing AI, possibly in the form of a convention”.¹²⁴

124. The added value would be to get a specific legally binding instrument on the design, development and application of AI based on the Council of Europe’s standards on human rights, rule of law and democracy. It would harmonise rules and obligations across states on AI deployment, as well as establish a clear agreement regarding AI research and development procedures. Successful examples of such innovative legal frameworks developed by the Council of Europe in the past include the [Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data \(CETS No. 108\)](#), and the [Budapest Convention on Cybercrime \(CETS No.185\)](#).

¹²³ Council of Europe, Committee of Ministers, Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, April 2020.

¹²⁴ For the adopted resolution and recommendation in this regard, see <https://pace.coe.int/en/news/8059/establishing-a-legally-binding-instrument-for-democratic-governance-of-ai>.

125. At the same time, it may be considered premature to attempt to draft a convention containing detailed legal obligations in relation to AI systems. An overly prescriptive and rigid approach could lead to a rejection of the instrument and a lack of willingness on the part of signatories to ratify and actually implement the convention in practice. Conversely, states that do ratify the convention might discourage developments in the field. Overly rigid rules could stymie innovation and curtail research into, and the development and deployment of new technologies and cutting-edge solutions to existing problems, many of which could save lives and benefit society as a whole.
126. The latter concern could be addressed by ensuring that the rights and obligations that are set out in the convention are not overly prescriptive or detailed. Alternatively, it could also be addressed by adopting a framework convention on AI, which would provide for broad core principles and values to be respected as regards the design, development and application of AI to be enshrined in a binding instrument, in line with Council of Europe's standards and leave a broad margin of discretion to states parties in their respective implementation. Under the so-called "framework convention and protocol approach", parties could agree on a more general treaty – a framework convention – and when in the future they wish to do so, decide to elaborate more detailed protocols or other instruments to enact specific provisions. This regulatory technique, which has a number of benefits compared to single "piecemeal" treaties in international law, could be particularly appropriate in the field of AI. In this context, it should be carefully considered whether such a structure based on a more general treaty and the possible elaboration of additional specific instruments (such as protocols) would increase the complexity of the resulting legal framework and make compliance more challenging.
127. A framework convention would require the states to agree mutually on the scope of the legal framework and the procedure to be complied with to offer effective safeguards in the design, development and application of AI systems, based on Council of Europe's standards. It could contain the commonly agreed upon core principles and rules for AI research, development and implementation, in the interests of human society. It could also contain specific rules on procedural safeguards, preventive measures, jurisdiction, international co-operation. For instance, it could include provisions to allow for the exchange of information, or for already existing independent competent authorities like the ones dedicated to data protection or competition supervision at the national level to be mobilised. The framework convention could also set forth the rules and procedures necessary for states to implement it.
128. An existing example of such a framework convention at Council of Europe level is the Framework Convention for the protection of national minorities ([FCNM](#)). The FCNM is a legally binding instrument under international law and provides for a monitoring system, but the word "framework" highlights the scope for member States to translate the Convention's provisions to their specific country situation through national legislation and appropriate governmental policies. Another example, albeit not officially carrying the term "framework" in its title, is the Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine, in short, the Convention on Human Rights and Biomedicine (the so-called "[Oviedo Convention](#)"). This Convention was adopted in 1997 to respond to potential misuses of scientific advances in the field of biology and medicine. It draws on the principles established by the ECHR and aims at protecting the dignity and identity of all human beings. It sets out fundamental principles applicable to daily medical practice and also deals specifically with biomedical research, genetics and transplantation of organ and tissues. It is further elaborated and complemented by additional protocols on specific subjects, for instance, on the prohibition of cloning human beings.
129. Irrespective of the choice of a convention or framework convention, it should be noted that the addressees are states and state bodies, in particular the Council of Europe's member States. Accession can, however, be opened to other states, which is of course a considerable added value and

can significantly contribute to the global reach and effectiveness of the instrument¹²⁵. The strength of treaties lies in their formality and the fact that they are legally binding on those states which have ratified them. States becoming parties to a convention incur legal obligations which are enforceable under international law.

130. The foregoing does not rule out the important role that private actors could play in the implementation of specific regulations implemented at the national level on the basis of broad international commitments. In particular, they could take up a prominent role in the design of co-regulatory mechanisms by which states would, in close interaction with private actors, give further shape to their international commitments.
131. Whereas no general conclusion can be drawn as to how long it takes to prepare and negotiate a Convention, and for it to be ratified and enter into force (this may range from a couple of months as in the case of the European Convention on Spectator Violence and Misbehaviour at Sport Events and in particular at Football Matches ([ETS No. 120, 1985](#)) or the Additional Protocol to the Oviedo Convention on a Prohibition of Cloning Human Beings ([ETS No.168, 1998](#)) to several years, depending on the nature and complexity of the issues to be solved, but also on the political will of member States), a potential weakness of international treaties is the slowness of the ratification process. Even if a member State votes in favour of adoption in the Committee of Ministers text, it is under no obligation to ratify. There is therefore no way to guarantee that any or all member States will ratify the treaty.

8.3 Non-binding legal instruments

8.3.1 Council of Europe

132. A distinction should be made between non-binding (or soft law) instruments at the level of the Council of Europe and at the national level. The former are already relied upon in several sectors (cf. Chapter 4), but could be complemented with a general soft law instrument, such as a recommendation or declaration, that consolidates common principles. Such a soft law instrument could operate as a stand-alone document or complement a binding instrument to further operationalise its provisions. Other options include the drafting of guidance documents to increase the understanding of the relationship between the protection of human rights, democracy and rule of law, and AI (e.g. by providing information about case law of the ECtHR), and to thereby contribute to strengthening protection at national level. Such 'manuals' or 'guides' could be developed through broad multi-stakeholder consultation with governments, private companies, civil society organisations and representatives of the technical community and academia. These documents should be evolving, updated periodically, and fleshed out collaboratively in light of new developments. Precedents include the [Manual on Human Rights and the Environment](#) and the [Guide to Human Rights for Internet Users](#).

8.3.2 Member State level

133. Soft law mechanisms approved by national competent authorities could be encouraged by a Council of Europe legal framework, to further operationalise it and demonstrate compliance. These soft law instruments could consist of approved guidelines, codes of conduct, labelling mechanisms, marks and seals, as well as certification mechanisms. Whereas soft-law measures cannot, due to their non-binding character, meet the objectives of ensuring that AI applications protect and advance human rights, democracy and the rule of law, they can make important contributions thereto. The advantages of a soft law approach include flexibility, adaptability, immediacy of implementation, broader appeal, and capacity to be reviewed and amended quickly
134. Private actors, including academic institutions and standard-setting bodies, can help ensure that such soft-law instruments are practically effective. Organisations developing and deploying AI can

¹²⁵ In general, non-member States of the Council of Europe can accede to CoE convention through an invitation by the Committee of Ministers and after having obtained the unanimous agreement of the Parties to the Convention. Some Council of Europe's Convention have become global standards: "Convention 108" has 55 States parties, whereas the "Budapest Convention" has 64 States parties.

incorporate soft-law instruments into their governance structure, procurement process, operation, and auditing practices (as they already do with many standards and certifications related, for example, to security). In addition, rating agencies could also play a role, for instance by providing an annual ranking of private organisations complying with soft-law requirements based on sound evidence.

135. However, it should be stressed that while self-regulation might be a complementary method of implementing certain principles and rules, it cannot substitute the positive obligations that member States have under the ECHR to effectively protect and safeguard human rights, democracy and the rule of law in relation to AI. Voluntary, self-regulatory and ethics-based approaches lack effective mechanisms of enforcement and accountability, and should therefore, on their own, not be considered as a sufficient and effective means to regulate AI. Moreover, certification mechanisms are not immune to errors and mistakes. Hence, for these mechanisms to be effective, a number of conditions should be fulfilled.

8.3 Other type of support to member States such as identification of best practices

136. There are numerous ways in which best practices can be identified or encouraged (many of which are familiar to or implemented by member States or economic actors). A European Benchmarking Institute could be a highly effective, efficient, and trustworthy source of identification, definition, and consensus around the underlying evidence that should guide sound best practices. Such evidence can, in turn, serve as the basis for a wide range of best practices that can be efficiently and effectively propagated by sound technical standards and certifications. The added value of such an Institute in respect of other standard setting organisations such as the ISO and the IEC would have nonetheless to be carefully considered.

137. In addition, a uniform model developed at the level of the Council of Europe for a human rights, democracy and rule of law impact assessment could be extremely helpful in harmonising member States' implementation of common values in relation to AI systems. As concerns practical mechanisms that can help to both implement and enforce the legal framework, we refer to Chapter 9 of this feasibility study.

8.4. Possible complementarity between the horizontal and cross-cutting elements that could form part of a conventional-type instrument and the vertical and sectoral work that could give rise to specific instruments of a different nature.

138. Chapter 4 has described the Council of Europe's sectorial work on AI systems, which is expected to develop further in the coming years. The horizontal elements which could form part of a convention-type instrument would help finetune sectorial work and provide impetus to the development of specific instruments in areas where the analysis of the impact of AI systems and of the required policy responses is advancing. A potential horizontal binding legal instrument could include explicit references to the existing or future instruments in the different areas of work of the Council of Europe.

139. Another mechanism to ensure complementarity could be the setting up of a joint certification scheme/body, comparable to the one existing in the pharmaceutical sector (the [European Directorate for the Quality of Medicines \(EDQM\) and HealthCare](#) and its Pharmacopoeia). Such joint certification mechanism/body could be tasked with providing more detailed guidelines regarding human rights, democracy and rule of law impact assessments and common quality standards at European level. Moreover, it could be responsible for supporting the implementation and monitoring the application of quality standards for AI systems, just like EDQM does for safe medicines and their safe use.

140. In conclusion, given the evolving nature and the challenges posed by AI, an appropriate legal framework would likely need to consist of a combination of binding and non-binding legal instruments, that complement one another. A binding horizontal instrument, i.e. a convention or framework convention, could consolidate general common principles that would be contextualised to apply to the AI environment and include more concrete provisions to safeguard the rights, principles and obligations identified in Chapter 7. This instrument, which could include appropriate follow-up mechanisms and

processes, could be combined with additional (binding or non-binding) sectoral Council of Europe instruments establishing further sector specific principles and detailed requirements on how to address specific sectoral challenges of AI. This combination would allow the required level of guidance to private actors who wish to undertake self-regulatory initiatives to be provided. This approach would also allow for the flexibility required for technological development, as revisions to the vertical instruments could be undertaken with relatively less formality and complexity.

9. POSSIBLE PRACTICAL AND FOLLOW-UP MECHANISMS TO ENSURE COMPLIANCE AND EFFECTIVENESS OF THE LEGAL FRAMEWORK

9.1 The Role of Compliance Mechanisms

141. The ultimate effectiveness of any legal framework will depend on the breadth of its adoption and compliance. Practical mechanisms (such as impact assessments, lifecycle auditing, and monitoring, certification methods, and sandboxes) are one way of driving such compliance and of helping member States to understand and monitor adherence to the legal framework. Such mechanisms confer further benefits beyond compliance, for example by increasing transparency around the use of AI and creating a common framework for promoting trust.

142. Any Council of Europe legal framework should formulate the abstract requirement to develop compliance mechanisms at a general level as well as what principles need to be fulfilled by any practical mechanisms to ensure compliance. It would be for state parties to decide how to enforce this through their legislative framework, including which practical mechanisms they choose to make mandatory or which actors or institutions they empower to provide independent, expert, and effective oversight. This would enable implementation to account for the existing roles of local institutions, regulatory culture, and legal requirements. Rather than mandating a single solution, this approach would further enable the creation of an AI assurance ecosystem, which would create the potential for diverse participation and the emergence of novel and innovative approaches to compliance. That said, collaboration between state parties should be considered paramount to protect against the risk of diverging approaches and the resulting fragmentation of markets.

143. Compliance mechanisms might be used to assess the design of an AI-enabled system, as well as its operational processes, contextual implementation and use case. On the question of when AI systems should be subject to such assessment, the CAHAI-PDG agreed on the fundamental importance of *ex ante* assessment and continuous assessment at various milestones throughout the AI project lifecycle, including after initial deployment and use. Compliance mechanisms should also evolve over time to account for the evolving nature of the system.

144. The ongoing assessment approach presents three salient advantages. First, it allows for a better understanding of the implications of any AI system (throughout its design, development, and deployment). Second, it facilitates decision making to reconsider future unforeseen uses of an AI system. Third, it monitors changes in the behaviour of the model *ex post* (which is particularly crucial in e.g. reinforcement learning contexts and dynamic learning systems). In particular, the procurement and use of pre-built AI-enabled solutions and technical advancements such as transfer learning applications presents challenges that need to be considered.

9.2 The Role of Different Actors

145. As outlined above, each member State should ensure national regulatory compliance with any future legal framework. Different actors should contribute in a complementary way to bring about a new culture of AI applications that are compliant with the legal framework's principles and local regulations to generate adequate incentives for compliance and oversight incentives, either as assurers, developers, or operators and users.

9.2.1 Assurers of systems

146. Member States should also be responsible for identifying and empowering independent actors to provide oversight. These independent actors should represent and be accountable to clearly identified stakeholder groups affected by practical applications of AI, and could be, as appropriate, an expert committee, academics, sectoral regulators or private sector auditors. Where they do not exist already, member States might consider setting up independent oversight bodies equipped with appropriate and adequate inter-disciplinary expertise, competencies, and resources to carry out their oversight function. Such bodies might be equipped with intervening powers and be required to report to parliament and publish reports about their activities regularly¹²⁶. They might also resolve disputes on behalf of citizens or consumers. For example, states could extend the mandate of existing ombudsmen institutions or create a new ombudsman institution to assess and resolve any complaints or appeals as a complement to binding judicial mechanisms. It is unreasonable to expect that any such body could cover all AI-based products and systems, and so consideration as to scope would be important.
147. Many AI systems are deployed across multiple jurisdictions. It is vital for adequate oversight to share information among the member States. Mechanisms of information sharing and reporting about AI systems could be included in each State's regulatory framework (e.g. information on certified AI systems, banned AI applications or the current status of a specific AI application). Private sector actors could also play a role in assuring systems.
148. In addition to auditing services, certification schemes can support a legal framework and promote an active role for the private sector to prevent and manage the risks of adverse human rights impacts associated with AI systems. Indeed, more generally, certification mechanisms are highly versatile and can provide evidence-based instruments upon which governance regimes can be flexibly developed to meet the needs of different domains and the allowances of national regulatory regimes. Standards and certifications can be developed for all stages of AI development and operations and may engage all agents involved in order to implement certain requirements. Incentives can be created for private actors to adopt such instruments promptly, including through the procurement practices of intergovernmental organisations and of national public sector entities. When duly implemented, they can help empower ordinary citizens by serving as the "currency of trust" that both experts and non-experts can relate to (as with nutritional labels or car safety crash-tests). The underlying evidence sought by such standards and certifications can also be used to spur, accelerate, and reward innovation through open, recurring, AI-innovation benchmarking initiatives.
149. Within certification schemes, professional training could include the legal framework as part of the training curricula. In broader terms, universities and civil society could be part of education policy to disseminate, research and instruct on AI's legal framework and technical developments. This approach would also confer further benefits in a global market economy.
150. Furthermore, professional certification at the level of developers and of systems may be another strategy for assuring that AI is used in line with the Council of Europe standards of human rights, democracy and the rule of law. This certification mechanism could be similar to the one already existing for doctors, lawyers or for (certain types of) engineers.

9.2.2 Developers of systems

151. Actors building AI-enabled systems (both private and public sector) should consider actions they can take to increase compliance with a future legal framework. For example, policies can be adopted to increase the visibility of where such technologies are being deployed, in particular by publishing public sector contracts, or by establishing public registers¹²⁷ or notification systems) or developing

¹²⁶ See the Recommendation of the Council of Europe Human Rights Commissioner on "[Unboxing AI: 10 steps to protect human rights](#)".

¹²⁷ Such registers already exist in the Netherlands and in the UK: <https://algoritmeregister.amsterdam.nl/>; <https://ai.hel.fi/en/ai-register/>.

norms and standardised tools for internal audit and self-certification (all the while acknowledging the limitations of this approach). Liability considerations should also be taken into account.

9.2.3 Operators and Users of systems

152. Operators and users of AI could generate demand for AI applications that comply with the future legal framework. This is particularly true of the public sector and its relative procurement power. The promotion of trust carriers, such as certification labels on AI systems' lifecycles, and periodic auditing and reporting, are market responses pushed by operators and users of AI systems' preferences and expectations. When operators and users of AI systems become better informed of their rights and redress mechanisms, the transaction cost of oversight is significantly reduced.

9.3 Examples of Types of Compliance Mechanism

153. There are many contexts where organisations are already required to meet standards or regulations, such as, for example, financial services and healthcare. Each of these has evolved into ecosystems of services that allow organisations to prove to themselves, their customers, and regulators that they met a required standard. Different mechanisms will work best in different contexts, depending on existing infrastructure, sectoral mechanisms and institutions. It should also be considered which components of an AI system can be subject to compliance, for example, the training data used, the algorithm construction, the weighting of different inputs or the accuracy of any outputs. Inclusive participatory processes should be conducted to establish the relevant regulatory and enforcement mechanisms in each case.

154. A future legal framework might specify that practical mechanisms adhere to a set of principles that promote the framework's core values. These might include:

- **Dynamic (not static):** assessment *ex ante* and at various points throughout the AI project lifecycle to account for choices made during the design, development and deployment processes and any changes in the application-behaviour of dynamic learning models.
- **Technology adaptive:** to support the future-proofing of any compliance mechanisms.
- **Differentially accessible:** understandable to experts and non-experts, in turn simplifying the process of any potential appeals and redress.
- **Independent:** conducted, or overseen, by an independent party.
- **Evidence-based:** supported on evidence produced by technical standards and certifications. For example, including data collected through best practices such as borderless, standardization or key metrics developed through benchmarking.

155. Any mechanisms need to be implementable in practice and account for existing governance infrastructure and technical limitations. The practical mechanisms outlined below should therefore be considered as a toolkit that presents ample opportunity for further regulatory innovation and refinement:

(1) Human rights impact assessments - Conducting human rights due diligence is recommended by the United Nations Guiding Principles on Business and Human Rights (UNGPs). Moreover, The Council of Europe' Recommendation on the human rights impact of algorithmic systems has recommended that public and private organisations perform a fundamental rights impact assessment. These assessments might explicitly validate conformity with principles outlined in a future legal framework. In specific contexts, 'integrated impact assessments' might be deemed more appropriate to reduce the administrative burden on development teams (bringing together, for example, human rights, data protection, transparency, accountability, competence, and equalities considerations).

(2) Certification & Quality Labelling - *Ex ante* obligations, administered by recognised bodies and independently reviewed, would help build trust. A distinction could be made between standards and certifications that can apply to (i) products / AI systems or (ii) organisations developing or using AI systems. An expiration date would ensure systems are re-reviewed regularly. Such schemes could be made voluntary or mandatory, depending on the maturity of the ecosystem.

Legal safeguards must ensure certifications are not used by companies to shield themselves from potential liability claims associated with their conduct. The certification process should be subject to regulation regarding auditors' qualifications, the standards adopted, and how conflicts of interests are managed. The certification process should strive for continuous improvement and be responsive to complaints. Ongoing multi-stakeholder standards development work would support this led by standard-setting bodies.

- (3) **Audits** - Regular independent assessments or audits of AI-enabled systems by experts or accredited groups is also a mechanism that should be exercised throughout the lifecycle of every AI-enabled system to verify their integrity, impact, robustness, and absence of bias. Audits will facilitate a move towards more transparent and accountable use of AI-enabled systems. Audits could certify organisations as a whole, rather than just specific use cases.
- (4) **Regulatory Sandboxes**¹²⁸ - Regulatory sandboxes, particularly those that enable closer regulatory support, present an agile and safe approach to testing new technologies and could be used in order to strengthen innovative capacity in the field of AI. Sandboxes could be of particular use where a timely, possibly limited market introduction appears warranted for public welfare reasons, e.g. in extraordinary crises such as a pandemic, or in cases where current legal frameworks have not been tested in practice that could lead to constrained innovation. Cross-jurisdictional sandboxes present further opportunities for collaboration, building on the model of the Global Financial Innovation Network¹²⁹.
- (5) **Continuous, automated monitoring** – Automated systems can be deployed in parallel to AI-enabled systems to continuously monitor and assess its operation to guarantee compliance of established norms.

9.4 Other types of support to ensure compliance, such as the identification of best practices

- 155. While like-minded common values are being promoted by some member States or international organisations in which they are active, a cross-border agreement of international legal framework on AI systems is needed because of the virtual impact of AI on member State's jurisdictions and the challenges of global competition on the market. There are numerous ways in which best practices can be identified or encouraged (many of which are familiar or already implemented by member States or economic actors). A European benchmarking institute could be a highly effective, efficient, and trustworthy source of identification, definition, and consensus around the underlying evidence that should guide sound best practices. Such evidence can, in turn, serve as the basis for a wide range of best practices that can be efficiently and effectively propagated by sound technical standards and certifications.
- 156. In addition, a uniform model developed at the level of the Council of Europe to carry out a human rights, democracy and rule of law impact assessment could be extremely helpful in harmonising member States' implementation of common values in relation to AI systems.
- 157. In conclusion, mandating practical mechanisms to enforce compliance should be considered only one part of a broader package of initiatives required to drive change. Member States could reinforce compliance mechanisms with several initiatives. For example, to invest in digital literacy, skilling up and building competencies and capacities of developers, policymakers and wider society to understand the human rights implications of AI-enabled systems; to drive the widespread adoption of norms such as

¹²⁸ Sandboxes shall be understood as concrete frameworks which, by providing a structured context for experimentation, enable in a real-world environment the testing of innovative technologies, products, services or approaches especially in the context of digitalisation for a limited time and generally in a limited part of a sector or area under regulatory supervision of the respective authority ensuring that appropriate safeguards are in place.

¹²⁹ <https://www.fca.org.uk/firms/innovation/global-financial-innovation-network>

open access to source code; or engaging with human rights civil society organisations as key stakeholders at various stages of development¹³⁰.

158. This more comprehensive work to develop best practices and norms within existing legal and regulatory regimes should be accompanied by ongoing discourse, collaboration, and best practice sharing between actors at national and international level. Centres of expertise would be well placed to facilitate collaboration on innovative solutions to inter-sectoral regulation projects¹³¹.

9.5 Follow-up mechanisms

159. In addition to the above-mentioned practical mechanisms, the CAHAI-PDG has taken note of the variety of follow-up mechanisms and processes, as well as measures for international co-operation which are envisaged under the Council of Europe's legal instruments, of which the features vary according to the type and contents of such instruments.
160. As regards follow-up mechanisms and processes, the CAHAI-PDG noted that they can include, for instance, the appointment of one or more entities – such as independent expert groups, conventional committees, standing committees, consultative committees and committees of parties – that can be in charge of tasks such as monitoring the implementation of a given convention, facilitating the effective use and implementation of a convention, and exchanging information and good practices on significant legal, policy or technological developments pertaining to a given area
161. As to potential measures of international co-operation, these could include the appointment of points of contact or the creation of networks among the state parties to advance mutual assistance and co-operation in criminal or civil matters.
162. While identifying precise solutions would be too premature at this stage, and bearing in mind that the concrete features of follow-up mechanisms and processes will depend on the nature and substantive elements of the chosen legal instrument(s), the CAHAI-PDG recommends that the CAHAI ensures that a future legal framework on AI includes appropriate follow-up mechanisms and processes, as well as measures for international co-operation, in line with the Council of Europe's legal standards and practice. This is of key importance to guarantee the effectiveness of the main principles, rights and obligations set out in Chapter 7 at international level, and to complement the practical and oversight measures described earlier in this chapter, which can be implemented at domestic level.

10. FINAL CONSIDERATIONS

163. This study has confirmed that AI systems can provide major opportunities for individual and societal development as well as for human rights, democracy and the rule of law. At the same time, it also confirmed that AI systems can have a negative impact on several human rights protected by the ECHR and other Council of Europe instruments, as well as on democracy and the rule of law. The study has noted that no international legal instrument specifically tailored to the challenges posed by AI exists, and that there are gaps in the current level of protection provided by existing international and national instruments. The study has identified the principles, rights and obligations which could become the main elements of a future legal framework for the design, development and application of AI, based on Council of Europe standards, which the CAHAI has been entrusted to develop.
164. A robust legal framework will likely consist of a combination of binding and non-binding legal instruments, that complement each other. A binding instrument, a convention or framework convention, of horizontal character, could consolidate general common principles – contextualised to apply to the AI environment and using a risk-based approach – and include more granular provisions in line with the rights, principles and obligations identified in this feasibility study.
165. This instrument could be combined with additional binding or non-binding sectoral Council of Europe instruments to address challenges brought by AI systems in specific sectors. This combination

¹³⁰ CAHAI(2020)21 rev PDG contributions p.45-46.

¹³¹ CAHAI(2020)21 rev PDG contributions p.32-33.

would also allow legal certainty for AI stakeholders to be enhanced, and provide the required legal guidance to private actors wishing to undertake self-regulatory initiatives. Moreover, by establishing common norms at an international level, transboundary trust in AI products and services would be ensured, thereby guaranteeing that the benefits generated by AI systems can travel across national borders. It is important that any legal framework includes practical mechanisms to mitigate risks arising from AI systems, as well as appropriate follow-up mechanisms and processes and measures for international co-operation.

166. The Committee of Ministers is invited to take note of this feasibility study and to instruct the CAHAI to focus its work on the elaboration of the main elements of a binding legal framework, in parallel with progress that can be made on sectoral instruments.